Historical Linguistics 2017

EDITED BY

Bridget Drinka

HISTORICAL LINGUISTICS 2017

CURRENT ISSUES IN LINGUISTIC THEORY

Current Issues in Linguistic Theory (CILT) is a theory-oriented series which welcomes contributions from scholars who have significant proposals that advance our understanding of language, its structure, its function and especially its historical development. CILT offers an outlet for meaningful contributions to current linguistic debate.

A complete list of titles in this series can be found on *benjamins.com/catalog/cilt*

Volume 350

Bridget Drinka (ed.)

Historical Linguistics 2017. Selected papers from the 23rd International Conference on Historical Linguistics, San Antonio, Texas, 31 July – 4 August 2017

HISTORICAL LINGUISTICS 2017

SELECTED PAPERS FROM
THE 23RD INTERNATIONAL CONFERENCE ON
HISTORICAL LINGUISTICS,
SAN ANTONIO, TEXAS, 31 JULY – 4 AUGUST 2017

Edited by

BRIDGET DRINKA
University of Texas at San Antonio

JOHN BENJAMINS PUBLISHING COMPANY
AMSTERDAM & PHILADELPHIA

Table of contents

Part VI. Languages in contact

Foreword & Acknowledgements

The 23rd International Conference on Historical Linguistics (ICHL23) took place in San Antonio, Texas, July 31–August 4, 2017, hosted by the University of Texas at San Antonio. We welcomed 300 participants from 40 countries to the conference, which featured eight plenary speakers, nine workshops, four "Tricentennial Panels," several special sessions, and a roundtable on New Directions for Historical Linguistics.

The plenary lectures represented a broad range of topics, languages, and theoretical orientations, from contact, maintenance, and diversification among the endangered languages of Amazonia to the ecology of language change in African American Vernacular English, from the role of Australian languages in the formulation of new theories of language change to an examination of contact-stimulated grammatical developments in Native American varieties. The eight plenary speakers were the following: **Henning Andersen** (UCLA), **Claire Bowern** (Yale), **Michela Cennamo** (University of Naples Federico II), **Patience Epps** (University of Texas at Austin), **Geoffrey Khan** (Cambridge), **Marianne Mithun** (University of California, Santa Barbara), **Salikoko Mufwene** (University of Chicago), **Rena Torres Cacoullos** (Pennsylvania State University).

In addition, nine workshops were organized on cutting-edge topics in historical linguistics, with 121 participants total:

– The Development of Aspect and Tense Systems *Jadranka Gvozdanivić*
– Paradigm Leveling *David Goldstein & Canaan Breiss*
– Logical Vocabulary and Logical Change *Moreno Mitrović*
– Arabic and Contact-induced Change *Christopher Lucas & Stefano Manfredi*
– The Loss of Inflection *Matthew Baerman, Oliver Bond, Greville Corbett & Helen Sims-Williams*
– Alignment Typology in Diachronic Perspective *Eystein Dahl*
– Atomizing Linguistic Change and the Nuclear Step: From Individual Realization to Emergence *Janet Giltrow, Marianne Mithun & Dieter Stein*
– New Historical Perspectives on Non-dominant Speakers as Agents of Contact-induced Language Change *Petros Karatsareas & Jonathan Kasstan*
– Germanic and Romance: Probing the Similarities and Differences *Sam Wolfe & Christine Meklenborg Salvesen*

We were proud that our series of panels entitled "Las lenguas de San Antonio a 300 años: Reconstructing the Linguistic Roots of a Multicultural City" was designated as an official event of the Tricentennial celebration of San Antonio, recognizing the 300th anniversary of the city's first settlement by Spanish settlers in 1718. The four panels, comprised of experts from around the world, focused on the linguistic heritage of San Antonio and South Texas, and how this diversity is connected to the global community:

– *Spanish Socio-historical Linguistics: Isolation and Contact*
– *African American Vernacular English and the Ecology of Language Evolution*
– *The History of Texas German*
– *Endangered Languages and Historical Linguistics*

Another special event of the conference was the Roundtable on "New Directions for Historical Linguistics: Impact and Synthesis, 50 Years Later." The roundtable was organized to celebrate the republication of *Directions for Historical Linguistics*, on the 50th anniversary of its original publication in 1968, and featured the keynote speakers William Labov and Gillian Sankoff, and invited speakers Elizabeth Traugott, Brian Joseph, Sarah Thomason, and Paul Hopper (via video).

Many thanks to all those who worked so hard to make this conference a success: our Local Organizing Committee, who helped plan every aspect of the conference, and all of the dedicated graduate interns and volunteers who worked tirelessly to solve every problem and to make participants feel welcome. Heartfelt thanks go to Deborah Thomas, to Assistant Conference Organizer Gabriela Lemos and to Technical Coordinator Kat Weigle, as well as to Christopher Espinoza and Brandon Bombardier, who filmed all the plenary sessions. Thanks also to Todd Krause of the Linguistic Research Center, who further edited these films which have now been archived at the LRC (https://liberalarts.utexas.edu/lrc/extras/ichl/index.php).

We would also like to recognize all those who lent their financial support to this effort: the National Science Foundation for a generous grant in support of the Panel on Endangered Languages and Historical Linguistics; Humanities Texas for their support of the History of Texas German Panel; Oak Ridge Associated Universities; Japan Society for the Promotion of Science; Casa de España; the City of San Antonio International Relations Office; the Austin Baroque Orchestra; Staples; and Walgreens. We also wish to thank the following publishing companies for providing much-appreciated financial support: John Benjamins, Cambridge University Press, Oxford University Press, Brill, De Gruyter, and Wiley. In addition, we would like to sincerely thank the University of Texas at San Antonio, the University of Texas at Austin, the University of Texas at Rio Grande Valley, Texas State University, Texas A & M San Antonio and Texas A & M College Station for their generous support of our efforts.

Finally, we extend heartfelt thanks to the members of the Scientific Committee, most of whom served both in the preliminary review process for the conference and in the double blind peer-review for the present volume: Henning Andersen, Jóhanna Barðdal, Brigitte Bauer, Yasmine Beale-Rivaya, Marina Benedetti, Hans Boas, Vit Bubenik, Michela Cennamo, Whitney Chappell, Bert Cornillie, Andrii Danylenko, Ulrike Demske, Thórhallur Eythórsson, Patience Epps, Jan Terje Faarlund, Livio Gaeta, Elly van Gelderen, Jadranka Gvozdanovic, Hans Henrich Hock, Brian Joseph, Ans van Kemenade, Geoffrey Khan, Sonja Lanehart, Juliet Langman, Danny Law, Eugenio Lujan, Na'ama Pat-El, Marc Pierce, Domenica Romagno, Joseph Salmons, John Charles Smith, Sarah Thomason, Elizabeth Traugott, Carola Trips, Nigel Vincent, Christopher Wickham. Special thanks also go to my graduate assistant, Brandon Bombardier, who lent much-appreciated assistance in the final editing process of this collection.

The present volume is a tangible reflection of the outstanding scholarly contributions that were made at the conference. Thanks to all supporters, contributors, and especially participants for demonstrating the immense role that Historical Linguistics should and does play in the complex world of social, political, and intellectual discourse.

Bridget Drinka
San Antonio, Texas
August 17, 2019

Introduction

For more than four decades, since the first International Conference on Historical Linguistics in Edinburgh in 1973, ICHL has continued to serve as the preeminent forum for historical linguistics in the world. Throughout the years, this conference has reflected the growth and diversification of the field, while also recognizing its time-honored foundations, through the introduction of new methodologies and tools, and the incorporation of theoretical constructs and perspectives from various related fields – sociolinguistics, cognitive science, network analysis, phylogenetics, and other disciplines. This burgeoning interdisciplinarity speaks to the vitality of historical linguistics, and to the contributions which it continues to make to the study of language, not only in addressing key issues within the discipline, such as how and why language change occurs, but also in providing explanations for the synchronic outcomes resulting from these changes.

The present volume brings together the work of some of the most influential historical linguists of our time. The papers collected here represent a breadth of topics, from the examination of ancient patterns of argument structure and alignment to the investigation of rates of convergence in the Neo-Aramaic varieties, from the minute analysis of a particle in flux in Old French to a comprehensive examination of the role of indexicality in language change. The volume thus reflects the broad range of interests, methodologies, and theoretical orientations of some of the most productive scholars active in historical linguistics today.

The volume has not been organized according to traditional categories of phonology, morphosyntax, etc., but, rather, according to themes that emerged from the papers themselves. These themes are presented below, with a brief synopsis of each paper's contribution to the scholarly discourse on the topic.

Part I centers around Case & Argument Structure, both in ancient languages and modern varieties, focusing on such topics as actionality, non-canonical subjects, and ditransitive constructions. Domenica Romagno, in her analysis of the accusative of respect in Ancient Greek, points to the fact that this construction tends to promote its most animate argument to subject position, leading to the alignment of syntactic roles and case marking according to the animacy hierarchy. Among her most noteworthy and original observations is the fact that the construction, which entails inalienable possession, is sensitive to the degree of telicity of the event as it interacts with the degree of affectedness of the possessor. Marina Benedetti and

Chiara Gianollo likewise explore the role of case assignment in Ancient Greek, focusing on those syntactic, semantic, and pragmatic features of the dative experiencer argument of *dokéō* 'seem' which suggest that it should be regarded as a non-canonical subject. The authors provide evidence of the telling mismatch of case between this argument and coreferents across clause boundaries, convincingly demonstrating that the dative argument is best construed as a non-canonical subject. Steve Rapaport analyzes parallel case syncretism in early Indo-European, and astutely observes that early IE languages tended to engage in the rare practice of merging oblique cases with other oblique cases – a trend seldom found outside the IE family. He accounts for this unusual outcome as due especially to phonological and prosodic developments, combined with a tendency towards the leveling of irregularity across the early IE languages. Rosa María Ortiz Ciscomani examines the intriguing development of the possessive dative in the ditransitive construction in medieval and modern Spanish, identifying its role as giving relevance to the possessor over the possessed. This possessed entity, the direct object, has in the meantime taken on progressively more abstract and more animate meanings, especially since the 16th century. This section concludes with Andrea Mojedano Batel's incisive analysis of an additional change within Spanish argument structure: the innovative assignment of Spanish *gustar* 'like' from NOMEXP-PPSTIM to DATEXP-NOMSTIM, based on the pattern of other 'liking' verbs of 13th–17th century Spanish. Her extensive corpus analysis demonstrates how both higher type frequency and the subcategorizing properties of syntactic stimuli played a role in producing this outcome.

In Part II, Alignment & Diathesis, Michela Cennamo carries out an impressive study of how new diathetic patterns developed in the transition from Latin to Italo-Romance, especially by means of auxiliaries and light-verb structures ('come', 'do/make', 'become'), reflexive passives, and indefinite/ impersonal SE. She makes the crucial observation that some of these "new" voice distinctions retain patterns of aspect and animacy already familiar in Latin, but that, alongside these persistent semantic categories, the development of the person hierarchy in connection with SE represents an innovation, as does the use of SE itself in its function as voice modulator. Mikel Martínez-Areta, in his analysis of the development of the Basque ergative, claims that this structure arose from the passive voice, and goes on to argue that a passivizing morpheme was not needed for this development. Martínez-Areta makes the provocative but convincing proposal that, because verbal roots of transitive verbs were unspecified as to voice in Proto-Basque, they could be interpreted as active or passive depending on lexical and syntactic context. He concludes that the irrealis mood helped establish the split ergativity existing in Proto-Basque.

In Part III, Patterns, Paradigms, & Restructuring, Henning Andersen's important contribution examines the central role of allomorphy as the link between

synchronic and diachronic domains, with allomorphs serving both as index and symbol: as an index, an allomorph points to its own conditioning features, and, hence, also to the morphological sign it belongs to. Andersen demonstrates that diachronic trends are detectable within the variable allomorphic patterns on the synchronic level, just as diachronic outcomes, such as gradual typological shifts, must reflect reanalyses of layers of synchronic patterns. He suggests that morphological inconsistencies can thus help us differentiate which patterns are productive at the synchronic level, and which pertain to the past but are maintained due to normative convention. Jessica Nowak carries out a well-designed examination of the nature of morphological paradigm extension through partial regularization strategies in ablaut patterns in German and Luxembourgish. Nowak finds that frequency played a significant role: German simplified the pattern of low-frequency verbs from three-vowel to two-vowel systems, while Luxembourgish reduced all strong verbs to a single pattern. Besides frequency considerations, she also finds other factors to be responsible for this outcome: the openness of the ablaut pattern, its similarity across paradigns between target and source, and the cue validity of the ablauting vowel. Livio Gaeta likewise studies paradigmatic developments of strong verbs in German, but focuses not on the preterites but on the past participles in Titsch, a Walser German dialect of the Aosta Valley of Italy. He makes the intriguing observation that the complexity mismatch across the verbal paradigm is resolved in this variety in a novel way, through exaptation or remodeling of the past participles of strong, class I verbs: when they are inflected, they take the *-en* ending; otherwise, they use the *-t* ending. The implications of this exaptive change are noteworthy: BE perfects, for example, take the nasal form, while HAVE perfects take the *-t* form. Johanna Wood's study of patterns found in measure nouns in Middle English completes this section. Wood discovers that English does not follow patterns established in other Germanic languages: it does not develop a Direct Partitive Construction (*a bucket water), but requires the preposition *of* in such structures. Wood provides the especially interesting example of the collective noun *pair*, which, unlike other nouns under consideration, did show some early signs of following the Germanic pattern, if only briefly. The difference in English may be due to the fact that *pair* was acquired via French, while the other Germanic languages borrowed the word directly from Latin.

In Part IV, Grammaticalization & Construction Grammar, Sam Wolfe maps the fascinating development of the particle sɪ from Classical Latin through 13th century French along a grammaticalization cline, tracing its path from an initial stage as an adverbial of manner to a ForceP expletive. What is remarkable about this development, Wolfe shows, is the fact that the reanalysis extends from a low position to the very highest level in the clause, in contrast to the pattern usually seen in such cases in which the extension remains within the VP, TP, or CP.

Significantly, this development is interconnected with many other morphosyntactic changes in French, especially concerning verb movement and the Extended Projection Principle. Grammaticalization is also at the center of the comprehensive analysis presented by Vit Bubenik and Leila Ziamajidi concerning the complex and multi-layered development of the analytic perfect in the West Iranian languages. In Middle Persian, the authors explain, a new system of analytic constructions grew up, using BE + PP for the new perfect. When the BE auxiliary came to be reduced to a suffix, forming a new synthetic preterite, an unambiguous perfect was recreated in Early New Persian by attaching personal suffixes to the perfect stem. Finally, in New Persian, evidential interpretations of the perfect also arose through grammaticalization, possibly also due to Turkic influence. Andrii Danylenko, in his data-rich, admirably meticulous analysis of the grammaticalization of resultatives formed with *-(v)ši-* in North Slavic, argues that the process is internally motivated: he sees the construction as not areally but only genetically related to the *-vęs-* resultatives of Baltic. The grammaticalization of the Slavic construction is driven, Danylenko claims, by sociolinguistic factors such as the social stability and size of the speech community, the types of social networks and shared information, and the amount of adult language contact. Dieter Stein, in his thought-provoking analysis of the nature of language change at the micro-level, argues that we will not understand the means by which change is implemented if we do not attend to the onset context in the singular speech acts of individual speakers. At the core of his proposal is the claim that speakers continually assess the knowledge state of their hearers and gauge the extent to which a deviancy from the remembered history of a particular form can be "risked". Hearers, for their own part, construct the meaning of the slightly-altered form based on their own remembered use history, and on extralinguistic and cognitive meaning assembled on the spot, and may incorporate the innovative usage into their own grammatical repertoire. Stein proposes that speaker innovation may well entail several local directionalities rather than one single directionality, and that the original motivations for the innovative usage may not be those that power its diffusion.

In Part V, Corpus Linguistics & Morphosyntax, John Sundquist carries out a comprehensive corpus analysis of the development of light verb constructions in thirteenth- to sixteenth-century Old Swedish utilizing statistical tools drawn from network science and ecology. Sundquist finds that the light verbs, representing a small, closed class of verbs with special properties, follow the principle of the "rich get richer" as they grow more frequent and diverse. The study points to how dynamic network analysis and biodiversity indexes can be utilized in innovative ways to detect linguistic trends. Using the Icelandic Parsed Historical Corpus (IcePaHC), Hannah Booth carries out a corpus analysis of expletives in the history of Icelandic, and, notably, finds that both *það* and *þar* were used as expletives in overlapping

contexts in early texts. Booth thus disputes claims that expletive *það* arose as a contact phenomenon, and points to the need to include the largely-neglected *þar* as an expletive, as well.

In Part VI, Languages in Contact, Geoffrey Khan explores the role played by contact and other sociolinguistic and systemic factors in the development of the Neo-Aramaic dialects. He discovers that change due to convergence can proceed at different rates across the NENA varieties: while most replica forms show only partial convergence, others replicate the model completely. Khan recognizes a hierarchy of features depending on their relative tendency to converge with contact languages, and notes that constraints on contact-induced change exist, such as socially-motivated resistance to borrowing, large populations and loose networks which may block diffusion, and other factors. He also makes the fascinating observation that contact can add complexity to a linguistic system, even while the goal may have been to reduce complexity. Skillfully pulling together themes which have been addressed elsewhere in this volume, Carola Trips utilizes Johanson's integrative approach to language contact and code copying and other models to carry out a corpus analysis focusing on the copying of Old French psych verbs into Middle English. She finds that English has developed two modes of expressing RECIPIENT arguments, one inherited from Germanic (the ditransitive structure) and one copied from French (*to* – PP). Integration conflicts are seen as diagnostic of contact, pointing to a deeper level of copying, beyond mere borrowing. Andrea Hoa Pham carefully traces the origin of the rare low, back, unrounded vowel /ɑ/ in the Quảng Nam dialect of south central Vietnam to the Hà Tĩnh dialects of the north. She convincingly argues that the introduction of this phoneme in the south was the result of migration that occurred in the 15th to 19th centuries. Pham goes on to claim that the original vowel was later restructured according to other linguistic processes. The present-day outcome is thus due to both contact and internal restructuring. Duaa AbuAmsha carries out an extremely interesting examination of the future tense formations in the Palestinian Arabic (PA) spoken by refugees from Jaffa who are now living in Gaza. She finds that those speakers who remained in Jaffa express the future with a particle *ra:ḥ* (< 'go') or the quasi-verb *bidd-* 'want'. AbuAmsha claims that Jaffan *ra:ḥ* was brought to Gaza and adopted by Gaza speakers, and that this particle then underwent further grammaticalization and loss of phonetic substance to become the prefix *ḥa-*, to be attached to the non-past stem. This future prefix is found in the PA of middle-aged and younger speakers in Gaza, whether Gazan or Jaffan. As AbuAmsha notes, both internal and external factors appear to have been involved in this change. In the final paper in this section, Mary Ann Walter examines the gender assignment of loanwords of neuter gender as they are brought into Bosnian/Croatian/Serbian (BCS). Walter finds that earlier borrowings from Turkish tended to have neuter gender and to enter BCS either unchanged or with feminine

endings, while later borrowings of neuter-like gender from western European languages took on masculine endings. Using optimality theory as an analytical tool, she concludes that an emerging social dispreference for words of Turkish origin may have played a role in the decline in words ending in /e/ overall – a trend which could have facilitated the decline of the neuter class altogether.

In summary, these contributions represent a broad range of theoretical orientations, methodologies, and perspectives. Taken together, they reflect the vibrancy of historical linguistics as a discipline and serve as a fitting representation of the high caliber of papers presented at ICHL23. It has been my pleasure to have served both as organizer of the conference and as editor of this collection of the impressive and significant work of my colleagues.

Bridget Drinka
San Antonio, Texas
August 17, 2019

Case & argument structure

Strategies for aligning syntactic roles and case marking with semantic properties

The case of the accusative of respect in ancient Greek

Domenica Romagno
Università di Pisa

In the present paper, we show that: (1) the accusative of respect, a case marker for inalienable possession in ancient Greek, represents a strategy for aligning syntactic roles and case marking with animacy; (2) this hypothesis can consistently account for a series of issues that remained unsettled; (3) the semantic properties of the predicate involved in the accusative of respect are fundamental to defining its function and specifically govern its distribution. Our analysis comprises Greek literary texts, from Homer to the fifth century B.C.

Keywords: ancient Greek, accusative case, inalienable possession, possessor ascension, syntactic roles, animacy hierarchy, semantic roles, actionality

1.   The unsettled question of the accusative of respect in ancient Greek

In the present paper, we address the function and the distribution of the so-called "accusative of respect" in ancient Greek, from Homer to the fifth century B.C.[1] The accusative of respect typically refers to a construction including an intransitive or passive predicate and a noun in the accusative case, that denotes something with respect to which the scope of the predicate is limited (Schwyzer and Debrunner, 1950), as illustrated in (1)–(4).

(1)   Hom. *Il.* 1.58
       <u>πόδας</u> ὠκὺς Ἀχιλλεύς
       pódas ōkỳs Achilleús
       'swift-footed Achilles' (lit. fast with respect to feet)

---

1.   For a more detailed analysis, see Romagno (2017).

© 2020 John Benjamins Publishing Company

(2)   Xen. *Mem*. 1.6.6

ἢ διὰ τὸ ἀλγεῖν <u>τοὺς</u> <u>πόδας</u> οὐ βαδίζοντα ὅπου ἂν βούλωμαι

ὲ̀ dià tò algeîn toùs pódas ou badízonta hópou àn boúlōmai

'or to be prevented from walking anywhere by sore feet'   (Marchand, 1923)

(3)   Hom. *Il*. 5.284–285

βέβληαι <u>κενεῶνα</u> διαμπερές, οὐδέ σ᾽ ὀΐω

béblēai keneôna diamperés, oudé s᾽ oΐō

δηρὸν ἔτ᾽ ἀνσχήσεσθαι

dēròn ét᾽ anschésesthai

'Thou art smitten clean through the belly, and not for long, methinks, shalt thou endure'   (Murray, 1924)

(4)   Xen. *An*. 2.6.1

οἱ μὲν δὴ στρατηγοὶ οὕτω ληφθέντες ἀνήχθησαν ὡς βασιλέα καὶ

hoi mèn dè stratēgoì hoútō lēphthéntes anéchthēsan hōs basiléa   kaì

ἀποτμηθέντες <u>τὰς</u> <u>κεφαλὰς</u> ἐτελεύτησαν

apotmēthéntes tàs kephalàs eteleútēsan

'The generals, then, after being thus seized, were taken to the King and put to death by being beheaded'.   (Brownson, 1961)

This construction (also known as "accusative of specification" or "accusative of reference") is frequent in ancient Greek and is so particularly a feature of this language that it well deserved its name of *accusativus graecus* or "Greek accusative". However, its function remains unclear: "il est difficile de définir l'accusatif de relation" (Jaquinod, 2006: 42). Among the various functions that the accusative case has in ancient Greek, the role of this specific accusative appears to be quite vague (Meillet and Vendryes, 1927: 505; see also Luraghi, 2003: 52ff.; Crespo, 1988; Marouzeau, 1969; Schwyzer and Debrunner, 1950; Chantraine, 1942; La Roche, 1861). The common theme connecting the various manifestations of the accusative of respect is uncertain and the principle underlying its distribution still has to be defined (Smyth, 1956; Lavidas, 2013). In particular, the question of an accusative depending on an intransitive or passive predicate has long baffled scholars; nonetheless, it remains unsettled: "it is […] a matter which needs some explanation when we find an accusative depending on a passive verb" (Courtney, 2004: 425).

2.   A case marker for inalienable possession

The accusative of respect can indicate the part of a whole, a quality or attribute (such as form, size, name, etc.), either a concrete or abstract property, a place, the site or the extent of something, and also result in quasi-adverbial expressions of duration and measure (La Roche, 1861; Schwyzer and Debrunner, 1950: 67ff.).

Previous studies almost exclusively focused on the referent of the nouns that take the accusative of respect (La Roche, 1861; Jacquinod, 1989, 2006; among others). These nouns denote:

–   body parts (or the whole body): e.g., χεῖρας, ὄμματα, χρῶμα cheîras, hómmata, chrôma
–   the heart, the soul, or other parts of the human being, related to spiritual, emotional or intellectual dimensions: e.g., φρένα, κῆρ phréna, kêr
–   qualities or distinctive properties: e.g., εἶδος, μέγεθος, κάλλος eîdos, mégethos, kállos
–   typically human activities or the ability to perform those activities: e.g., μάχην, πολέμια, ἀρετάς, νόον máchēn, polémia, aretás, nóon
–   notions such as lineage and family: e.g., γένος génos
–   'name' (ὄνομα ónoma) and 'denomination, title' (ἐπίκλησιν epíklēsin)
–   'feats': ἔργα érga

The occurrences of the accusative of respect in Homer are variously distributed among these classes. The same classes occur in the accusative of respect constructions in later texts, with slight differences involving single instances of each class and the proportions between classes (Romagno, 2017; Jacquinod, 2006: 63ff.; La Roche, 1861).

The typology of the entities denoted by the nouns that take the accusative of respect can be interpreted as belonging to the category of inalienable possession: the hands, for instance, cannot be logically alienated from their possessor, even if they are severed from the whole body: 'hands' implies 'someone's hands', like 'mother' and 'father' necessarily imply a relation with another entity (i.e., a child: 'someone's mother/father'); vice versa, a bicycle and a pen do exist independently of their possessor and are involved in a relationship of alienable possession (Lévy-Bruhl, 1916; Fillmore, 1968; Nichols, 1988; Nichols and Bickel, 2005; Heine, 1997; Haspelmath, 1999; Aikhenvald and Dixon, 2013).

While alienability is an open-class category (Nichols, 1988: 562 describes its membership as "infinite"), inalienability involves a closed set of nouns, that – as shown by crosslinguistic investigations – largely correspond to the categories involved in the accusative of respect construction.

The following passage by Xenophon (5) reveals the strict relationship between the accusative of respect and the notion of inalienable possession (cf. Frei, 1939: 188; Jacquinod, 1989: 43; 2006: 64ff.; 2016: 7):

(5)   Xen. *Mem*. 2.1.22

κεκοσκημέμην <u>τὸ</u> μὲν <u>χρῶμα</u>   καθαρειότητι, <u>τὰ</u> δὲ <u>ὄμματα</u> αἰδοῖ,
kekoskēmémēn tò mèn chrȏma   kathareiótēti, tà dè hómmata aidoî,
<u>τὸ</u> δὲ <u>σχῆμα</u>   σωφροσύνῃ, **ἐσθῆτι   δὲ λευκῇ**
tò dè schȇma   sōphrosýnēi, hesthȇti dè leukȇi
'and her limbs were adorned with purity, her eyes with modesty; sober was her
figure, and her robe was white'   (Marchant, 1923)

The symmetry between accusatives is broken by the clothing term, that takes
oblique case marking (dative): the accusative of respect is excluded from clothing
terms, as they do not fall under inalienable possession.

    Also ἔργα érga 'feats', which occurs three times in Homer, and its continuation
in the post-Homeric literature (e.g., πολέμια, μάχην, μάχας, ἰατρικήν, μουσικήν,
polémia, máchēn, máchas, iatrikḗn, mousikḗn, well-attested in Aeschylus,
Sophocles, Aristophanes, Thucydides, Pindar) can be attributed to the category of
inalienable possession (cf. Romagno, 2017; Jacquinod, 2006: 65ff.; 2016: 21). ἔργα
érga always refers to a distinctive property of the possessor, as shown by the fol-
lowing passage (6), in which it expresses a typical feature of Athena and perfectly
parallels κάλλος kállos 'beauty', that denotes a characteristic of Aphrodite:

(6)   Hom. *Il*. 9.389–390

οὐδ᾽ εἰ χρυσείη Ἀφροδίτη <u>κάλλος</u> ἐρίζοι,
oud᾽ ei chryseíēi Aphrodítēi kállos   erízoi,
<u>ἔργα</u> δ᾽ Ἀθηναίη γλαυκώπιδι ἰσοφαρίζοι
érga d᾽ Athēnaíēi glaukṓpidi isopharízoi
'not though she vied in beauty with golden Aphrodite and in handiwork were
the peer of flashing-eyed Athene'   (Murray, 1924)

πολέμια, μάχην, μάχας, ἰατρικήν, μουσικήν, polémia, máchēn, máchas, iatrikḗn,
mousikḗn, etc. refer to typically human activities or to the ability to perform those
activities.

2.1   Accusative of respect and inalienable possession: Key questions

As regards the relationship between accusative of respect and inalienable posses-
sion, two crucial issues require special attention: these have not been addressed in
previous studies and, nonetheless, have important implications for clarifying the
function and the distribution of the accusative of respect, as we will show below.
Specifically:

1. the subject of the construction, both in Homeric and later texts, refers to a highly animate entity, typically human;
2. kin terms, which represent one of the prototypical classes involved in the inalienable possession relationship, do not occur in the accusative of respect. The reason for this has not yet been provided. In the following sections, we will propose an explanation for this noteworthy exception.

3.   The function and distribution of the accusative of respect: Research questions

We have shown so far that the accusative of respect construction encodes the relationship of inalienable possession. However, crucial questions still remain open: (1) what is the principle underlying the selection of this specific type of construction to express the inalienable possession relationship in ancient Greek? (2) what is the role of the predicate in defining the function of the accusative of respect and how do the semantic properties of the predicate affect the distribution of this construction? (3) which dimensions are critically involved in the relationship between the predicate and the two main arguments of the construction? We will address these unsettled questions in the following sections.

4.   A strategy for aligning syntactic roles and case marking with animacy

To provide an answer to the first question, we propose that the accusative of respect represents a strategy to promote the most animate argument of the relationship (i.e., the possessor) to the subject position and, consequently, to align syntactic roles and case marking with animacy hierarchy. The subject of the construction, in fact – as observed in § 2.1 – both in Homeric and later texts, refers to a highly animate entity, almost exclusively a human being:[2] a possessor with high ranking in the animacy hierarchy and, more precisely, a human possessor represents one of the prototypical features of the inalienable possession (Heine, 1997; Aikhenvald and Dixon, 2013).[3]

---

2.   On the animacy hierarchy, see the seminal work by SILVERSTEIN (1976). See also: Comrie (1978), Dahl and Fraurud (1996), de Swart *et al.* (2008), among others.

3.   On the "grammar of possession" (Fillmore, 1968) and, specifically, of inalienable possession constructions, and on the typology of inalienable possession expressions across languages, see Fillmore (1968), Chappel and McGregor (1996), Heine (1997), Haspelmath (1999), Aikhenvald and Dixon (2013). On the notion of meronymy and its relationship with inalieble possession,

4.1   Accusative of respect and double accusative

In order to further explain and support our thesis, in this and the following sections we will address a series of unsettled issues by showing how the idea proposed here can consistently account for all of them. We start by focusing on a piece of evidence that has not yet received enough attention from a diachronic perspective. Many of the constructions defining the typology of the accusative of respect and, in particular, those expressing a part–whole relationship are well-attested, in Homer, in the transitive. This is a subset of the so-called "double accusative" constructions, a subset that belongs to the "whole and part schema" (σχῆμα καθ᾽ ὅλον καὶ μέρος schêma kath᾽ hólon kaì méros: cf. Delbrück, 1893; Brugmann, 1910; Schwyzer and Debrunner, 1950: 84 ff.; Hahn, 1954; Jacquinod, 1989):[4] the same predicate – mostly corresponding to the "hitting and breaking" type – occurs in the intransitive version of the accusative of respect construction and in the transitive version with double accusative. Typically, the transitive version shows an active pattern, in which both the possessor and the possessum bear accusative case (7); whereas, in the intransitive, the verb is formally a middle or an aorist in -η-/-θη- -ē-/-thē-, with passive value (cf. Romagno, 2010, 2014), the possessor is in the nominative case and the possessum is in the accusative (8):

(7)   Hom. *Il.* 501–502 = transitive construction with double accusative
<u>τόν</u> ῥ᾽ Ὀδυσεὺς ἑτάροιο χολωσάμενος **βάλε** δουρὶ
tón rh᾽ Odyseùs hetároio cholōsámenos bále dourì
<u>κόρσην</u>
kórsēn
'Him Odysseus, wroth for his comrade's sake, smote with his spear on the temple'   (Murray, 1924)

(8)   Hom. *Il.* 4.518–519 = intransitive construction with accusative of respect
χερμαδίῳ   γὰρ **βλῆτο** παρὰ σφυρὸν ὀκριόεντι
chermadíōi gàr blêto parà sphyròn okrióenti
<u>κνήμην</u> <u>δεξιτερήν</u>
knḗmēn dexiterén
'for with a jagged stone was he smitten on the right leg by the ankle'   (Murray, 1924)

---

see Rosen (1976) and Mirto & Rosen (1994). Here we propose that the accusative of respect, as opposed to other constructions encoding the inalienable possession relationship, represents a strategy to assign the subject role to the noun which is highest in animacy.

4.   These constructions can also be classified as ditransitive with neutral alignment, as the possessor and the possessum have the same marker: see Malchukov *et al.* (2010).

Remarkably, the intransitive construction tends to totally replace the transitive, in post-Homeric literature (Romagno, 2017; Lavidas, 2013: 5; Schwyzer and Debrunner, 1950: 84–85; see also Jacquinod, 1989): the double accusative construction expressing the part–whole relationship, which is frequent in Homer, becomes extremely rare in the subsequent periods of the language, in which the accusative of respect is almost exclusively attested in intransitive constructions.

The idea that the accusative of respect is a strategy for aligning syntactic roles and case marking with animacy hierarchy can account for the increasing frequency of the accusative of respect in intransitive constructions, as opposed to its almost total disappearance in the transitive constructions with double accusative, in post-Homeric texts.

The accusative of respect intransitive construction represents a case of possessor ascension,[5] in which the possessor, the most animate entity in the construction – unlike the double accusative, in which it is, indeed, in the accusative – is promoted to the subject role and marked with nominative case; while the possessum, less animate, remains in the accusative: the result of this strategy is an alignment of syntactic roles (specifically, subject and object) and case marking (specifically, nominative and accusative) with animacy hierarchy.

The fact that the language progressively abandons the double accusative construction to express the part–whole relation, whereas the accusative of respect, in its intransitive version, increases its frequency testifies to the tendency to assign the most animate argument to the subject role (the role prototypically taken by animate nouns) and to encode it into the nominative case (the prototypical case of animate subjects).[6] In the double accusative construction of the type explained above, in

---

5. It is worth remarking that possessor ascension is typical – even if not exclusive – of inalienable possession (cf. Heine, 1997: 163 ff.; Chappell and McGregor, 1996). The same phenomenon has also been called "possessor raising" or "possessor promotion": cf. Perlmutter and Postal, 1983; Heine, 1997: 163ff.; Haspelmath, 1999: 9ff.; Blake, 2002; Deal, 2013. On the terms "possession specification" and "external possessor construction", both proposed to refer to inalienable possession expressions, as also related to possessor ascension, see Heine (1997) and König and Haspelmath (1995), respectively. For a discussion of inalienable possession and possessor ascension, including terminological issues, see Heine (1997: 163 ff.; cf. also Blake, 2002: 100 ff. and Chappell and McGregor, 1996). On the notion of external possession, from different perspectives (from typological to psycholinguistic, from formal to cognitive-functional approaches), see the various contributions in Barshi and Payne (1999). For a discussion on limitations on possessor ascension depending on the verb class, cf. Levin (1993), Tenny (1994: 213 ff.), Heine (1997: 163 ff. and, in particular, 168 ff.). For an attempt to address the question of body-part possessor ascension from a neuropsychological perspective, see Kemmerer (2003).

6. On the relationship between nominative/accusative and ± animate arguments, in different languages, see Winter (1971), Lazzeroni (2002a, 2002b).

fact, the possessor takes the accusative, which is atypical for animate arguments (cf. Bossong, 1998; Aikhenvald, Dixon and Onishi, 2001; Romagno, 2005, 2006, 2007; among others). In the accusative of respect intransitive construction, the possessor instead takes the nominative, which is the protypical case of animate arguments (cf. Carruba, 1992; Van Valin and LaPolla, 1997; Lazzeroni, 2002a, 2002b; de Swart et al. 2008): then, the encoding of the part–whole relationship and, on the whole, of the inalienable possession relationship selected a strategy that shows a "canonical" alignment of syntactic roles and case marking with animacy.

4.2   The noteworthy exception of kin terms

In § 2.1, we mentioned a noteworthy exception that has not yet been explained: kin terms, such as mother and father, which represent a prototypical category of inalienable possession (Fillmore, 1968; Heine, 1997; Aikhenvald and Dixon, 2013), are excluded from the accusative of respect, which, nonetheless – as we have shown above – is a typical construction to express the inalienable possession relationship in ancient Greek. We believe that the reason can be identified in the high animacy of kinship nouns. Their incompatibility with the accusative of respect confirms the idea that this construction represents a strategy to assign the subject role to the entity that is highest in animacy and, consequently, to align case markers and syntactic roles with the animacy hierarchy. In the accusative of respect, in fact, the possessor, more animate, becomes the subject of the construction and takes the nominative case; the possessum, less animate, remains in the accusative, as illustrated in (9) and (10):

(9)   Hom. *Il.* 9.9
Ἀτρεΐδης δ᾽ ἄχεϊ   μεγάλῳ βεβολημένος ἦτορ
Atreḯdēs d᾽ ácheï megálōi bebolēménos êtor
'But the son of Atreus, stricken to the heart with sore grief'   (Murray, 1924)

(10)   Hom. *Il.* 4.272
Ἀτρεΐδης δὲ παρῴχετο γηθόσυνος κῆρ
Atreḯdēs dè paróicheto gēthósynos kêr
'and the son of Atreus passed on, glad at heart'   (Murray, 1924)

The high animacy of kin terms is inconsistent with the prototypical possessum status and, therefore, conflicts with the animacy relationship between the possessor and the possessum established by the accusative of respect.

4.3 The strange case of an accusative depending on a passive and the lack of double nominative in Homer

The possessum is lower in animacy, but totally affected by the event that the predicate denotes; the possessor, instead, is only partially affected by the event but higher in animacy. In expressions such as τόν ῥ' Ὀδυσεὺς βάλε κόρσην tón rh' Odyseùs bále kórsēn (see Example (7), in § 4.1) 'Odysseus hit him on the temple', the possessor (τόν tón = him) is only partially affected, while the temple (κόρσην kórsēn) undergoes the effect of the event entirely.

The possessum corresponds to the entity which is highest in affectedness and, therefore, could be a perfect candidate for the subject role in passive constructions (Romagno, 2006; Fici Giusti, 1994: 27ff., 43ff.). Nonetheless, in these constructions, the nominative subject is assigned to the possessor, because it is the most animate argument in the construction, as in (11):

(11) Hom. *Il.* 5.284
      βέβληαι <u>κενεῶνα</u>
      béblēai keneôna
      'You got hit in the flank (κενεῶνα keneôna = accusative of respect)'.
      <div align="right">(cf. Example (3))</div>

The apparently insurmountable difficulty by which an affected entity is marked with the accusative case in a passive construction can be overcome. It has been affirmed that "the fact that the accusative with an active verb should remain accusative with a passive verb is the crux of the whole matter" (Hahn, 1954: 241). What the explanation proposed here can account for is the fact that the less animate argument (the possessum), which takes the accusative case in the active construction, remains in the accusative in the passive, while the possessor (higher in animacy), takes the nominative.

Analogously, the idea that the accusative of respect represents a strategy to promote the most animate argument of the construction to the subject position can also solve the aporia – referred to by Hahn (1954) as irreducible – whereby there is no double nominative construction paralleling the double accusative of the part–whole schema: "we have absolutely no trace in Homer of a nominative body-part noun used originally as one of two subjects […] of a passive verb […] no such instances exist" (Hahn, 1954: 282). Body parts (the possessum) are less animate than the possessor (a human being): only the most animate argument is promoted to

nominative subject, in order to match the animacy relationship between possessor and possessum with syntactic roles and case marking.[7]

5.   A case marker for undergoer arguments

To answer the other unsolved questions addressed in § 3, we investigated the role of the predicate and, in particular, of its semantic properties, in defining the function and the distribution of the accusative of respect (cf. Romagno, 2017; Bossaglia, 2005). In fact, although it has been defined as "de tous les types d'accusatifs, le seul qui complète une forme nominale" (Jacquinod, 2016: 21), the accusative of respect clearly refers to the whole event, as it involves the relationship between the predicate and its argument(s). We seek to understand how this relationship affects the distribution of the accusative of respect and accounts for its function. To this purpose, we analyzed the event types[8] involved in the constructions including an accusative of respect in Ancient Greek, from Homer to the fifth century B.C., and found that the accusative of respect occurs in combination with the following categories (cf. Romagno, 2017):

I.   Adjectives; this construction is extremely frequent in Homeric and post-Homeric texts and occurs also in inscriptions (cf. (12) and (13)), as well as previous Examples (1) and (10)):

(12)   Hdt. 3.4.1
ἦν   τῶν ἐπικούρων Ἀμάσιος ἀνὴρ <u>γένος</u> μὲν **Ἁλικαρνησσεύς**,
ên   tôn   epikoúrōn Amásios anèr génos mèn Halikarnēsseús,
οὔνομα δέ οἱ Φάνης, καὶ <u>γνώμην</u> **ἱκανὸς** καὶ <u>τὰ πολεμικὰ</u> **ἄλκιμος**.
oúnoma dé hoi Phánēs, kaì gnṓmēn hikanòs kaì tà polemikà álkimos.
'There was among Amasis' mercenaries a man who was a Halicarnassian by birth, a clever man and a good soldier, whose name was Phanes'

(Godley, 1920)

(13)   IG 14.134
ἀμέπτως <u>τὸ(ν)</u> <u>βίον</u>
améptōs tò(n) bíon
'irreproachable in his life'.

II. Predicates[9] expressing a state, property or condition (cf. (14)–(16), as well as Example (2)):[10]

(14) Hom. *Il.* 6.481
… **χαρείη** δὲ <u>φρένα</u> μήτηρ
… chareíē dè phréna métēr
'… and may his mother's heart wax glad"
(Murray, 1924; more lit. "and may the mother feel joy at heart')

(15) Hom. *Od.* 208–209
αἰνῶς μὲν <u>κεφαλήν</u> τε καὶ <u>ὄμματα</u> <u>καλὰ</u> **ἔοικας**
ainôs mèn kephalén te kaì ómmata kalà éoikas keínōi
κείνῳ
keínōi
'moltissimo per la testa e i begli occhi somigli a lui', (Di Benedetto, 2010)
'you resemble him very much in your head and your beautiful eyes'

(16) Pl. *Resp*. 453b
ἔστιν οὖν ὅπως οὐ πάμπολυ **διαφέρει** γυνὴ ἀνδρὸς <u>τὴν φύσιν</u>;
éstin oûn hópōs ou pámpoly diaphérei gynè andròs tèn phýsin?
'Can it be denied then that there is by nature a great difference between men and women?'.
(Shorey, 1969)

---

9. We use the term predicate as related to functional dimensions, independently of formal distinctions. This predicate category, for instance, comprises both active and middle forms, perfect forms, aorist in -η-, both finite and nonfinite verb forms, etc. Analogously, passive predicates include middle forms, aorist in -η-/-θη-, finite verb forms, participles, etc. What is relevant here is the logical structure of the predicate and the semantic representation of the event (cf. Vendler, 1967; Dowty, 1979; Van Valin and LaPolla, 1997, among others).

10. Significantly, with verbs such as στάζω stázō 'to drip', the predicate involved in the accusative of respect intransitive construction encodes the anticausative member of a causative/anticausative alternation, that expresses a state or condition of the subject (or, in different cases, a change of state of the subject: cf. the following predicate category, in the main text): Soph. *Aj*. 9–10 ἔνδον γὰρ ἀνὴρ ἄρτι τυγχάνει, <u>κάρα</u> / **στάζων** ἱδρῶτι καὶ <u>χέρας</u> ξιφοκτόνους éndon gàr hanèr árti tynchánei, kára / stázōn hidrôti kaì chéras xiphoktónous 'for the man has just now gone in, dripping with sweat from his head and from his hands that have killed with the sword' (Jebb, 1893). Cf. also Eur. *Supp.*, 586–587: the text is corrupted, but if the conjecture στόμα stóma is accepted (Oates and O'Neill, 1938), this represents one of the extremely rare cases of non-human possessor, as the noun in the accusative of respect refers to a horse's mouth (μοναμπύκων τε φάλαρα κινεῖσθαι στόμα / ἀφρῷ καταστάζοντα Καδμείαν χθόνα. monampýkōn te phálara kineîsthai stóma / aphrôi katastázonta Kadmeían chthóna.). στάζω stázō 'to drip ('faire tomber goutte à goutte' or 'tomber goutte à goutte': Chantraine, DELG) is a case of "labile non-direct opposition", in Haspelmath's terms (Haspelmath, 1993): the same verb, with no extra marker, encodes either the causative member of the causative/anticausative alternation, that denotes an externally caused event (e.g., *the sun melts the snow*), or the anticausative member (also called "inchoative": Haspelmath, 1993), that denotes the event with no external cause (e.g., *the snow melts*; see also Levin and Rappaport Hovav, 1995; Kulikov, 1998).

III. Change of state predicates (either achievements or accomplishments: cf. Vendler, 1967; Bertinetto, 1986) (cf. (17)–(18)):

(17) Hom. *Il.* 5.354
[…] **μελαίνετο** δὲ χρόα καλόν
[…] melaíneto dè chróa kalón
'[her] beautiful skin became dark'' (lit. '[she] became dark at the beautiful skin')

(18) Andoc. 1.61
τὴν κλεῖν **συνετρίβην** καὶ τὴν κεφαλὴν **κατεάγην**
tḕn kleîn synetríbēn kaì tḕn kephalḕn kateágēn
'I broke my collar-bone and fractured my skull'.   (Meidment, 1968)

IV. Passive predicates (cf. (19) and (20), as well as Examples (3), (4), (8), (9)):[11]

(19) Xen. *An.* 4.5.12
οἵ τε **διεφθαρμένοι** ὑπὸ τῆς χιόνος τοὺς ὀφθαλμοὺς
hoí te diephtharménoi hypò tês   chiónos toùs ophthalmoùs
'those whose eyes had been blinded by the snow'   (Brownson, 1961)

(20) Eur. *Med.* 8
ἔρωτι θυμὸν   **ἐκπλαγεῖσ**᾽ Ἰάσονος
érōti thymòn ekplageîs̱᾽   Iásonos
'her heart smitten with love for Jason'.   (Kovacs, 1994)

We are able to identify a common principle underlying the distribution of the accusative of respect among these different event types: they all imply the representation of a state, that is, they have a state predicate in their logical structure (in Dowty's terms: Dowty, 1979; see also Vendler, 1967; Van Valin and LaPolla, 1997; Van Valin, 2007). The state may be represented as either inherent (types I, II and, in more marked contexts, IV) or acquired as a consequence of a change of state (types II and IV), and the representation may either include or exclude an external cause (cf. Romagno, 2017: 77–80). The argument of a state predicate refers to an entity that is in a given state/condition or undergoes a change of state/condition: therefore, it has the semantic role of an undergoer, prototypically, of a patient or theme.[12]

---

11. This category mostly comprises change of state verbs: e.g., ἐκκόπτω ekkóptō "to cut"; διαφθείρω diaphtheírō "to damage, to ruin"; πλήσσω pléssō "to hit, to beat, to punch", also in combination with verbal prefixes: ἐκπλήσσω, καταπλήσσω ekpléssō, katapléssō "to knock out, to shock, to scare"; δαΐζω daízō "to break"; πίμπλημι pímplēmi "to fill up", etc.

12. Specific instances of semantic roles can be subsumed under two main macroroles: actor and undergoer. The prototipical actor is an agent and the prototypical undergoer is a patient (or a theme). On the notion of macrorole, see, in particular, the contribution of the Role and Reference Grammar: Foley and Van Valin (1984), Van Valin (1990), Van Valin and LaPolla (1997). Semantic

It is, then, clear that the principle underlying the distribution of the accusative of respect is strictly related to the functional-semantic dimensions involved in this construction: the accusative of respect requires an undergoer argument, as it implies the representation of a state or a change of state. It does not occur, in fact, in combination with verbs that do not entail a state or a change of state, such as 'to walk' or 'to hear', in expressions like *'to walk with respect to legs/feet' or *'to hear with respect to ears'. In these cases, the possessum shows oblique case marking, as in (21) and (22):

(21) Soph. *OT* 1385
  ὀρθοῖς ἔμελλον ὄμμασιν τούτους ὁρᾶν;
  orthoîs émellon ómmasin toútous horân?
  'was I to look with steady eyes on this folk?'   (Jebb, 1887)

(22) Aesch. *Supp.* 210
  ἴδοιτο δῆτα πρευμενοῦς ἀπ' ὄμματος
  'May he indeed behold you, and with a gracious eye'.   (Smyth, 1926)

On the other hand, in cases like 'to be glad at heart' or 'to hit/to be hit on the head' – as shown in the Examples (1)–(20) – the possessum takes the accusative of respect, instead of an oblique case (unlike English or Italian, for instance, in which, in this kind of expression, the possessum is encoded into a locative prepositional phrase: e.g., 'on the head', 'in the flank', 'at heart'). This establishes a correspondence between argument coding (accusative case) and semantic role (undergoer): the accusative, in fact, is prototypical for the semantic role of undergoer, as it typically encodes the more affected argument. The accusative of respect, then, appears to function as a prototypical accusative case, as it applies only to undergoer arguments.

The reason for having an undergoer argument cannot be reduced to the "object status" that the possessor and the possessum have in the Homeric double accusative construction of the whole-part schema: this condition, in fact, does not apply only to the predicate types that occur in that construction and, in parallel, in the corresponding intransitive/passive version (the "hitting and breaking" predicates, such as

---

roles are also called thematic roles or thematic relations, in different frameworks, depending on whether the focus is on the semantic role of the arguments or on the grammatical relation that they take to the verb. On the *continuum* of semantic roles and on terminological issues, from different perspectives, see Fillmore (1968), Jackendoff (1976, 1987), Chomsky (1981), Dowty (1991), Van Valin and LaPolla (1997), among others. We are thankful to an anonymous reviewer for pointing out that the functional-semantic perspective adopted in the present paper perfectly matches the syntactic perspective that considers the subject of the intransitive accusative of respect constructions examined above a promoted object: this could also constitute the syntactic link with the double accusative construction.

βάλλω bállō 'to hit', πλήσσω pléssō 'to hit, to beat, to punch', (ἀπο)τέμνω (apo)témnō 'to cut', etc.), but – as shown above – follows a more general principle that governs the distribution of the accusative of respect among different event types, including not only changes of state, but also adjectives and (one-argument) state verbs.[13]

5.1   A restriction on telic predicates

Finally, we will address a further issue, which has not yet been noticed. We have shown in § 5 that the accusative of respect encodes an affected entity.[14] Therefore, the notion of change of state is one of the critical dimensions involved in it, as the argument that undergoes a change of state corresponds to the most affected entity. Consequently, the notion of telicity plays a role in the accusative of respect construction, as the representation of a change of state is proper to telic verbs, such as 'to die', 'to murder', 'to break', 'to cut', as opposed to atelic verbs, such as 'to walk' or 'to see', that do not have any delimitation or final state (Vendler, 1967; Bertinetto, 1986).

The affected entity that takes the accusative of respect corresponds to the possessum of an inalienable possession relationship: therefore, the possessor, typically human, is also affected by the event that the verb denotes. This event, though, does not entail a total and definite change of state of the whole possessor, as the scope of the predicate is defined by the term that takes the accusative of respect. The accusative of respect, then, requires a further restriction on the selection of the event type, a restriction that depends on the degree of telicity of the event as related to the degree of affectedness of its participant(s): inherently telic verbs that do not allow a limitation on the affectedness of their animate, typically human, undergoer argument (e.g., 'to kill', 'to murder', 'to die') are not compatible with the accusative of respect.[15]

---

13.   The argument discussed here may also contribute to clarifying the relationship between the accusative of respect and the body part-accusative in partitive apposition (specifically, in the double accusative construction): it has been proposed that the former developed from the latter (Brugmann, 1910; Hahn, 1954; for a first hint of the same view, see Delbrück, 1893); however, we have shown that the body part-accusative is a subset of a more complicated scenario, in which different instances of the accusative of respect can be accounted for in a consistent explanation, that also captures diachronic patterns.

14.   On the notion of affectedness in relation to the "external possessor construction", in a typological perspective, see König and Haspelmath (1997), Haspelmath (1999).

15.   Significantly, there is typological evidence on limitations on the body part-possessor ascension, related to the degree of telicity and affectedness (e.g., Levin, 1993, on English verb classes).

This limitation refers to the possibility that the event affects the possessor only partially (e.g., in a body part, but not in the whole body), like in cases such as 'to break', 'to cover', 'to hit' (as in (23) and (24)):

(23) = (11) Hom. *Il*. 284
βέβληαι κενεῶνα
béblēai keneôna
'you got hit in the flank'   (cf. (3), in § 1)

(24) = (18) Andoc. 1.61
τὴν κλεῖν **συνετρίβην** καὶ τὴν κεφαλὴν **κατεάγην**
tền kleîn synetríbēn kaì tền kephalền kateágēn
'I broke my collar-bone and fractured my skull',   (Meidment, 1968)

and/or involves a less definite change of state and, consequently, a lower degree of telicity. The notions of murdering or dying, for instance, entail a total and definite change of state; whereas the notion of improving or those of becoming dark and spoiling/rotting can even be encoded into a gradual completion predicate, and, then, refer to a gradual approach to the telos (e.g., Italian *la situazione è migliorata di parecchio*, English *the situation has improved by a lot* vs. Italian **il nemico è morto di parecchio*, English **the enemy has died by a lot*: cf. Bertinetto and Squartini, 1995; Beavers, 2013), cf. (25) and (26):

(25) = (17) Hom. *Il*. 5.354
μελαίνετο δὲ χρόα καλόν
melaíneto dè chróa kalón
'[her] beautiful skin became dark' (lit. '[she] became dark at the beautiful skin')

(26) Xen. *An*. 4.5.12
οἵ τε ὑπὸ τοῦ ψύχους τοὺς δακτύλους τῶν ποδῶν **ἀποσεσηπότες**
hoí te hypò toû psýchous toùs daktýlous tồn podồn aposesēpótes
'or those whose toes had rotted off by reason of the cold'.   (Brownson, 1961)

6. Conclusions

In the present paper, we have shown that the accusative of respect, which involves an inalienable possession relationship, represents a strategy to promote the most animate argument of the construction (i.e., the possessor) to the subject position and, consequently, to align syntactic roles and case marking with the animacy hierarchy. This idea can also:

1. account for the increasing frequency of the accusative of respect in intransitive constructions, as opposed to its almost total disappearance in the transitive constructions with double accusative, in post-Homeric texts;

2. explain the noteworthy and previously unsettled exception by which kin terms, which represent a prototypical category of inalienable possession, are excluded from the accusative of respect;

3. overcome the apparently insurmountable difficulty by which an affected entity is marked with the accusative case in a passive construction;

4. solve the apparently irreducible aporia whereby there is no double nominative construction paralleling the Homeric double accusative construction of the part–whole schema.

In addition, we have shown that:

I.   the semantic properties of the predicates involved in the accusative of respect are fundamental to defining its function and specifically govern its distribution;

II. a common principle underlies the combination of the accusative of respect with different predicate types: this principle relies on specific semantic dimensions, involving argument roles and actionality;

III. there is a restriction on the selection of the telic verbs involved in the construction, that depends on the relationship between degree of telicity of the event and degree of affectedness of the possessor.

References

Aikhenvald, A. Y., Dixon, R. M. W. & Onishi, M. (Eds.). (2001). *Non-canonical marking of subjects and objects*. Amsterdam/Philadelphia: Benjamins. https://doi.org/10.1075/tsl.46

Aikhenvald, A. & Dixon, R. (Eds.). (2013). *Possession and ownership*. Oxford: Oxford University Press.

Barshi, I. & Payne, D. (Eds.). (1999). *External possession*, Amsterdam/Philadelphia Benjamins.

Beavers, J. (2013). Aspectual classes and scales of change. *Linguistics*, 51, 681–706. https://doi.org/10.1515/ling-2013-0024

Bertinetto, P. M. (1986). *Tempo, aspetto e azione nel verbo italiano*. Firenze: Accademia della Crusca.

Bertinetto, P. M. & Squartini, M. (1995). An attempt at defining the class of gradual completion verbs. In P. M. Bertinetto, V. Bianchi, J. Higginbotham, & M. Squartini (Eds.), *Temporal Reference, Aspect and Actionality, vol. I: Semantic and Syntactic Perspectives*. (pp. 11–26). Torino: Rosenberg and Sellier.

Blake, B. (2002). *Relational Grammar*. London/New York: Routledge. https://doi.org/10.4324/9780203221761

Bossaglia, G. (2005). L''accusativo greco' in Omero e nella letteratura successiva. BA Thesis, University of Pisa.

Bossong, G. (1998). La marquage différentiel de l'objet dans les langues d'Europe. In J. Feuillet (Ed.). *Actance et Valence dans les langues de l'Europe* (pp. 193–258). Berlin/New York: Mouton de Gruyter.

Brugmann, K. (1910). Der sogenannte Akkusativ der Beziehung im Arischen, Griechischen, Lateinischen, Germanischen, *Indogermanische Forschungen*. 27: 121–151. https://doi.org/10.1515/9783110242676.121

Carruba, O. (1992). Le notazioni dell'agente animato nelle lingue anatoliche (e l'ergativo). In O. Carruba. *Per una Grammatica Ittita / Towards a Hittite Grammar*. (pp. 61–98). Pavia: Iuculano.

Chantraine, P. (1942). *Grammaire Homérique, I*. Paris: Klincksieck.

Chantraine, P. DELG, *Dictionnaire étymologique de la langue grecque*. (repr. 2009). Paris: Klincksieck.

Chappel, H. & McGregor, W. (Eds.). (1996). *The grammar of inalienability. A typological perspective of body part terms and the part–whole relation*. Berlin/New York: de Gruyter. https://doi.org/10.1515/9783110822137

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris Publications.

Comrie, B. (1978). Ergativity. In W. P. Lehmann (Ed.). *Syntactic Typology: Studies in the Phenomenology of Language*. pp. 329–394. Austin: University of Texas Press.

Courtney, E. (2004). The 'Greek' accusative. *The Classical Journal* 99, 4, 425–431.

Crespo, E. (1988). The semantic and syntactic functions of the accusative. In A. Rijksbaron et al. (Eds.). *In the Footsteps of Raphael Kühner* (pp. 99–121). Amsterdam: Gieben.

Dahl, Ö. & Fraurud, K. (1996). Animacy in grammar and discourse. In T. Fretheim and J. K. Gundel (Eds.). *Reference and referent accessibility* (pp. 47–65). Amsterdam/Philadelphia: Benjamins. https://doi.org/10.1075/pbns.38.04dah

de Swart, P., Lamers, M. & Lestrade, S. (2008). Animacy, argument structure and argument encoding. *Lingua* 118, 131–140. https://doi.org/10.1016/j.lingua.2007.02.009

Deal, A. R. (2013). Possessor raising. *Linguistic Inquiry* 44, 3, 391–432. https://doi.org/10.1162/LING_a_00133

Delbrück, B. (1893). *Vergleichende Syntax der Indogermanischen Sprachen*. Vol. 1. Strassburg: Trübner.

Dowty, D. (1979). *Word Meaning and Montague Grammar*. Dordrecht: D. Reidel. https://doi.org/10.1007/978-94-009-9473-7

Dowty, D. (1991). Thematic Proto-Roles and Argument Selection. *Language* 67, 547–619. https://doi.org/10.1353/lan.1991.0021

Fici Giusti, F. (1994). *Il passivo nelle lingue slave. Tipologia e semantica*. Milano: F. Angeli.

Fillmore, (1968). The case for case. In E. Bach & R. T. Harms (Eds.). *Universals in Linguistic Theory*. (pp. 1–25). London: Holt, Rinehart and Winston.

Foley, W. A. & Van Valin, R. D. , Jr. (1984). *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.

Frei, H. (1939). Sylvie est jolie des yeux. In A. Sechehaye et al. (Eds.). *Mélanges de linguistique offerts à Charles Bally* (pp. 185–192). Geneve: Georg.

Hahn, A. (1954). Partitive Apposition in Homer and the Greek Accusative. *Transactions and Proceedings of the American Philological Association* 85, 197–289. https://doi.org/10.2307/283476

Haspelmath, M. (1993). More on the typology of inchoative/causative alternation. In B. Comrie & M. Polinsky (Eds.). *Causatives and Transitivity* (Studies in Language Companion Series, 23) (pp. 87–120). Amsterdam/Philadelphia: Benjamins. https://doi.org/10.1075/slcs.23.05has

Haspelmath, M. (1999). External possession in a European areal perspective. In I. Barshi & D. Payne (Eds.). *External possession* (pp. 109–35). Amsterdam/Philadelphia: Benjamins. https://doi.org/10.1075/tsl.39.09has

Haspelmath, M., Dryer, M. S., Gil, D. & Comrie, B. (2005). *The World Atlas of Language Structures (WALS)*. Oxford: Oxford University Press.

Heine, B. (1997). *Possession*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511581908

Jackendoff, R. (1976). Toward an Explanatory Semantic Representation. *Linguistic Inquiry* 7, 1, 89–150.

Jackendoff, R. (1987). The Status of Thematic Relations in Linguistic Theory, *Linguistic Inquiry* 18, 3, 369–411.

Jacquinod, B. (1989). *Le double accusatif en grec d'Homère à la fin du V siècle avant J.C.*, Louvain: Peeters.

Jacquinod, B. (2006). Le domaine de l'accusatif de relation. In E. Crespo et al. (Eds.). *Word classes and related topics in Ancient Greek* (pp. 59–68). Louvain: Peeters.

Jacquinod, B. (2016). Entre tradition et modernité. *Syntaktika* 50, Numéro spécial: Hommage à Bernard Jacquinod, pp. 1–43.

Kemmerer, D. (2003). Why can you hit someone on the arm but not break someone on the arm? A neuropsychological investigation of the English body-part possessor ascension. *Journal of Neurolinguistics* 16, 16–36. https://doi.org/10.1016/S0911-6044(01)00042-2

König, E. & Haspelmath, M. (1997). Les constructions à possesseur externe dans les langues d'Europe, in J. Feuillet (Ed.). *Actance et valence dans les langues de l'Europe. (Empirical Approaches to Language Typology/EUROTYP, 20- 2.)* (pp. 525–606). Berlin: Mouton de Gruyter.

Kulikov, L. I. (1998). Passive, anticausative and classification of verbs: the case of Vedic. In L. Kulikov & H. Vater (Eds.). *Typology of verbal categories: Papers presented to Vladimir Nedjalkov on the occasion of his 70th birthday* (pp. 139–153). Tübingen: Niemeyer. https://doi.org/10.1515/9783110913750.139

La Roche, J. (1861). *Homerische Studien. Der Akkusativ in Homer*. Wien: Gerold.

Lavidas, N. (2013). Accusative. In: *Encyclopedia of Ancient Greek Language and Linguistics*. https://doi.org/10.1163/2214-448X_eagll_COM_00000004

Lazzeroni, R. (2002a). Ruoli tematici e genere grammaticale. Un aspetto della morfosintassi Indoeuropea. *Archivio Glottologico Italiano* 87, 1, 3–19.

Lazzeroni, R. (2002b). Il nome greco del sogno e il neutro indoeuropeo. *Archivio Glottologico Italiano* 87, 2, 145–162.

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago Press.

Levin, B. & Rappaport Hovav, M. (1995). *Unaccusativity. At the syntax/lexical semantics interface*. Cambridge, MA/London: MIT Press.

Lévy-Bruhl, L. (1916). L'expression de la possession dans les langues mélanesiénnes. *Mémoires de la Société de Linguistique* 19, 96–104.

Luraghi, S. (2003). *On the meaning of prepositions and cases. The expression of semantic roles in Ancient Greek*. Amsterdam/Philadelphia: Benjamins. https://doi.org/10.1075/slcs.67

Malchukov, A., Haspelmath, M. & Comrie, B. (2010). Ditransitive constructions: A typological overview. In A. Malchukov, M. Haspelmath, & B. Comrie (Eds.). *Studies in ditransitive constructions. A comparative handbook* (pp. 1–64). Berlin / New York: Mouton De Gruyter.

Marouzeau, J. (1969). *Lexique de la terminologie linguistique*, 3rd ed. Paris: Geuthner.

Meillet, A. & Vendryes, J. (1927). *Traité de grammaire comparée des langues classiques*, Paris: Champion.

Mirto, I. & Rosen, C. (1994). *Meronyms as predicates: An apparent 'inalienable possession' construction*. Ithaca, N.Y: Ms. Cornell University.

Nichols, J. (1988). On alienable and inalienable possession. In W. Shipley (Ed.). *In honor of Mary Haas* (pp. 475–521). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110852387.557

Nichols, J. & Bickel, B. (2005). Possessive classification. In M. Haspelmath et al. (Eds.). *The World Atlas of Language Structures (WALS)* (pp. 242–245). Oxford: Oxford University Press.

Perlmutter, D. & Postal, P. (1983). The Relational Succession Law. In D. Perlmutter (Ed.). *Studies in Relational Grammar 1* (pp. 30–80). Chicago: University of Chicago Press.

Romagno, D. (2005). La codificazione degli attanti nel Mediterraneo romanzo: Accordo del participio e marcatura dell'oggetto. *Archivio Glottologico Italiano* 90, 90–113.

Romagno, D. (2006). Gradiente di transitività e codifica dell'oggetto. Dall'accusativo preposizionale al partitivo. *Archivio Glottologico Italiano* 91, 203–222.

Romagno, D. (2007). Canonical and non-canonical marking of core arguments in European languages. A typological approach. In P. Ramat & E. Roma (Eds.). *Europe and the Mediterranean as linguistic areas. Convergencies from a historical and typological perspective* (pp. 289–315). Amsterdam/Philadelphia: Benjamins. https://doi.org/10.1075/slcs.88.13rom

Romagno, D. (2010). Anticausativi, passivi, riflessivi: considerazioni sul medio oppositivo. In I. Putzu, G. Paulis, G. F. Nieddu, & P. Cuzzolin (Eds.). *La morfologia del greco fra tipologia e diacronia* (pp. 430–441). Milano: FrancoAngeli.

Romagno, D. (2014). The aorist in –ην in Homeric Greek: At the Morphosyntax/Semantics Interface. A thorough analysis of Iliad and Odyssey. *Archivio Glottologico Italiano* 99, 155–186.

Romagno, D. (2017). The accusative of respect in Ancient Greek: Animacy hierarchy, semantic roles and event type. *Studi e Saggi Linguistici* 55, 65–91.

Rosen, C. (1976). *Accusativus graecus*: A relational analysis. In R. L. Brown et al. (Eds.) *Proceedings of the 1976 Mid-America Linguistics Conference*, pp. 285–295.

Schwyzer, E. & Debrunner, A. (1950). *Griechische Grammatik*, Bd. II. München: Beck.

Silverstein, M. (1976). Hierarchy of features and ergativity. In R. M. W. Dixon (Ed.). *Grammatical categories in Australian languages*. (pp. 112–171). Canberra: Australian Institute of Aboriginal Studies.

Smyth, H. W. (1956). *Greek Grammar* (revised by Gordon M. Messing). Cambridge, MA: Harvard University Press.

Tenny, C. L. (1994). *Aspectual Roles and the Syntax-Semantics Interface*. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-011-1150-8

Van Valin, R. D. , Jr. (1990). Semantic parameters of split intransitivity, *Language*, 66, 221–260. https://doi.org/10.2307/414886

Van Valin, R. D. , Jr. (2007). *Exploring the Syntax-Semantics Interface*. Cambridge: Cambridge University Press.

Van Valin, R. D. , Jr. & La Polla, R. J. (1997). *Syntax: structure, meaning and function*, Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139166799

Vendler, Z. (1967). *Linguistics in Philosophy*. Ithaca (NY): Cornell University Press.

Winter, K. (1971). *Formal frequency and linguistic change: some preliminary comments*, «Folia Linguistica», 5, pp. 55–61.

Texts and translations

Andocides, *On the Mysteries*: Meidment, K. J. (1968). *Andocides. Minor Attic Orators in two volumes 1, Antiphon Andocides*, with an English translation by K. J. Maidment, Harvard University Press, Cambridge, MA and William Heinemann Ltd., London.

Aristophanes, *Clouds*: Hickie, W. J. (1853). *Aristophanes. Clouds. The Comedies of Aristophanes*, by William James Hickie, vol. I, Bohn's Classical Library, London.

Aristophanes, *Clouds*: Hall, F. W. & Geldart, W. M. (1907). *Aristophanes Comoediae*, ed. by F. W. Hall and W. M. Geldart, vol. 2, Clarendon Press, Oxford.

Demosthenes: Vince, C. A. & Vince, H. (1926). *Demosthenes with an English translation by C. A. Vince and J. H. Vince*, Harvard University Press, Cambridge, MA and William Heinemann Ltd. London.

Euripides, *Medea*: Kovacs, D. (1994). *Euripides, Vol. I, Cyclops, Alcestis, Medea*, edited and translated by David Kovacs, Loeb Classical Library, Harvard University Press, Cambridge, MA.

Euripides, *Suppliants*: Oates, W. J. & O'Neill, E. (1938). *Euripides. The Complete Greek Drama*, edited by Whitney J. Oates and Eugene O'Neill, Jr. in two volumes. Vol. 1. The Suppliants, translated by E. P. Coleridge, Random House, New York.

Herodotus: Codley, A. D. (1920). *Herodotus, with an English translation by A. D. Godley*, Harvard University Press, Cambridge, MA.

Homer, *Iliad*: Murray, A. T. (1924). *Homer. The Iliad with an English Translation*, 2 volumes, Harvard University Press, Cambridge, MA.

Homer, *Odyssey*: Di Benedetto, V. (2010). *Omero, Odissea, a cura di Vincenzo Di Benedetto, traduzione di Vincenzo Di Benedetto e Pierangelo Fabrini*, BUR, Rizzoli, Milano.

Homer, *Odyssey*: Murray, A. T. (1919). *Homer. The Odyssey with an English Translation*, 2 volumes, Harvard University Press, Cambridge, MA.

IG, XIV = *Inscriptiones Graecae Siciliae et Italiae, additis Galliae, Hispaniae, Britanniae, Germaniae inscriptionibus*, ed. by Georg Kaibel, Berlin, 1890.

Plato, *Republic*: Shorey, P. (1969). *Plato in Twelve Volumes, Vols. 5 and 6, translated by Paul Shorey*, Harvard University Press, Cambridge, MA and William Heinemann Ltd., London.

Sophocles, *Ajax*: Jebb, R. (1893). *Sophocles. The Ajax of Sophocles. Edited with introduction and notes, by Sir Richard Jebb*, Cambridge University Press, Cambridge.

Sophocles, *Oedipus Tyrannus*: Jebb, R. (1887). *The Oedipus Tyrannus of Sophocles. Edited with introduction and notes by Sir Richard Jebb*, Cambridge University Press, Cambridge.

Xenophon, *Anabasis*: Brownson, C. L. (1961). *Xenophon, Anabasis, Translated by Carleton L. Browson*, Loeb Classical Library, Harvard University Press, Cambridge, MA.

Xenophon, *Memorabilia*: Marchant, E. B. (1923). *Xenophon in seven volumes, vol. IV, Translated by E. C. Marchant*, Loeb Classical Library, Harvard University Press, Cambridge, MA and William Heinemann, Ltd., London.

Criteria for subjecthood and non-canonical subjects in Classical Greek

Marina Benedetti and Chiara Gianollo

Università per Stranieri di Siena / Università di Bologna

The dative-marked argument of the verb *dokéō* 'seem' in Classical Greek displays syntactic, semantic and pragmatic properties that qualify it as a non-canonical subject. To substantiate this claim, three phenomena are analyzed, all involving coreference resolution across clause boundaries: long-distance reflexivization, interclausal coreference, and case mismatch in participial constructions. For the latter phenomenon, the observed mismatch between case marking and referential properties is captured by positing the same coreference mechanism for finite clauses and for a class of participial constructions that qualify as a full clausal domain.

Keywords: Classical Greek, non-canonical subjects, participles, coreference, case mismatch

1. Non-canonical subjects in Classical Greek

The need for a non-monolithic notion of subject, which crops up in the traditional distinction between grammatical and logical – or notional – subject, has been firmly established in more refined linguistic frameworks in the last decades.

The issue of a "Multi-Factor concept of 'Subject'" (using an expression by Keenan in his famous 1976 paper) is consistent with the empirical observation that a vast number of languages attest constructions in which an argument not formally encoded as a subject nevertheless shares behaviour / control properties and semantic properties with "canonical" subjects (i.e. with arguments formally encoded as subjects, as far as, e.g., morphological case, verb agreement, and word order are concerned). In the following we shall adopt the label "Non-canonical subject" to refer to such arguments.[1]

---

1. The crosslinguistic validity (and universality) of categories such as subject or object is a theoretically debated issue; let us simply recall here the discussion in Croft (1990: 13 ff. et passim); Dryer (1997); Aikhenvald et al. (2001) and the recent synthesis and proposal by Barðdal & Eythórsson (2018).

After Andrews' pioneering paper on Icelandic (Andrews, 1976) and the studies by Masica and Klaiman on South-Asian languages (Masica, 1976; Klaiman, 1980), investigations of non-canonical subjects have been carried out on different languages and languages families, producing considerable results, and also awakening a lively debate, as shown by the immense bibliography (cf. n. 2). The research and the debate have involved also ancient Indo-European languages like Germanic (Icelandic deserves a special mention here), Latin, Greek, Hittite, Indo-Aryan, Slavic, Celtic), thus posing interesting challenges also on prehistoric reconstruction.[2]

In general, raising, control phenomena, coreferential deletion, constraints on reflexivization etc. have proven to be valid behavioural criteria for subjecthood in a variety of languages; but, of course, they do not apply to all languages (and anyway not in the same way and with the same manifestations). The identification of a set of tests for subjecthood in individual languages is a demanding task, still open to research.

Focusing on Classical Greek, which is the object of this paper, evidence for non-canonical subjects (or "oblique", or "quirky" subjects; for an overview on the various labels cf. Seržant & Kulikov, 2013) has been put forward – in greater or lesser detail – in several works, such as Conti (2008, 2009, on the genitive case); Barðdal & Eythórsson (2009); Lühr (2011); Barðdal et al. (2012); Dahl & Fedriani (2012); Barðdal & Smitherman (2013); Danesi et al. (2018).

Much work has been done, and much is still to be done. As Sevdali remarks, some tests for subjecthood that have proven to apply to Icelandic – a language which has provided a good training ground for this kind of research – do not apply to Ancient Greek, because the relevant environments do not exist in this language or because subjects behave similarly to objects with respect to some of these tests. She claims, quite drastically, that, in order to investigate quirky subjects in Ancient Greek "we would have to invent an entirely new set of diagnostics […] distinct from diagnostics used for other languages, those with strict word order and expletive subjects such as Icelandic and Old English." (Sevdali, 2013, p. 18).

In this paper, we wish to take a deeper look into some hitherto not discussed phenomena related to possible subject behaviour of dative arguments in Classical

---

2.   We list here only a few bibliographical references, regarding both general issues and phenomena of individual languages and language families: Hock (1990); Barðdal (2000); Aikhenvald et al. (2001); Eythórsson & Barðdal (2005); Barðdal & Eythórsson (2009); Fedriani (2009, 2014); Luraghi (2010); Seržant & Kulikov (2013); Le Mair et al. (2017) (with rich references). In the debate, of course, scepticism or critical remarks have been expressed on the evaluation of some adduced subjecthood properties in individual languages. For Greek cf., e.g., Sevdali (2013: 18), for Latin, Jøhndal (2012: 24 ff.). For a synthesis and critical discussion on oblique subjects in the history of Indo-Aryan cf. Verbeke et al. (2015).

Greek. While we agree with Sevdali that the validity of subject tests has to be assessed language-internally and that Ancient Greek is particularly challenging in this respect, especially as concerns control phenomena, we argue that at least some classical tests (reflexivization, interclausal coreference) can be applied in some contexts also in this language and may reveal non-canonical subjecthood phenomena. For the purposes of this work, our data will be presented in a theoretically informed way, but without subscribing to a specific formal framework, since most syntactic models share the assumptions we will need to adopt.

2.   A case study: The dative argument of *dokéō*

Our case study is represented by occurrences of a dative argument in combination with the verb *dokéō* 'seem', in Classical Greek, in the construction types exemplified below:[3]

(1)   ποιέειν αὐτίκα μοι δοκέει   (H. 3.71.8)
      'it seems (good) to me to act immediately'
      poiéein autíka   moi   dokéei
      act-INF immediately me-DAT.SG seem-PRS.3.SG

(2)   ἐδόκεε δέ οἱ ἐκ τῶν αἰδοίων τῆς θυγατρὸς ταύτης φῦναι ἄμπελον
                                                                                 (H. 1.108.2)
      'It seemed to him that there grew a vine from his daughter's genitals'
      edókee   dè   hoi   ek   tôn   aidoíōn
      seem-IPFV.3.SG PTC him-DAT.SG from DET-GEN.PL genital-GEN.PL
      tês   thugatròs   taútēs   phûnai   ámpelon
      DET-GEN.SG daughter-GEN.SG DEM-GEN.SG be born-INF vine-ACC.SG

(3)   καλῶς μοι δοκεῖς λέγειν   (X. Cyr. 5. 31)
      'you seem to me to speak well'
      kalôs moi   dokeîs   légein
      well me-DAT.SG seem-PRS.2.SG speak-INF

---

3.   We shall concentrate on constructions involving a dative argument; other types are not relevant for the present discussion, e.g. the non-raising nominative construction (*dokéō* 'I think' + infinitive). Our corpus includes Herodotus' *Histories*; Thucydides' *Histories*; Xenophon's *Anabasis* and *Cyropaedia* (occasionally, passages from other authors); the following forms have been collected (through the electronic resources of Thesaurus Linguae Graecae®): all personal forms in the present indicative; the third person singular of all other tenses and moods. We shall follow a qualitative and not a quantitative approach; what is here relevant is the presence of certain – hitherto mostly unobserved – phenomena. Translations are adapted from the digital Loeb edition. Glosses and abbreviations according to the Leipzig Glossing Rules, with following additional abbreviations: AOR = aorist; PLUSPFV = PLUSQUAMPERFECT; PTC = particle.

In all these examples *dokéō* occurs in a matrix clause accompanied by an infinitive complement clause.

In (1) and (2) we have impersonal constructions. In (1) the dative argument of *dokéō* (*moi*) controls the (PRO) subject of the infinitive (*poiéein*). In (2), the infinitive complement clause ([…] *phûnai ámpelon*) has its own independent subject, distinct form the dative argument of *edókee* (i.e. *hoi*), and showing up in the accusative case (so-called accusative and infinitive construction).

On the other hand, (3) is a personal (raising) construction; the nominative argument controlling agreement on *dokéō* (the "grammatical subject", here formally manifested through the person and number features on the 2sg. verb) is an argument licensed by the infinitive (*légein*) and raised to subject of the matrix clause.

There are good reasons to more deeply investigate the dative argument of *dokéō* as a good candidate for non-canonical subjecthood (as already suggested by Lühr, 2011). Among them, semantic reasons are very relevant: in fact, *dokéō* may be qualified as a verb of cognition; it thus belongs to a semantic class that appears to frequently attest oblique subject constructions, as results from crosslinguistic evidence (on the relationship between semantic verb classes and non-canonical subjects, cf. Onishi, 2001; Barðdal et al. 2012, with further references).

Moreover, an element of interest in the *dokéō*-construction is that it offers, in the raising type (cf. (3) above), the possibility of comparing, so to say, two competing "subjects": the one displaying subject coding properties (nominative, verb agreement), and the dative one displaying, as we claim, subject behavioural properties.

In what follows, we would like to more precisely characterize the subject-like properties of the dative experiencer selected by *dokéō*, and to discuss how to capture the notion of "logical subject" in more rigorous semantic, syntactic, and pragmatic terms.

Three phenomena will be taken into account, all involving coreference resolution across clause boundaries: long-distance reflexivization, interclausal coreference and case mismatch in participial constructions. The first two of them will be illustrated briefly in Section 3 and 4, respectively, whereas the third will be the object of a more detailed investigation in Section 5. All these phenomena point to behavioural and/or pragmatic subject properties of the dative argument of *dokéō*. Such properties are prototypically attributed to arguments that also exhibit subject coding properties. With *dokéō*, conversely, these properties belong to the dative-marked argument, not to the nominative-marked one. The fact that, thanks to these properties, the dative-marked argument ends up being more salient in discourse, with a number of syntactic consequences, is for us a manifestation of non-canonical subjecthood, according to a multi-factor notion of subject that also considers its pragmatic properties.

For our argument to be viable it is essential that the above-mentioned properties clearly single out subjects in Ancient Greek, and do not extend to canonical

(direct or indirect) objects. The validity of commonly held assumptions in this regard has been tested by means of (i) a cross-check of the properties of the canonical indirect object marked with dative case appearing with the verb *dídōmi* "to give" in Herodotus' *Histories*, and (ii) a thorough annotation of the properties of null subjects and reflexives in selected chapters from Herodotus (7.1–9) and Xenophon (*Cyr.* 6.1.1–38). These texts have been chosen as a "control sample" among those which display the *dokéō*-properties relevant for our discussion. The aim is to test the claim that the properties which the dative argument of *dokéō* shares with subjects are not also shared by unambiguous objects (such as dative Recipients / Beneficiaries and accusative Patients). As we will see in the further sections, the only exceptions are found in cases where the direct or indirect object has special semantic-pragmatic properties that speak for its non-canonical nature.

3.   Reflexivization

In Classical Greek, the pronoun *heautón*[4] may occur as a long-distance reflexive, i.e. it may refer to an antecedent in a higher clause. It is widely assumed that the antecedent of this long-distance reflexive generally coincides with the (nominative) subject of the higher clause (see, for example, Crespo et al. 2003, p. 46; according to Peels, 2007, p. 37, "*heōutón* is highly subject oriented"; cf. also Tronci, 2012). With reference to the forms like *heautón* in Classical Greek, Kiparsky (2012, p. 4) states that they "require a subject antecedent within the same finite domain, either in the same clause or across an infinitive clause boundary".[5]

Exceptions to this requirement have occasionally been observed, and ascribed to some sort of subject-like character of such antecedents: "the pronoun refers to what is, in sense, though not in grammar, the subject" (Powell, 1933, p. 220); "the logical but not the grammatical subject" (Powell, 1934, p. 172); "a dative virtual subject" (Cooper 1998: p. 501).[6]

---

4.   On this form, compound of the inherited reflexive *he* (< *swe*) and the pronoun *autón* cf. Petit (1999).

5.   A further cooccurring restriction sometimes observed in traditional grammars is that the subject of the matrix clause (and antecedent of *heautón*) must denote the person, distinct from the speaker, whose point of view is being presented in the *heautón*-clause (Kühner-Gerth, 1898, p. 561 f.; Humbert, 1960, p. 62f.). This would point to what is now called a "logophoric" function. On the relationship between long-distance reflexives and logophoricity cf. Siewierska (2004, p. 201 ff.); for Greek, cf. the discussion in Peels (2007), Kiparsky (2012).

6.   A pragmatic perspective is introduced by Peels (2007, p. 37): "when the antecedent is not a subject, it is still an experiencer/topic".

The dative argument of *dokéō* belongs to the non-nominative arguments that may be the antecedent of long-distance *heautón*, as in (4):

(4) **oἱ**ᵢ ὅ τε χαρακτὴρ τοῦ προσώπου προσφέρεσθαι ἐδόκεε **ἐς ἑωυτόν**ᵢ

(H. 1.116.3)

'the feature of [the boy's] face seemed to him to be similar to his own'

hoiᵢ   *hó*   *te*   *kharaktḕr*   *toû*   *prosōpou*
he-DAT.SG DET-NOM.SG PTC feature-NOM.SG DET-GEN.SG face-GEN.SG
prosphéresthai edókee   *es* **heōutôn**ᵢ
be-similar-INF seem-IPFV.3SG to REFL-ACC.SG

In the personal construction in (4) *dokéō* combines both with a nominative and with a dative argument: the antecedent of the reflexive is the dative argument (*hoi* 'to him') and not the raised subject (*hó kharaktḗr* 'the feature').[7] This clearly confirms that nominative marking is no necessary condition on the antecedent of *heautón*.

We may thus observe that, with respect to long-distance reflexivization, the dative argument of *dokéō* displays properties that are usually associated with subjects (and which, as far as we know, are not shared by "canonical" indirect objects).[8]

We should also remark, however, that this property of the dative argument of *dokéō* can be safely assessed only in personal constructions. No decisive argument, instead, can be drawn from impersonal ones. In impersonal constructions we do find instances of coreference between the dative argument of *dokéō* and a reflexive in the embedded clause: in (5), for example, the reflexive pronoun (*heōutôn*) of the infinitival clause has its antecedent in the dative argument of *dokéō*, i.e. *toîsi basileûsi* 'to the kings' (in the matrix clause). However, following what we observed concerning (1) above, we analyze the example in (5) (and similar ones, such as Hdt. 4.111.8), as a control construction. The dative argument of *dokéō* (*toîsi basileûsi*) controls the (PRO) subject of the embedded infinitive, which in turn binds the reflexive in its domain: hence, the binding relation is local, between the PRO subject and the anaphor within the infinitival clause.

---

7. This fact recalls Keenan's remark on a sort of reduced subjecthood of derived subjects (cf. Keenan, 1976, p. 323). On the other hand, the raised subject is not ruled out as a possible antecedent of *heautón*, as shown by X. *Cyr.* 7.1.6.6.

8. On similar phenomena in Latin cf. Pieroni (2007), who shows that Latin oblique Experiencers can be the antecedent of reflexives.

(5) **τοῖσι** δὲ **βασιλεῦσι**$_i$ δόξαι ἐν τῇ **ἑωυτῶν**$_i$ [scil. χώρῃ] (**PRO**$_i$) κεῖσθαι ἀποθα-
     νόντας   (H. 4.11.16)
     'the kings preferred (lit. 'it seemed to the kings') to lie slain in their own country'
     toîsi   dè   basileûsi$_i$ dóksai   en têi   heōutôn$_i$
     DET-DAT.PL PTC king-DAT.PL seem-INF in DET-DAT.SG REFL-GEN.PL
     *[khṓrēi]   (**PRO**$_i$) keîsthai apothanóntas*
     land-DAT.SG   lay-INF dead-ACC.PL.M

The ability of direct objects and indirect objects to control null infinitival sub-
jects is a well-known property of Ancient Greek (Sevdali, 2013; Goldstein, 2016,
p. 263–268, Danesi et al. 2018, p. 50; for similar facts in Latin cf. Jøhndal, 2012,
p. 54–56), which is by no means restricted to subject-like arguments. For example,
as we see in (6), the dative argument of a verb of command like *epitássō* (which is
a canonical indirect object), is able to control the PRO subject of its complement
clause, which in turn binds the reflexive in its domain:

(6) χαλεπώτερον γὰρ ἄν, ἔφη φάναι, εἰ **αὐτῇ**$_i$ ἐπέταττον (**PRO**$_i$) ἀμελεῖν τῶν **ἑαυτῆς**$_i$
                                                                        (X. Oe. 9.18.6)
     'it would have been harder, she said (lit.: she declared having said), had I
     required her to neglect her own possessions'
     khalepṓteron   gàr án éphē   phánai ei autêi
     harder-NOM/ACC.N.SG PTC PTC say-AOR.3SG say-INF if her-DAT.SG
     *epétatton   (**PRO**$_i$) ameleîn   tôn   heautês*
     command-IPFV.1SG   neglect-INF DET-GEN.PL REFL-GEN.SG

Summing up, coreference of the dative argument of *dokéō* with a reflexive in the
embedded clause (a) offers conclusive evidence for the subject-like status of the
dative argument in personal (raising) constructions such as (4), where the anaphor
is bound by the dative argument of *dokéō*; (b) is not relevant for the present dis-
cussion in impersonal (control) constructions such as (5), in which the anaphor is
not directly bound by the dative argument of *dokéō* but by the PRO subject of the
infinitival clause.

4.   Interclausal coreference with finite clauses

The dative argument of *dokéō* may co-refer with a (*pro*) nominative subject in a
coordinate or subordinate clause, as shown in (7)–(11):

(7) Ἔδοξέ τε **τοῖσι Πελοποννησίοισι**$_i$ ταῦτα εἶναι ποιητέα **καὶ (pro**$_i$**) ἔταμον** ὅρκιον

(H. 9.26.17)

'It seemed to the Peloponnesians that this should be done and [they] swore a
compact'

*édokse   dè   **toîsi   Peloponnēsioisi**$_i$ taûta*
seem-AOR.3SG PTC DET-DAT.PL Peloponnesian-DAT.PL DEM-NOM/ACC.PL.N

*eînai   poiētéa   kaì   (**pro**$_i$)   étamon   hórkion*
be-INF to_be_done -NOM/ACC.PL.N and   take-AOR.3PL oath-ACC.SG

(8) Κάρτα τε ἔδοξε **τῷ βουκόλῳ**$_i$ πρὸς τὰ παρεόντα εὖ λέγειν **ἡ γυνή**$_j$, καὶ αὐτίκα
(pro$_i$**) ἐποίεε** ταῦτα   (H. 1.113.1)

'Thinking that his wife counselled him exceeding well in his present strait, the
cowherd straightway did as she said'

*kárta te   édokse   **tôi   boukólōi**$_i$ pròs   tà*
very and seem-AOR.3SG DET-DAT.SG cowherd-DAT.SG towards DET-ACC.PL

*paréonta   eû   légein   **hē**   **gunḗ**$_j$ kaì*
present_things-ACC.PL well speak-INF DET-NOM.SG wife-NOM.SG and

*autíka   (**pro**$_i$)   epoíee   taûta*
immediately   do-IPFV.3SG DEM-ACC.PL.N

(9) Κῦρος$_i$ μὲν οὕτως ἔλεξε· τῷ δὲ **Κυαξάρῃ**$_j$ **(pro**$_i$**)** ἔδοξέ τε εὖ λέγειν, καὶ τοῦ μὲν
πλείους μεταπέμπεσθαι οὐκέτι **(pro**$_j$**)** ἐμέμνητο, παρεσκευάζετο δὲ τὰ ὅπλα

(X. Cyr. 2.1.10.2)

'Thus Cyrus spoke. And to Cyaxares it seemed that he spoke to the point (lit.:
And he seemed to Cyaxares to speak to the point); and he (=Cyaxares) no longer
talked of sending for reinforcements, but he set about procuring the arms'

Kûros$_i$ mèn houtōs élekse;   tōi   dè
Cyrus-NOM.SG PTC so   speak-AOR.3SG DET-DAT.SG PTC

***Kuaksárēi**$_j$ (**pro**$_i$) édoksé   te   eû   légein,   kaì toû*
Cyaxares-DAT.SG   seem-AOR.3SG PTC well speak-INF and DET-GEN.SG

*mèn pleíous   metapémpesthai oukéti   (**pro**$_j$) emémnēto,*
PTC more-ACC.PL send-INF   no_longer   meditate-PLUSPFV.3SG

paraskeuázeto   dè   tà   hópla
prepare-IPFV.3SG PTC DET-ACC.PL arm-ACC.PL

(10) Ταῦτα εἰπὼν **(pro**$_j$**)** ἔδοξε **τῷ Κλεάρχῳ**$_i$ ἀληθῆ λέγειν, καὶ **(pro**$_i$**)** εἶπεν (sc.
Κλέαρχος) […]   (X. An. 2.5.24.2)

'In these things that he said he (sc. Tissaphernes) seemed to Clearchus to be
speaking the truth; and Clearchus said (…)'

*taûta   eipṑn (**pro**$_j$)   édokse   **tōi**   **Kleárkhōi**$_i$*
DEM-ACC.PL say-PTCP.N.SG.M seem-AOR.3SG DET-DAT.SG Clearchus-DAT.SG

*alēthê   légein, kaì   (**pro**$_i$) eîpen*
true-ACC.PL say-INF and   say-AOR.3SG

(11) ἡνίκα δὲ ἔδοξε **τῷ Κύρῳ**ᵢ καιρὸς εἶναι, **(pro**ᵢ**)** ἐξῆρχε παιᾶνα (X. Cyr. 7.1.25.4)
'Then, when it seemed to Cyrus to be just the right time, he began the paean'
*hēníka dè édokse **tōi**   **Kúrōi**ᵢ   kairòs   eînai,*
when seem-AOR.3SG DET-DA.SG Cyrus-DAT.SG right_time-NOM.SG be-INF
*(pro*ᵢ*) eksêrkhe   paiâna*
        begin-AOR.3SG paean-ACC.SG

As is well known, in the investigation of pro-drop languages, the conditions which license null subjects and the strategies for recovering their antecedents launch interesting challenges to linguistic research. Although such an investigation is still missing for Classical Greek, the passages above are worth noting because they are instructive as to the nature of the dative argument of *dokéō*. In all of them, a *pro* (nominative) subject of a finite clause finds its antecedent in the dative argument of *dokéō* located in a preceding coordinate (in (7)–(10)) or subordinate (in (11)) finite clause. This clearly proves that identity in case marking does not provide the unique strategy for recovering the antecedent of zero-anaphoras.

Particularly interesting is the case in which *dokéō* occurs in a personal construction, being thus accompanied both by a dative and by a nominative (raised) argument, as in (8)–(10): remarkably, here the dative argument (and not the nominative one) is – despite the case mismatch – the antecedent of the *pro* (nominative) subject of the coordinate clause. So, in (8), the null subject of *epoíee* corefers with the dative *tōi boukólōi* and not with the nominative *hē gunḗ*.[9] In (9), the null subject of *emémnēto*, *paraskeuázeto* corefers with the dative *Kuaksárēi* and not with the (null) nominative subject of *édokse* (referring, in the scene described, to Cyrus). Similarly, in (10), the null subject of *eîpen* corefers with the dative *tōi Kleárkhōi*, and not with the (null) nominative subject of *édokse* (referring, in the scene described, to Tissaphernes).

It should be noted that, in all these cases, no agreement feature and no semantic constraint can be invoked as a disambiguating factor, as all nominals involved are 3sg. and all denote humans. Rather, the notion of topicality seems to be relevant here: in the examples quoted, the dative argument introduces the topic of the immediately following narrative, and the coreferent zero anaphora signals topic continuity. So, the cowherd (dat. *tōi boukólōi*) and not his wife (nom. *hē gunḗ*) is the topic of the sentence in (8); there follows (in the coordinate clause as well as in following sentences) a sequence of zero anaphoras, with the cowherd as their referent. Similar remarks apply to the other passages.

---

9. A similar passage is in Hdt. 3.119.27.

Summing up, the dative argument of *dokéō* may corefer with null nominative arguments across clause boundaries: it appears to be a highly accessible antecedent for null (nominative) subjects of a coordinate or matrix clause.

Accessibility may be influenced by various parameters (with different relevance in each individual language) but is undoubtedly closely associated with subject properties. In as theory-neutral terms as possible, we can state the generalization governing the resolution of null subjects as follows (12):

(12)   A null subject (a *pro*) should be construed with the most accessible compatible antecedent, whereby accessibility is a dynamic multifactorial notion
(Givón, 1983; Ariel, 1988, 2001; Dimitriadis, 1996)

The antecedent's degree of accessibility is sensitive to different linguistic factors, such as its grammatical role, syntactic position and the recentness of its last mention. In particular, *pro* is used in situations of topic continuity, a subset of the situations where a referent is salient enough to be recoverable.

The fact that coreference appears related to topic continuity has been observed repeatedly for null subject languages. In particular, work on Modern Greek has shown that topic continuity obtains when *pro* is used, whereas an overtly realized pronoun is used to interrupt the topic chain and perform a topic switch (cf. Dimitriadis, 1996; Tsimpli et al. 2004; Torregrossa et al. 2015). Similar observations have been made for Italian (Frascarelli, 2007). Existing literature, in particular Goldstein (2016, p. 144–165), and our own corpus work confirm that a similar condition holds for Classical Greek as well: a null subject is used in cases of interclausal topic continuity, whereas cases of topic switch are signalled by the presence of an overt subject, typically accompanied by the particle *dé*.

Since the dative argument of *dokéō* has topic status, it is more salient than the co-occurring nominative argument of *dokéō* for interclausal coreference, and becomes the most prominent – and thus preferred – antecedent for zero anaphora. Following Keenan (1976, p. 318), we take topicality to be a prototypical pragmatic property of subjects. Importantly, the properties that we observe for the dative argument of *dokéō* do not extend to canonical oblique arguments. Crespo et al. (2003, p. 361) remark that the subject has to be overtly expressed if it is coreferent with an oblique argument of the previous clause. Goldstein (2016, p. 149) adds that in these cases the subject, typically sentence-initial, is marked by the particle *dé*. Our cross-check of the behaviour of the dative argument of *dídōmi* "to give" (a canonical indirect object) in Herodotus is in line with these observations: if the referent of the dative argument becomes a subject in the following clause, the subject is overtly expressed and accompanied by *dé* (10 cases) or *ge* (1 case).[10]

---

10.   In Herodotus, we found only one example (Hdt. 9.94.18) where a null subject is coreferent with a preceding dative argument of *dídōmi*. This can be explained by the fact that the whole

As discussed e.g. by Haspelmath (2001), depending on the notion of subject adopted, behavior in interclausal coreference may or may not qualify as a subjecthood test. Given the multi-factor notion of subject that we adopt, which crucially includes pragmatic properties, we consider these phenomena relevant to define the subject status of an argument, and we conclude that they characterize the dative argument of *dokéō* as a non-canonical subject.

5.   Interclausal coreference with participial clauses

In this section we discuss a further phenomenon of interclausal coreference, this time involving coreference between a finite and a non-finite clause, the latter represented by a participle. This case study will require more elaboration, since our analysis involves a proposal on the internal structure of the participial clause, which we will need to present in some detail.

The dative argument of *dokéō* may co-refer with the nominative subject (expressed or unexpressed) of a participle, as in the following passages, where a matrix clause with the verb *dokéō* is accompanied by a participial adjunct clause, i.e. a non-finite subordinate clause with an adverbial function.

The examples in (13)–(14) are cases where the participle has no overt subject (indicated with $Ø_i$) and inflects in the nominative.

(13)   [$Ø_i$ **Βουλόμενος** δὲ κατάσκοπόν τινα πέμψαι ἐπὶ Λυδίας (..)] ἔδοξεν **αὐτῷ**$_i$ ἐπιτήδειος εἶναι **Ἀράσπας**$_j$ ἐλθεῖν ἐπὶ τοῦτο   (X. Cyr. 6.1.31)
'Now, wishing to send somebody as a spy into Lydia (..) Araspas seemed to him to be the proper person to go on this mission'
[$Ø_i$ ***boulómenos***   *dè   katáskopón tina   pémpsai epì*
wish-PTCP.NOM.SG.M   PTC   spy-ACC.SG someone-ACC.SG send-INF   into
Ludías]   édoxen   autôi$_i$ *epitédeios   eînai*
Lydia-GEN.SG seem-AOR.3SG DEM-DAT.SG apt-NOM.SG.M   be-INF
Ἀράσπας$_j$ ἐλθεῖν ἐπὶ τοῦτο
Aráspas   *eltheîn epì toûto*
Araspas-NOM.SG.M go-INF for DEM-ACC.SG

passage in which this example occurs shows an alternation between two contrastive topics, and no ambiguity may occur because one referent is plural and the other is singular. Moreover, temporal expressions signal the topic switch in the example at hand.

(14) [ἅπαντα γὰρ Ø$_i$ **προσκεψάμενοι** καὶ **ἐπιλογισθέντες** ὅτι οὔτε πλήθεϊ ἕξουσι χρᾶσθαι οἱ βάρβαροι οὔτε ἵππῳ,] ταύτη **σφι**$_i$ ἔδοξε δέκεσθαι τὸν ἐπιόντα ἐπὶ τὴν Ἑλλάδα (H. 7.177.4)

'For after due survey they reckoned that the foreigners could not make use of their multitude, nor of their horsemen; and therefore they resolved (lit.: it seemed to them) that here they would encounter the invader of Hellas'.

[*hápanta   gàr   Ø$_i$ **proskepsámenoi**   kaì   **epilogisthéntes***
all-ACC.PL PTC   consider-PTC.NOM.PL.M and reckon-PTC.NOM.PL.M
hóti oúte   plēthei   heksousi   khrâsthai hoi
that neither multitude-DAT.SG can-FUT.3PL employ-INF DET-N.PL
*bárbaroi   oúte híppōi,]   taútēi   **sphi**$_i$*
barbarian-NOM.PL nor cavalry-DAT.SG DEM-DAT.SG them-DAT.PL
édokse   dékesthai   tòn   epiónta   epì
seem-AOR.3SG encounter-INF DET-ACC.SG attack-PTCP.ACC.SG against
tền   Helláda
DET-ACC.SG Hellas-ACC.SG

The context leaves no space for ambiguity: despite the case mismatch, the participle corefers with the dative argument of *dokéō*, which represents its referential anchor.

In (15)–(16) we see cases where the subject of the participle is explicitly realized. Syntactic and semantic facts lead us to assume that the nominative phrases *ho Kûros* 'Cyrus' and *hḗ* 'she' belong to the participial clause (for the same structural assumption cf. Haug, 2017). They are, thus, the overt counterpart of the null category indicated as Ø$_i$ in (13)–(14), and can help us identify its nature.

(15) [Ἐκ δὲ τούτου **ἐπιθυμῶν ὁ Κῦρος**$_i$ ἤδη κατασκευάσασθαι καὶ αὐτὸς ὡς βασιλεῖ ἡγεῖτο πρέπειν], **ἔδοξεν αὐτῷ**$_i$ τοῦτο σὺν τῇ τῶν φίλων γνώμῃ ποιῆσαι

(X. *Cyr.* 7.5.37)

'After this, as Cyrus wished (lit.: wishing) to establish himself as well in the forms that he thought were proper for a king, it seemed to him necessary to do it with the approval of his friends'

[*ek   dè   toútou   **epithumôn**   **ho***
after PTC DEM-GEN.SG wish-PTCP.NOM.SG.M DET-NOM.SG.M
***Kûros**$_i$ ḗdē   kataskeuásasthai kaì autòs   hōs*
Cyrus-NOM.SG.M already establish-INF   also himself-NOM.SG as
*basilê   hēgeîto   prépein]   **édoksen   autôi**$_i$*
king-DAT.SG think-IPFV.3SG be_proper-INF seem-AOR.3SG DEM-DAT.SG
toûto   sùn   têi   tôn   phílōn
DEM-ACC.SG with DET-DAT.SG DET-GEN.PL friend-GEN.PL
gnṓmēi   poiêsai
approval-DAT.SG do-INF

(16) [καὶ **ᾗ**ᵢ οὐκ **ἔχουσα** διαφυγεῖν (ἔμπροσθε γὰρ αὐτῆς ἦσαν ἄλλαι νέες φίλιαι, ἡ δὲ αὐτῆς πρὸς τῶν πολεμίων μάλιστα ἐτύγχανε ἐοῦσα),] ἔδοξέ **οἱ**ᵢ τόδε ποιῆσαι
                                                                        (Hdt. 8.87.2)

'She could not escape, for other allied ships were in front of her and hers was the nearest to the enemy. So she resolved to do this'

*[kaì **hē**ᵢ   ouk ékhousa   diaphugeîn (émprosthe gàr*
and DET-NOM.SG.F not be_able.PTCP.NOM.SG.F escape-INF in front   PTC

autês   êsan   allai   nées   phíliai,
DEM-GEN.SG be-IPFV.3PL other-NOM.PL ship-NOM.PL allied-NOM.PL

hē   dè   autês   pròs tôn   polemíōn   málista
DEM-N.SG PTC DEM-GEN.SG near DET-GEN.PL enemy.GEN.PL above_all

*étugkhane   eoûsa)]   édokse   **hoi**ᵢ*
happen-IPFV.3SG be-PART.NOM.SG.F *seem*-AOR.3SG she-DAT.SG

tóde   poiêsai
DEM-ACC.SG do-INF

The peculiarity of these examples consists in the case mismatch between the participle (in the nominative) and the coreferent argument on the matrix clause (in the dative). The coreference is non-canonical for Classical Greek because it is not accompanied by case sharing between the referential anchor in the main clause and the participle form.

Compare, for a contrast, the example in (17), with the participle agreeing in case with the nominative subject (a *pro*) of the matrix clause.

(17) **Ø**ᵢ **βουλόμενος** τὰ ἐπὶ τούτοις παρασκευάζειν […] ἔκπλουν ποιεῖται
'wishing to make preparations for the next actions, he sails out'   (T. 1.65.1)
*Ø**ᵢ boulómenos   tà   epì toútois   paraskeuázein [..]*
    wish-PTCP.NOM.SG.M DET-ACC.PL.N after DEM-DAT.PL prepare-INF

*(pro**ᵢ*) ékploun   poieîtai*
    sailing _out-ACC.SG do-IND.PRES.3SG

Compare also (18), where, again, *dokéō* appears, but the participle this time agrees in case with the dative argument in the matrix clause (showing that case mismatch is by no means obligatory with this verb):

(18) ἀλλ' ἑτεροῖός **σφι**ᵢ δοκέοι εἶναι μὴ **ὁρῶσι**ᵢ.   (H. 1.100.1)
'but he might seem to them different if they didn't see [him]'
*all(à) heteroîós   **sphi**ᵢ   **dokéoi**   eînai [mè*
but   different-NOM.SG them-DAT.PL.M seem-OPT.3SG be-INF not
***orôsi**ᵢ]*
see-PTCP.DAT.PL.M

Such phenomena have occasionally been pointed out in historical grammars, and described with reference to the notion of "logical subject".[11] Examples of this kind are rare, but we believe that it is not by chance that they occur with predicates like *dokéō*. In our corpus we do not find similar constructions occurring with other verbs.

We argue that these phenomena of non-canonical coreference are attested because from a semantic-pragmatic point of view the dative argument has some subject properties. More precisely, it is salient in discourse and it is an aboutness topic.

In our analysis, therefore, the fact that a participle in the nominative corefers with a dative argument provides a hint to the non-canonical subject status of the dative argument of *dokéō*.

We can now spell out in more detail how the intended coreference comes about. We analyze adverbial adjunct participles in (13)–(16) as full clauses (CPs). Classical Greek participles express a rich array of grammatical categories: voice (the aorist distinguishes three forms: active, middle-passive and passive), aspect (imperfective vs. perfective), and tense (e.g. future participle).[12] We argue that, as a consequence, in certain syntactic configurations Greek participles may be construed with their own lexical subject, whose reference may coincide with the subject referent of the main clause, or may be distinct from it.

That is, while the subject of certain participial constructions is necessarily controlled (a PRO, cf. Haug, 2017), in other constructions, among which we class our examples of non-canonical coreference, the subject of the participle is referential, and is realized either by an overt nominal phrase or by a covert category that we identify as *pro*. In this way, we reach a unification with respect to the cases of coreference across finite clauses seen in Section 4.

Extending Sevdali's (2013) proposal on Classical Greek infinitival clauses to participles, we argue that adjunct participial clauses may come in two forms, a "strong", more clause-like, one and a "weak", structurally simpler, one. The hypothesis that non-finite forms are structurally heterogeneous has been put forward for various modern languages (cf. Pires, 2006 for gerunds and infinitives, and the papers collected in Nikolaeva, 2007) and has been defended for participles in Classical and New Testament Greek by Bary & Haug (2011) and Goldstein (2016).[13]

---

11. So, Kühner-Gerth (1904, p. 493) explain this kind of data by paraphrasing *dokeî moi* 'it seems to me' with *egṑ hēgoûmai* '*I* think', whereas Cooper (1988, I, p. 836 f.) claims that the dative here represents a "virtual subject".

12. The same holds for infinitives, which however, unlike participles, are not inflected for case, gender, number (agreement features).

13. In addition to the adverbial function, Greek participles can also have an attributive function or a predicative function, cf. Goldstein (2016) for an overview of uses. Our proposal on the internal structure of participial clauses is limited to adverbial participles, but see Goldstein (2016: Chapter 7) for an extension of his similar proposal also to other functions.

Our hypothesis makes clear predictions concerning, in particular, the subject of participles: corresponding to the strong-weak divide, there are two different ways for adjunct participles to assign case to their subjects.

With "weak" adjunct participial clauses, coreference is required between the participle's subject and the matrix clause's subject. The syntactic mechanisms ensuring coreference is usually captured in terms of control (cf. Haug, 2017); according to this analysis, the participle's subject is a null anaphoric pronoun PRO. An example of this strategy is (18).

In "strong" adjunct participial clauses, on the other hand, the participle's subject may have disjoint reference with respect to the matrix clause's subject. The participle's ability to licence a referential subject is linked to its clause-like status, i.e. to the fact that it represents a separate domain for tense, modality, negation, information structure, clitic positioning. Goldstein (2016: Chapter 7) reviews these factors in detail and, following Bary & Haug's (2011) analysis, argues that the independence of the participle's semantic tense with respect to the tense of the matrix clause is a determining factor ensuring clausal status. The "strong" type is quite uncontroversially instantiated by genitive absolute participial clauses. We argue that also some adjunct nominative participial clauses can display this structural property, the Examples (13)–(16) being cases in point.

In strong adjunct participial clauses, the subject may remain unexpressed (13–14), in which case we assume that it is syntactically instantiated by a phonetically null referential pronominal category (a *pro*), or may be expressed by a full nominal phrase (15–16). According to this analysis, we reduce coreference between the nominative argument of the participle and the dative argument of *dokéō* to an effect of standard pronominal resolution strategies. What is special in the cases of non-canonical coreference is that the referential chain is established between an argument of the subordinate participial clause in the nominative and a dative argument in the matrix clause.

The semantic-pragmatic properties of the dative experiencer make it even more prominent than the nominative argument of *dokéō* (present in its personal construction, cf. 3). This is clearly shown by (13). In (13) we have a nominative participle with an unexpressed $\varnothing_i$ = *pro* subject, and *dokéō* as matrix verb with both its nominative and dative arguments expressed. Now, the nominative argument of the matrix clause ('Araspas') is 'by-passed' in the establishment of the coreferential chain between the *pro* subject of the participle and an element of the matrix clause: the referential anchor is the dative-marked overt pronoun (in turn, resolved to an element of the broader context, i.e. Cyrus) that realizes the argument of *dokéō*.

We take this to be a test for the non-canonical subject status of the dative argument. The kind of coreferential chain seen in (13) is possible due to the rules that govern the identification of null subjects cross-linguistically, mentioned in Section 4. In our example, the dative experiencer clearly outweighs the (raised)

nominative subject in terms of saliency. Shared topicality (yielding topic conti-
nuity) is clearly an important factor favouring coreference in our example, as in
interclausal coreference with finite clauses seen in Section 4.

We conclude, again, that saliency and topicality are prototypical properties of
subjects, which in this case belong to a non-canonically marked subject. In this
respect, the dative argument of *dokéō* qualifies as a non-canonical subject.

Our proposal follows Keydana (1997) in assuming that in the instances of
non-canonical coreference seen in (13)–(16) the nominative on the participle is
used as the unmarked case for subjects / topics and is a device that facilitates par-
ticipant tracking (Buijs, 2013). The presence of the nominative signals an open
predication (not a closed one, as it would be with absolute genitive constructions,
cf. Haug, 2010), that is, the need to establish a referential chain with an element of
the matrix clause.[14]

To sum up, our proposal is that the few cases where non-canonical coreference
is observed are motivated by the subject-like properties of the dative argument of
dokéō serving as referential anchor in the matrix clause, i.e. by its non-canonical
subjecthood, and are made possible by the syntactic properties of the Classical
Greek participle.

6. Conclusions

We presented and analysed a series of phenomena concerning an argument that, as
we argue, qualifies as non-canonical subject in Classical Greek, i.e. the dative experi-
encer argument of *dokéō* 'seem'. This experiencer argument is marked with dative (no
subject coding properties) but has some syntactic and pragmatic subject properties.
In particular, we discussed the cases where it is the referential antecedent of long
distance reflexives, of the null subject of finite clauses and of the (null or expressed)
subject of a nominative adjunct participle. We conclude that the above phenomena
may provide a valid test for non-canonical subjecthood in Classical Greek.

In our analysis we capture the mismatch between case marking and referential
properties observed in participial constructions by positing the same coreference
mechanism for finite clauses and for a class of participial clauses. The mechanism
relies on semantic-pragmatic properties of the dative experiencer argument, which
characterize it as a non-canonical subject.

---

14. Compare the observations, about a similar phenomenon, in Kühner-Gerth (1904, p. 108),
suggesting an explanation along these lines: a participial construction in the nominative, found
in place of an expected genitive absolute construction, is explained as the attempt to highlight
the role of the "logical" subject.

Acknowledgements

Abbreviations

AOR	aorist	IND	indicative	PRS	present
DAT	dative	INF	infinitive	PTC	particle
DEM	demonstrative	M	masculine	PTCP	participle
DET	determiner	N	neuter	RELF	reflexive
F	feminine	NOM	nominative	SG	singular
FUT	future	OPT	optative mood	1	first person
GEN	genitive	PL	plural	2	second person
IPFV	imperfective	PLUSPFV	plusquamperfect	3	third person

References

Aikhenvald, A. Y., Dixon, R. M. V. & M. Onishi. (2001). (Eds.), *Non-canonical Marking of Subjects and Objects*. Amsterdam/Philadelphia: Benjamins. https://doi.org/10.1075/tsl.46

Andrews, A. (1976). The VP complement analysis in Modern Icelandic. *Proceedings of the North Eastern Linguistic Society*, 6, 1–21.

Ariel, M. (1988). Referring and accessibility. *Journal of Linguistics* 24(1), 65–87. https://doi.org/10.1017/S0022226700011567

Ariel, M. (2001). Accessibility theory. An overview. In T. J. M. Sanders, J. Schilperoord, W. Spooren (Eds.), *Text Representation: Linguistic and psycholinguistic aspects* (pp. 29–87). Amsterdam: Benjamins. https://doi.org/10.1075/hcp.8.04ari

Barðdal, J. (2000). Oblique subjects in Old Scandinavian. *NOWELE* 37, 25–51. https://doi.org/10.1075/nowele.37.02bar

Barðdal, J. & Eythórsson, Th. (2009). The origin of the oblique subject construction: an Indo-European comparison. In V. Bubenik, J. Hewson & S. Rose (Eds.), *Grammatical Change in Indo-European Languages* (pp. 179–193). Amsterdam: Benjamins. https://doi.org/10.1075/cilt.305.19bar

Barðdal, J. & Eythórsson, Th. (2018). What is a Subject: The Nature and Validity of Subject Tests. In J. Barðdal, N. Pat-El & S. M. Carey (Eds.), *Non-Canonically Case-Marked Subjects*. The Reykjavík-Eyjafjallajökull papers (pp. 257–274). Amsterdam: Benjamins.

Barðdal, J. & Smitherman, Th. (2013). The Quest for Cognates: A Reconstruction of Oblique Subject Constructions in Proto-Indo-European. *Language Dynamics and Change*, 3(1), 28–67.

Barðdal, J., Smitherman, Th., Bjarnadóttir, V., Danesi, S., Jenset, G. B. & McGillivray, B. (2012). Reconstructing constructional semantics. The dative subject construction in Old Norse-Icelandic, Latin, Ancient Greek, Old Russian and Old Lithuanian. *Studies in Language*, 36(3), 511–547. https://doi.org/10.1075/sl.36.3.03bar

Bary, C. & Haug, D. T. T. (2011). Temporal anaphora across and inside sentences: The function of participles. *Semantics and Pragmatics*, 4, 1–56. https://doi.org/10.3765/sp.4.8

Buijs, M. (2013). Participle. In G. K. Giannakis (Ed.), *Encyclopedia of Ancient Greek Language and Linguistics*, Leiden: Brill.

Conti, L. (2008). Synchronie und Diachronie des altgriechischen Genitivs als Semisubjekt. *Historische Sprachforschung*, 121, 94–113.

Conti, L. (2009). Weiteres zum Genitiv als Semisubjekt im Altgriechischen: Analyse des Kasus bei impersonalen Konstruktionen. *Historische Sprachforschung*, 122, 182–207.

Cooper, G. L. (1988). *Attic Greek Prose Syntax*. 2 Volumes. After K. W. Krüger. Ann Arbor: University of Michigan Press.

Crespo, E., Conti, L., Maquieira, H. (2003). *Sintaxis del Griego Clásico*. Madrid: Gredos.

Croft, W. (1990). *Typology and Universals*. Cambridge: Cambridge University Press.

Dahl, E. & Fedriani, C. (2012). The argument structure of experience: experiential constructions in Early Vedic, Homeric Greek and Early Latin. *Transactions of the Philological Society*, 110(3), 342–362. https://doi.org/10.1111/j.1467-968X.2012.01313.x

Danesi, D., Johnson, C. A. & Barðdal, J. (2018). Where Does the Modality of Ancient Greek Modal Verbs Come From? The Relation between Modality and Oblique Case Marking. *Journal of Greek Linguistics* 18, 45–92. https://doi.org/10.1163/15699846-01801005

Dimitriadis, A. (1996). When Pro-Drop Languages Don't. Overt Pronominal Subjects and Pragmatic Inference. In L. Dobrin, K. Singer & L. McNair (Eds.), *Proceedings of the 32nd Meeting of the Chicago Linguistics Society* (pp. 33–47).

Dryer, M. S. (1997). Are grammatical relations universal? In J. L. Bybee, J. Haiman, S. A. Thompson (Eds.), *Essays on Language Function and Language Type: Dedicated to T. Givón* (pp. 115–143). Amsterdam: Benjamins. https://doi.org/10.1075/z.82.09dry

Eythórsson, Th. & Barðdal, J. (2005). Oblique subjects: a common Germanic inheritance. *Language* 81, 824–881. https://doi.org/10.1353/lan.2005.0173

Fedriani, C. (2009). The "Behaviour-Before-Coding" Principle: Further Evidence from Latin. *Archivio Glottologico Italiano* 94(2), 156–184.

Fedriani, C. (2014). *Experiential Constructions in Latin*. Leiden: Brill.

Frascarelli, M. (2007). Subjects, topics and the interpretation of referential *pro*: An interface approach to the linking of (null) pronouns. *Natural Language and Linguistic Theory* 25(4), 691–734. https://doi.org/10.1007/s11049-007-9025-x

Givón, T. (1983). Topic continuity in discourse: An introduction. In T. Givón (Ed.), *Topic Continuity in Discourse. A quantitative cross-language study* (pp. 1–41). Amsterdam: Benjamins. https://doi.org/10.1075/tsl.3.01giv

Goldstein, D. (2016). *Classical Greek Syntax. Wackernagel's Law in Herodotus*. Leiden: Brill.

Haspelmath, M. (2001). Non-canonical marking of core arguments in European languages. In Aikhenvald et al. (Eds.), pp. 53–83. https://doi.org/10.1075/tsl.46.04has

Haug, D. T. T. (2010). *PROIEL Guidelines for Annotation*. Ms., University of Oslo, http://folk.uio. no/daghaug/syntactic_guidelines.pdf

Haug, D. T. T. (2017). Backward control in Ancient Greek and Latin participial adjuncts. *Natural Language and Linguistic Theory* 35(1), 99–159. https://doi.org/10.1007/s11049-016-9339-7

Hock, H. H. (1990). Oblique subjects in Sanskrit? In M. K. Verma & K. P. Mohanan (Eds.), *Experiencer Subjects in South Asian Languages* (pp. 119–139). Stanford: CSLI Publication.

Humbert, J. (1960). *Syntaxe grecque*. 3rd. ed. Paris: Klincksieck.

Jøhndal, M. L. (2012). Non-finiteness in Latin. PhD Dissertation, Cambridge University.

Keenan, E. L. (1976). Towards a universal definition of "Subject of". In C. N. Li (Ed.), *Subject and Topic* (pp. 303–333). New York: Academic Press.

Keydana, G. (1997). *Absolute Konstruktionen in altindogermanischen Sprachen*. Göttingen: Vandenhoeck & Ruprecht.

Kiparsky, P. (2012). Greek Anaphora in Cross-Linguistic Perspective. *Journal of Greek Linguistics*, 12, 84–117. https://doi.org/10.1163/156658412X649977

Klaiman, M. H. (1980). Bengali Dative Subjects. *Lingua* 51(4), 275–295. https://doi.org/10.1016/0024-3841(80)90096-0

Kühner, R. & Gerth, B. (1898; 1904). *Ausführliche Grammatik der Griechischen Sprache*³, II-1; II–2. Hannover/Leipzig: Hannsche Buchhandlung.

Le Mair, E., Johnson, C. A., Frotscher, M., Eythórsson, Th. & Barðdal, J. (2017). Position as a behavioral property of subjects. The case of Old Irish. *Indogermanische Forschungen* 122(1), 111–142.

Lühr, R. (2011). Zur Validität linguistischer Theorien in der Indogermanistik. In Th. Krisch, Th. Lindner (Eds.), *Indogermanistik und Linguistik im Dialog. Akten der XIII. Fachtagung der Indogermanischen Gesellschaft* (pp. 321–330). Wiesbaden: Reichert.

Luraghi, S. (2010). Experiencer predicates in Hittite. In R. I. Kim, E. Riecken, N. Oettinger & M. J. Weiss (Eds.), *Ex Anatolia lux* (pp. 249–264). Ann Arbor: Beech Stave Press.

Masica, C. P. (1976). *Defining a Linguistic Area. South Asia*. Chicago: University of Chicago Press.

Nikolaeva, I. (2007) (Ed.). *Finiteness. Theoretical and Empirical Foundations*. Oxford: Oxford University Press.

Onishi, M. (2001). Non-canonically marked subjects and objects: Parameters and Properties. In Aikhenvald, A. Y. et al. (Eds.), pp. 1–51. https://doi.org/10.1075/tsl.46.03oni

Peels, S. (2007). ἑωυτόν as a long-distance anaphor in Herodotus' histories. Master Thesis, Leiden University.

Petit, D. (1999). **Sṷe- en grec ancien : la famille du pronom réfléchi. Linguistique grecque et comparaison indo-européenne*. Louvain: Peeters.

Pieroni, S. (2007). Soggetto e riflessivo. In N. La Fauci & S. Pieroni (Eds.), *Morfosintassi latina. Punti di vista* (pp. 27–39). Pisa: ETS.

Pires, A. (2006). *The minimalist syntax of defective domains : gerunds and infinitives*. Amsterdam: Benjamins. https://doi.org/10.1075/la.98

Powell, J. E. (1933). Studies on the Greek reflexive – Herodotus. *The Classical Quarterly* 27, 208–221. https://doi.org/10.1017/S0009838800016980

Powell, J. E. (1934). Studies on the Greek Reflexive – Thucydides. *The Classical Quarterly* 28, 159–174. https://doi.org/10.1017/S0009838800019996

Seržant, I. A. & Kulikov, L. (2013). (Eds.). *The Diachronic Typology of Non-Canonical Subjects*, Amsterdam: Benjamins. https://doi.org/10.1075/slcs.140

Sevdali, Ch. (2013). Case transmission beyond control and the role of Person. *Journal of Historical Syntax*, 24, 1–52.

Siewierska, A. (2004). *Person*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511812729

Torregrossa, J., Bongartz, C. & Tsimpli, I. (2015). Testing accessibility: A cross-linguistic comparison of the syntax of referring expressions. Extended abstract for the *Proceedings of the 89th Annual Meeting of the Linguistic Society of America*, https://journals.linguisticsociety.org/proceedings/index.php/ExtendedAbs/article/view/3046

Tronci, L. (2012). Sur le réfléchi en grec ancien: notes préliminaires. In Dupraz, E. (Ed.), *Anaphore et anaphoriques: variété des langues, variété des emplois* (pp. 151–166). Mont-Saint-Aignan: Publications des Universités de Rouen et du Havre.

Tsimpli, I. M., Sorace, A., Heycock, C. & Filiaci, F. (2004). First language attrition and syntactic subjects: A study of Greek and Italian near-native speakers of English. *International Journal of Bilingualism*, 8, 257–277. https://doi.org/10.1177/13670069040080030601

Verbeke, S., Kulikov, L. & Willems, K. (2015). Oblique case-marking in Indo-Aryan experiencer constructions: historical roots and synchronic variation. *Lingua* 163, 23–39. https://doi.org/10.1016/j.lingua.2015.05.009

Parallel syncretism in early Indo-European

Steve Rapaport

University of Edinburgh

Oblique-oblique case mergers are now known to be exceedingly rare outside Indo-European, yet very common within it (Baerman, Brown, and Corbett, 2001, 2005). Explanations of these mergers given after 2001 now need to include factors that are, individually or collectively, unique to Indo-European. In this paper, a set of such Indo-European specific factors is proposed, mainly phonological and prosodic, and its implications explored, particularly for the Germanic, Italic and Celtic families. We begin with Proto-Indo-European accent and ablaut morphology and PIE's existing syncretism of the ablative case. Timelines and examples in Proto-Germanic help flesh out the explanation.

Keywords: PIE, stress, syncretism, case mergers, ablaut, historical phonology

1. Introduction

Proto-Indo-European had 8 noun cases, and most modern Indo-European languages have fewer. The loss of these cases began in prehistoric times, with Proto-Germanic and Proto-Italic merging their locative, dative, and ablative cases, while Proto-Balto-Slavic and Proto-Hellenic merged their ablative and genitive cases. What all of these have in common is a tendency to merge oblique cases with other oblique cases.

Recently it has been shown (Baerman, Brown, & Corbett, 2001, 2005) that the early tendency of Indo-European languages to merge oblique noun cases with other oblique noun cases is typologically rare, and quite specific to Indo-European (IE) families. The Baerman et al. survey of 30 genetically diverse languages and later their broader WALS contributions (Baerman & Brown, 2013) of around 200 diverse languages show that other languages that merge cases tend to merge core cases[1] (nominative and accusative or ergative and absolute) with other core cases

---

1. *Core case*: one of accusative or nominative (or ergative or absolute). *Oblique case*: any case that is not a core case.

(Baerman's *type 1* merger), or occasionally merge the marked core case (often accusative) with oblique cases (e.g. dative, instrumental, ablative, genitive) (Baerman's *type 2* merger). Oblique-oblique (Baerman's *type 3* mergers) of the type found in IE are, according to this survey, "exceedingly sparse outside of Indo-European" languages (Baerman, 2005, p. 52).

In their chapter on case syncretism in WALS Online (Baerman & Brown, 2013), the authors include 198 languages (the WALS 200 language sample): 65 of these have three or more distinct noun cases. Of those 65, 18 have syncretisms marked *type 1*, 22 marked *type 2*, and none marked *type 3*.[2] There are some Indo-European languages included, but only modern examples showing *type 1* and *type 2* syncretism. Baerman and Brown conclude that "the elaborate networks of semantic relationships that have been sometimes constructed to account for syncretic patterns in Indo-European find little direct support from other language families".

Yet mergers between pairs of the ablative, dative, genitive, instrumental and locative cases in early Indo-European nominative paradigms (Proto-Germanic, Proto-Italic, Proto-Celtic, Proto-Balto-Slavic, Proto-Hellenic) are common. There are only two, partial, oblique-oblique case syncretisms present in the Proto-Indo-European paradigm itself. If the other case syncretisms are not driven by semantic relationships, then they must be driven by something else, something common to Indo-European languages but rare elsewhere. Semantic similarity or metaphor of cases tends to be more or less the same everywhere. This paper therefore seeks a historical explanation of typologically rare oblique-oblique mergers that will apply to Indo-European families displaying them, but will not apply equally well outside Indo-European. The explanation that works is simply a combination of sound change, analogy, and an initial prosodic grouping of the oblique cases.

In what follows, I will first review the historical morphology of PIE and its pitch-accent and ablaut grouping of the oblique cases separate from the core cases. I will then review how sound changes and a shift to stress accent in some of these proto-languages made the oblique cases so different from the core cases as to disrupt the paradigm. Mergers of final vowel sounds and levelling by analogy happen in several early IE languages at around the same time and account for many of the early oblique-oblique mergers.

---

2.   In response to the query of one of the anonymous reviewers as to whether BB&C tested languages with similarly complex nominal morphologies, my response is yes: I found the following results in WALS combining cases and syncretism (https://wals.info/combinations/49A_28A#2/16.7/148.9): 5 languages with 6–9 cases had type-1 (core cases only) syncretism (Wambaya, Yidiny, Mangarrayi, Yukaghir, Greenlandic), and 5 languages with 6–9 cases had type-2 syncretism (Burushaski, Nenets, Krongo, Araona, and Ngiyambaa), along with 4 additional Indo-European examples.

The existing syncretisms in PIE of the ablative case with the dative plural and genitive singular cases plus analogical levelling clearly help drive many of the mergers at this early stage. These two factors together, along with phonological changes, will be used in this paper to explain oblique-oblique mergers in PIE.

2.   How the core and oblique cases became distinct in early IE

I claim that two specifically PIE-related factors are responsible for the early oblique-oblique mergers, a phonological separation of the oblique and core cases (Factor I), and the existence of split syncretism dating back to PIE (Factor II). These factors, to be explored in some detail (see Table 5 for summary), are still not sufficient to create a similar syncretism in any other, non-Indo-European language. There must be other PIE-specific and morphology-specific factors which are also important. Among these is the fact that PIE was predisposed to these factors by its own morphology at the time: PIE was a highly synthetic-fusional language, with noun inflection simultaneously indicating case, gender, and number. At the same time, it had a clear distinction between words, rather than a root-inflection clumping of entire phrases as is often found in polysynthetic languages. There was a strict limit of one accented syllable per word. PIE had inflectional morphology depending on both accent and ablaut, and several marked phonemes (three laryngeals, four syllabic sonorants) which were essential to that ablaut system. If one wished to replicate the early PIE case system collapse in another language, one would need to start with a system of a similar type, as well as similar specific morpho-phonological grouping of the oblique cases, and the existing partial syncretism linking three oblique cases.

Looking first at the grouped oblique cases, the Proto-Indo-European (PIE) oblique athematic[3] noun cases are, in Fortson's terminology, the prosodically *weak* cases – the genitive, ablative, dative, instrumental and locative, as opposed to the *strong* cases (nominative, vocative, accusative)[4] differing only in ablaut and accent, with rightward movement of the accent in weak cases (Fortson, 2004, p. 103). Since Fortson's prosodic-based weak cases are almost exactly equivalent to the syntactic-/semantic-oriented oblique cases,[5] I will use the terms *weak* and *oblique*, *strong* and

---

3.   *Athematic nouns*: PIE nouns with non-vowel-terminated stems, which tended to shift accent rightward for weak cases.

4.   *Strong case*: A core case plus the vocative, as defined (phonologically) by Fortson (2004, p103). *Weak case*: any other noun case.

5.   Only two cases do not conform: the locative, which groups metrically with the oblique cases in the plural but with the core cases in the singular, and the vocative, which does not merge with anything at this stage.

core, interchangeably (Table 1). This near-merger of morphological/prosodic and syntactic/semantic case terms is, notably, unique to Indo-European languages. It is one of the IE-specific drivers to these early case mergers.

Table 1. PIE Core cases ~=strong; oblique ~=weak; (vocative, locative exceptional; initial syncretism in ablative with gen.sg and dat.pl as per Ringe (2006), Sihler (1995), Fortson (2004).)[6]

tooth: athematic, hysterokinetic[7]	singular	plural	syntax/prosodic
nominative	*$h_3$dónts*	*$h_3$dóntes*	core/strong
accusative	*$h_3$dóntm̥*	*$h_3$dóntm̥s*	core/strong
vocative	*$h_3$dónt*	*$h_3$dóntes*	oblique/strong
genitive	*$h_3$dn̥tés*	*$h_3$dn̥tóHom*	oblique/weak
ablative	*$h_3$dn̥tés*	*$h_3$dn̥tmós*	oblique/weak
dative	*$h_3$dn̥téy*	*$h_3$dn̥tmós*	oblique/weak
locative	*$h_3$dónti*	*$h_3$dn̥tsú*	oblique/mixed
instrumental	*$h_3$dn̥téh*	*$h_3$dn̥tbʰí*	oblique/weak

In PIE there seems to be just one (pitch) accent per word, either on root, stem or ending: effectively a *culminative*[7] pitch accent (similar to that of descendants Ancient Greek and Sanskrit). According to Lehmann (1985), noun roots were, in earlier PIE, originally all monosyllabic, and added accented or unaccented clitics for case, which became suffixes, often advancing the accent rightward on those nouns in those cases. Because of the accent-shifting nature of the oblique clitics, the strong (core) cases usually have first-syllable accent, while the weak cases are the cases that, in an increasing majority of nouns, have a rightward-shifted accent. This is true (as in Table 2) for the *amphikinetic*, *hysterokinetic*, and *proterokinetic* declensions. The athematic exceptions, nouns that do not shift accent rightward for the oblique cases, mostly pertain to the fourth, *acrostatic* declension, which is overtaken in time by the shifting *proterokinetic* declension nouns that do shift. (Fortson, 2004, p. 108; Ringe, 2006, p. 44). Over time, then, the regular nouns are those that shift the pitch accent rightward for the oblique cases.

Later, many daughter languages changed this pitch accent to a culminative stress accent, on the same syllable. Proto-Italic (Allen, 1973; Stuart-Smith,

---

6.   See Ringe (2006) for further details.

7.   Culminative accent: Each (phonological) word has a single strongest syllable bearing the main stress (Hayes, 1995, p. 24)

Table 2. Examples of the four PIE declensions in athematic nominals, showing ablaut and accent movement with strong and weak cases (after Ringe, 2006, p. 47, 197)

Declension	Case	Root	Stem	Ending	Example nom.sg. ~ gen.sg.	Gloss
proterokinetic	strong	é			*mén-ti-s	'thought'
	weak		é		*mn̥-téy-s	'of thought'
hysterokinetic	strong		é		*ph₂-tḗr	'father'
	weak			é	*ph₂-tr-és	'of father'
amphikinetic	strong	é			*léy-mō	'lake'
	weak			é	*li-mn-és	'of lake'
acrostatic	strong	ó			*nókʷ-t-s	'night'
	weak	é			*nékʷ-t-s	'of night'

2004), Proto-Germanic (W. Lehmann, 1961), and Proto-Celtic (Thurneysen, 1946, sec. 36), for example, all developed stress accent, though their predecessor was pitch-accented. A pitch accent adds specific pitch characteristics to a vowel, whereas a stress-accent emphasises a full syllable using pitch, duration and amplitude (Allen, 1973, p. 74; Hayes, 1995, p. 49; Stuart-Smith, 2004, p. 184). Stress accent implies contrastive strong and weak spoken syllables. These in turn give rise to stress-sensitive sound changes including Verner's Law and Dybo's Law (Matasović, 2012). Oblique-case athematic nouns in PIE were the weak cases, with their right-leaning pitch accents becoming stress accents. Weak cases became, for a time, metrically different from the nominative, accusative, and vocative nouns.

2.1   How the core and oblique stems became more phonologically distinct

Among the first sound changes from PIE to each of its daughters were the loss of the laryngeals (h_1, h_2, h_3) and changes to the syllabic sonorants (r̥, l̥, m̥, n̥) where either a vowel inserted itself before or after each sonorant (e.g. in Proto-Germanic, usually epenthetic *u, e.g. PIE *$h_1$dn̥t *tooth-* > PGmc *tund-), or the sonorant was replaced by a vowel (e.g. in Sanskrit, generally becoming [a], e.g. PIE *$h_1$dn̥t- > Skt dat-) (Ringe, 2006). Note that a sonorant's new epenthetic vowel sometimes ended up in a different place from the syllable's previous nucleus, e.g. PIE *wréh₂d- > PGmc *wrōt- (*root*, strong) but *wr̥h₂d- > *wurt- (*root*, weak) (see Ringe, 2006, p. 44, for additional examples). Beyond Germanic, a similar shifting syllabic peak is also found in Celtiberian, where PIE syllabic *r̥ and *l̥ become *ri* and *li* before a stop, yet likely became *ar* and *al* before vowel or *s*, and syllabic *m̥ and *n̥ became *am* and *an* (Wodtko, 2003, Sections 2–4, 20).

The resulting *wrōt-* ~ *wurt-* -type metathesis makes the alternating forms for different cases appear less related. These changes to the stems grouped and alienated

the oblique case nouns from the core case of the same noun, making the ablaut relationships within the paradigm collapse (Table 3).[8]

Table 3. Collapse of ablaut rules due to loss of laryngeals and epenthesis or substitution of syllabic sonorants

	NOM/ACC/VOC (strong stem)	GEN/ABL/INS/DAT/ LOC (weak stem)		Strong stem (After loss of laryngeals and syllabic sonorants)	Weak stem
PIE 'tooth'	*$h_1$dónt-	*h_1dṇt-	>	PGmc *tanþ- Skt dánt-	PGmc *tund- Skt dat-
PIE 'root'	*wréh₂d-	*wṛh₂d-	>	PGmc *wrōt-	PGmc *wurt-

According to Ringe's proposed chronology, Proto-Germanic's laryngeal and syllabic sonorant sound changes all precede the first stress-sensitive sound changes (Verner's Law), which, in turn, precede the fixing of the stress on the first syllable.[9] Ringe deduces that the order of changes is the following: loss of laryngeals and changes of syllabic sonorants => shift to stress accent => stress-sensitive changes (e.g. Verner's Law) => shift to initial stress (Ringe, 2006, p. 144).

So, for example (Ringe, 2006 with my filled-in reconstructions), using the common Proto-Germanic a-stems,

1. *$h_3$émōs ~ *h_3ṃsés > *ámōs ~ *ṃsés 'shoulder ~ of shoulder' (loss of laryngeals)
2. *ámōs ~ *ṃsés > *ámōs ~ *amsés (loss of syllabics, ablaut collapses, levelling of stem)
3. *ámōs ~ *amsés > *áma- ~ *amsá- (stress accent begins, Verner's law in effect)
4. *áma- ~ *amsá- > *áma- ~ *ámsa- (stress shifts to first syllable)

Similarly, Dybo's Law is in operation in Proto-Italo-Celtic. Matasović (2012) states it as "the shortening of vowels in pre-accented position" and more specifically as:

*H > Ø after *i, *u, *Ra before a stressed syllable not beginning with a nasal

---

8. Terms and abbreviations used:

NOM	nominative	DAT	dative
ACC	accusative	GEN	genitive
VOC	vocative	PIE	Proto-Indo-European
ABL	ablative	PGmc	Proto-Germanic
INS	instrumental	Skt	Sanskrit.
LOC	locative		

9. Verner's Law applies to non-initial syllables such as *ph₂tér > *fader.

Dybo's law is, like Verner's, stress sensitive, and so there is stress accent at this stage. In Proto-Italo-Celtic, which shows evidence of Dybo's law, laryngeals are lost just before accented syllables, leaving the preceding vowel short as in PIE *g^huH-tú- 'invoked' > OIr. *gŭth* 'voice' and also PIE *wiHro-* 'man' > PCelt *wĭro-* > OIr. *fer* (Matasović, 2009). By the time of Primitive Irish, three oblique cases had been merged into the dative (Blažek, 2008). Again the ordering, shown by the effects of Dybo's law, is loss of marked sounds, then stress accent behaviour, then shift to initial stress, then merging of oblique cases. The merging of cases sometimes comes earlier, as we will see below.

In summary, the original accent/ablaut system of PIE accentually grouped noun stems in all but one declension, effectively, by core case versus oblique case. A shift to stress accent in several families then grouped the resulting words metrically, while sound changes and metrical changes eventually made some of the core and oblique stems appear unrelated. Instead of an ablaut rule to shift between core and oblique noun cases, most nouns by now had irregular, unpredictable and different pronunciations for the weak cases. The resulting unpredictable divergence of sound and accent between strong and weak cases is, I claim, the first phase of the unusual oblique-oblique case syncretism between the Proto-Indo-European and early stages of the daughter languages.

The next phase is a natural levelling of the irregularity, which proceeded differently in the various IE languages, but usually dropped one of either the core-case stems or the oblique-case stems. In the *$h_3$émōs ~ *h_3m̥sés* example above, the core-case stem remains and the obliques are replaced at stage 2.

3.   Other proposed influences leading to oblique-oblique mergers

The increasing phonological non-uniformity of the core and oblique case stems is a possible driver of the typologically unusual oblique-oblique mergers. But that driver is also unlikely to be just one feature or structure. Given that whatever types and range of structures existed in past versions of PIE must also have existed in other language families, past and present, then if we find some single feature responsible for all these early IE case mergers, it surely must exist in other languages past and present, too. Many other language families with that feature should experience oblique-oblique case mergers. Yet oblique-oblique mergers outside Indo-European are exceedingly rare. The most likely explanation is that *two or more factors combined* in Indo-European to make an unusual circumstance that would drive these mergers. The usually-mentioned candidates for drivers of case mergers include syntactic grouping, semantic grouping, and existing syncretism.

3.1   Syntactic and semantic grouping

Oblique-oblique mergers in Italic languages are often explained by the influence of semantics (Luraghi, 1987) or syntax (Covington, 1999). Covington's explanation of the merger of the ablative, locative and instrumental in Proto-Italic is that semantic-only distinctions of case are eliminated, leaving only cases that are syntactically distinct. He points out that Keenan and Comrie's noun-phrase accessibility hierarchy (Keenan & Comrie, 1977) groups the ablative, locative, and instrumental together by their syntactic identity, making all three cases equally oblique, in exact fashion. Where one case may appear in a syntax tree, so might the other two.

The Keenan and Comrie accessibility hierarchy, based on syntactic location of nouns in a sentence, is elaborated in Covington to clarify the syntactic roles of each level. It is reproduced from Covington here (Table 4).

Table 4. Covington's expanded Keenan & Comrie hierarchy

SU	Nominative	Argument of 1-place (intransitive) verbs
DO	Accusative	Additional argument of 2-place (transitive) verbs
IO	Dative	Additional argument of 3-place (ditransitive) verbs
OBL	Abl./Loc./Ins.	Noun phrase in other relationship to main verb
GEN	Genitive	Noun phrase modifying noun phrase

Covington attributes the ABL + LOC + INS merger to their syntactic grouping, and the elimination of semantic-only case distinctions, accompanied by a rise in the use of prepositions.

But this explanation is not, in the end, feasible. First, Keenan and Comrie's hierarchy of accessibility is universal, but these syntactically-grouped mergers remain typologically rare and peculiar to IE languages. Also, similar mergers that do not follow this syntactic grouping (ABL+ LOC + DAT > DAT) took place in Germanic (Ringe, 2006), and again similarly and separately (ABL + LOC + INS + DAT > DAT) in Old Irish (Rovers, 2016) suggesting that the syntactic similarity is not the only driver of these mergers.

Luraghi's (1987) semantic-spread explanations run into a similar problem. She provides an extensive look at how the semantics of the oblique cases in Italic, Germanic, Ancient Greek and Hittite languages are metaphors of each other and how they can spread in usage – a correct observation but generally universal. However, we need to find a driver that is IE-specific and is likely to target just the oblique cases.

3.2   Existing syncretism of the ablative

There are few existing case/gender/number syncretisms in reconstructed PIE, but one particularly relevant one is the ablative case, which was never distinct, even in our earliest reconstruction of PIE. This case was already syncretic (see Table 1 above) with the genitive in the singular (in athematic nouns), and with the dative in the plural (Ringe, 2006, pp. 41–42; Fortson, 2004, p. 106). We can predict that levelling will cause the ablative to merge with either the genitive or the dative over-whelmingly more frequently than chance, especially in athematic nouns. Athematic nouns were more common in the early days of Indo-European, when these case mergers were taking place. This factor is also peculiar to Proto-Indo-European.

I therefore propose that the Indo-European peculiarities that caused these widespread yet typologically-rare mergers throughout Indo-European were just the two factors listed in Table 5:

Table 5. Two necessary factors to explain early oblique-oblique syncretisms in Indo-European languages

Factor I. Phonological split of cases
PIE sound changes cause core and oblique cases to diverge, at first in ways that allow for recognition of the relationship between core and oblique (see Table 1).
After loss of laryngeals and epenthesis or substitution of syllabic sonorants, the ablaut rules cease to make sense.
Since oblique cases are accented later in the word while core and vocative cases are accented earlier, stress-sensitive sound changes will enhance differences between core and oblique cases and will further separate the core and oblique stems.
Factor II. Pre-existing syncretism
Because the PIE ablative is already syncretic with the genitive in the singular and with the dative in the plural, levelling will sometimes fully merge ablative case with either the dative or the genitive case.

We will look at what this proposal predicts, then see if the predictions are borne out.

Predictions (assuming these two factors drive the mergers)

1. Upon loss of the accent/ablaut system, we may see merger of the phonologically similar ablative, locative, instrumental, dative and genitive stems, or replace-ment of these stems by the core case stems, along with levelling.

2. We would also expect merging (by levelling) of the ablative with the already partly-merged dative or genitive, depending on if the levelling follows the sin-gular or plural part of the paradigm.

3. The vocative, though technically oblique, will never merge with another oblique case, since the vocative is grouped phonologically and prosodically with the core cases. We would predict zero mergers at this stage between oblique and core cases.

This is, in fact, precisely what we do see. In Table 6, each early IE language is shown at the approximate point in time of its first oblique-oblique merger. The actual timeframe is different within each language, as each developed these changes at a different pace.

Table 6. Case-merging in Early IE languages, compared with prosody change

Old IE Language	Pitch accent > stress accent	Accent fixed to first syllable	Cases merged
Vedic Sanskrit	No	No	None[a]
Prakrit (fem.)[b]	Yes[c]	No	LOC+INS+ABL+DAT+GEN>INS
P.-Balto-Slavic	Both[d]	No	ABL+GEN > GEN[e]
Proto-Hellenic	No	No	ABL+GEN > GEN[f]
Proto-Celtic	Yes[g]	Yes[h]	None[i]
but Goidelic	Yes	Yes	LOC+INS+ABL+DAT > DAT[j]
Proto-Italic	Yes	Yes	LOC+ABL+INS >ABL[k]
Proto-Germanic	Yes[l]	Yes	LOC+ABL+DAT > DAT[m]

a. Only syllabic *r* is present on the surface, while syllabic *m̥* and *n̥* surface as *a* in Vedic, according to Steriade (1988). Thanks to reviewer 1 for correcting this detail.
b. The fact that feminine nouns merged cases before masculine ones in Prakrit is more evidence in favour of purely phonological, not semantic or syntactic, merger.
c. Schwarzschild (1956, p. 183) attributes the merger of all these oblique cases in Pali and Prakrit feminine nouns to sound change and analogy, "helped by a certain amount of syntactic liberty".
d. Long vowels in PBS acquired a new sort of pitch accent, while the former pitch accent became a stress accent.
e. Proto-Balto-Slavic (Kortlandt, 2010)
f. Proto-Greek (Sihler, 1995)
g. As with Proto-Germanic's Verner's Law, the stress-sensitive nature of Dybo's Law in Proto-Celtic (Matasović 2012) suggests that stress was by now sufficient to consider some syllables "unstressed".
h. Accent becomes first syllable after Dybo's Law and postconsonantal laryngeal loss before pretonic high vowels, but before the Proto-Italic sound changes mentioned by Schrijver (2015)
i. Blažek (2008, sec. 4.2) shows at least traces of all 8 PIE noun cases (possibly excepting instrumental) in the o-stem (athematic) nouns in Celtiberian, though Gaulish may have lost locative and ablative and Lepontic seems to have lost locative, ablative and instrumental.
j. From Blažek (2008, sec. 4.2), Goidelic u-stems, r-stems and n-stems preserve late PIE case endings phonologically but still lose Loc. Instr. and Abl. cases entirely, though some are preserved intact in Celtiberian
k. Proto-Italic (Sihler, 1995)
l. The existence of Verner's Law indicates that some syllables are "unstressed".
m. Proto-Germanic (Ringe, 2006)

The ablative and genitive do indeed merge via levelling of the singular to the plural in Balto-Slavic and Hellenic, but *before* pitch accent is lost.[10] This merger is therefore not reliant on factor I (phonology and accent) but may still be driven by factor II (the existing syncretism of the ablative and genitive singular, in this case).

---

10. A reviewer points out that Greek keeps its pitch accent beyond the point of further case mergers. Further work to come on this, but for now this is still consistent with Factor II.

The shift from pitch to stress appears to happen before the other obliques begin merging in the Proto-Italic, Proto-Germanic, Primitive Irish, and Middle Indo-Aryan. But further investigation of the timeline of sound changes (Ringe, 2006, pp. 104–105) shows that, at least in Germanic, the oblique and core noun stems sound quite unpredictably different in all these families by the time the crucial accent shift from pitch to stress occurs. Factor I, the phonological split of cases, can be seen as grouping the oblique cases as phonologically separate from the core cases.

The next change is even more momentous: once a language becomes stress-sensitive, there seems to be a strong tendency in early Indo-European languages to shift the stress to the first syllable.[11] This shift happens shortly after the change to stress accent in Proto-Germanic, Proto-Italic, and Proto-Celtic, and even Thessalian, with evidence from Dybo's Law and Verner's Law left behind to show that sound changes happened after the change to stress accent. The shift of stress to the first syllable is, in Lehmann's (1961) view, the defining factor in the inception of Proto-Germanic.

The changes shown in Table 6 occur (roughly) in left-to-right order in PIE. Note that all mergers follow the predictions above, and that the further along in the sound and prosodic changes each family is, the more cases there are that merge.

Never does the vocative merge with an oblique case, though it is, arguably, an oblique case. (The vocative is not an argument of a verb.) It either remains separate (as in Gaelic) or later merges with a core case (as in Early Latin and some Germanic languages) exactly as predicted by factor I, since the vocative shares prosody with the core cases.

There is an obvious correlation between accent grouping and the PIE daughter-language case mergers: case merger happens once either stress accent or fixed first-syllable stress occurs. Proto-Balto-Slavic and Proto-Hellenic (Table 6) do not (yet) have either of these changes, and show, at most, a merger of the ablative with the genitive, which is a levelling by analogy of the existing singular number syncretism in PIE, as predicted by factor II.

4.   Examples of case mergers

If these oblique-oblique mergers are mostly governed by phonological principles and analogical levelling, why is PIE so prone to them? This proclivity is, I claim, due to the combination of existing syncretisms and oblique endings that can merge phonologically in ways which tend to cascade like dominoes.

---

11.   Even the Greek dialects, with pitch accents in the last three syllables as per the "law of limitation", included Thessalian, which appears (Probert, 2008, p. 73) to have shifted from pitch to stress accent and moved to initial-syllable stress.

The PIE thematic dative and locative endings tend to syncretize as *-i* due to sound change plus stress shift, with athematic dative **-ey* and locative **-éy*, and dative **-ōi* and locative **-oi* in thematic stems. Diachronically, a raising rule turns *[ey]* to *[i]* while a stress shift can easily change *[-ōi]* to *[-oī]*. In Proto-Germanic, unaccented *[ey]* raised to *[i]*, making the dative singular the same as the locative singular (Ringe, 2006, p. 122). In other languages the change was similar: for athematic stems, PIE dative **-ey* > PGmc *-i*, OCS *-i*, Koine Gk. *-i*, while PIE locative **-éy* > PGmc *ī*, Koine Gk. *-í*, Ved. *-í*. Even Hittite shows dative *šiun-i* 'for the deity' and locative *nēpiš-i* 'in heaven'. Cf. (Lundquist & Yates, 2018, p. 5) for details.

The ablative, dative, and locative merge with a combination of pre-existing syncretism, phonological merger and levelling: in Proto-Germanic (for example), the dative and locative singular merge phonologically, while the ablative singular levels to match the -ī ending from the dative and locative. The locative plural takes on the ending of the (already identical) ablative and dative plural, to complete the merger. Consider the adjective PIE **gʷr̥h₂éwih₂* > PGmc **kuruz* 'heavy' (Ringe, 2006, p. 122) (Table 7 and 8).

Table 7. Part of paradigm for PIE **gʷréh₂us* 'heavy'.
(Athematic, proterokinetic. feminine adjective)

Feminine	Singular	Plural
ablative	**gʷr̥h₂uyéh₂s*	**gʷr̥h₂uyéh₂mos*
dative	**gʷr̥h₂uyéh₂ey*	**gʷr̥h₂uyéh₂mos*
locative	**gʷr̥h₂uyéh₂i*	**gʷr̥h₂uyéh₂su*

This becomes PGmc **kuruz*, and the ablative and locative forms join the dative (Table 8).

Table 8. Part of paradigm for PGmc **kuruz* 'heavy'. (Feminine adjective)

Feminine	Singular	Plural
dative (+abl+loc)	**kurjī*	**kurjaimaz*

In another example in Prakrit, the genitive, already syncretic in the singular with the ablative, can merge by levelling, and does so in Prakrit as the accent shifts to initial, away from the endings.

The exceptions to these phonological mergers of the oblique cases can tell us as much as the more common tendencies. IE languages that do not merge the instrumental with the dative and locative may also be keeping them separate for phonological reasons. The athematic instrumental PIE **-eh₁* (following (Lundquist & Yates, 2018, p. 5) is not prone at this point to phonological mergers with the merged

dative and locative *-i*. It is sometimes *-ī*, as in PGmc **kurjī* above, but generally remains instead as accented or lengthened *-ā* or *-ē* for some time, as in Ved. *pad-ā́* 'with the foot', *mánas-ā* 'with thought', or OAv. *zərədā(-cā)* 'with heart', *manaŋh-ā* 'with thought' (< PIE **men-es-eh₁*). The instrumental singular is sometimes lost as a noun case but fossilized into adverbial forms, as in Latin *vald-ē* 'very', Greek (Elean) *taut-ē* 'here'.

In future work, I plan to explore the loss of the instrumental further and show how similar phonological and prosody-driven mergers occurred in the oblique case endings in Goidelic, Prakrit, and Ancient Greek.

5.   Conclusion

I believe this paper provides evidence that sound change, prosodic grouping, and existing syncretism may, alone, be sufficient to drive the typologically unusual oblique-oblique case syncretism in early Indo-European languages. Future research may show a more specific timeline for this influence in individual daughters, and similar phonological drivers that also show parallel similarities between families, for later Indo-European noun case mergers.

Acknowledgements

Abbreviations

ABL	ablative	Oav.	Old Avestan
ACC	accusative	OCS	Old Church Slavonic
PL	plural	OIr.	Old Irish
DAT	dative	Pcelt	Proto-Italo-Celtic
GEN	genitive	PIE	Proto-Indo-European
INS	instrumental	PGmc	Proto-Germanic
LOC	locative	SG	singular
Gk.	Greek	Ved.	Vedic
IE	Indo-European	VOC	vocative
NOM	nominative		

References

Allen, W. S. (1973). *Accent and rhythm* (1st ed.). Cambridge: Cambridge University Press.

Baerman, M. & Brown, D. (2013). Case syncretism. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from http://wals.info/chapter/28

Baerman, M., Brown, D. & Corbett, G. (2005). *The syntax–morphology interface: A study of syncretism* (1st ed.). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486234

Baerman, M., Brown, D. & Corbett, G. G. (2001). *Case syncretism in and out of Indo-European*. Chicago: Chicago Linguistic Society.

Blažek, V. (2008). Gaulish language. *Studia minora Facultatis Philosophicae Universitatis Brunensis. Graeco-Latina Brunensia* vol. 57, issue 13, pp. 37–65. Brno: Brněnské Univerzity.

Covington, M. (1999). *Why the ablative, locative, and instrumental cases fell together in Latin* (pp. 1–5). University of Georgia. https://pdfs.semanticscholar.org/6ced/f5a88197c98ccbcabac-c822f0445783921d6.pdf (Retrieved April 5, 2015 from Semantic Scholar website)

Fortson, B. (2004). *Indo-European language and culture: An introduction*. Oxford: Blackwell.

Hayes, B. (1995). *Metrical Stress Theory: Principles and case studies*. Chicago: University of Chicago Press.

Kortlandt, F. (2010). Indo-European morphosyntax. In *Studies in Germanic, Indo-European and Indo-Uralic* (pp. 81–104). Amsterdam: Rodopi.

Lehmann, W. P. (1961). A definition of Proto-Germanic: A study in the chronological delimitation of languages. *Language*, 37(1), 67–74. https://doi.org/10.2307/1938717

Lehmann, W. P. (1985). Proto-Indo-European derivational morphology in chronological perspective. *Oceanic Linguistics Special Publications*, 20 (1985), 67–76.

Lundquist, J. & Yates, A. (2018). The morphology of Proto-Indo-European. In J. Klein et al. (Eds.), *Handbook of Comparative and Historical Indo-European Linguistics*, 1–85. Berlin/New York: de Gruyter.

Luraghi, S. (1987). Patterns of case syncretism in Indo-European languages. *Papers from the Seventh International Conference on Historical Linguistics*, 48, 355–371. Philadelphia: Benjamins. https://doi.org/10.1075/cilt.48.28lur

Matasović, R. (2009). *Etymological dictionary of proto-Celtic*. Leiden & Boston: Brill.

Matasović, R. (2012). Dybo's law in Proto-Celtic. *Zeitschift für celtische Philologie*, 59, 129–141. https://doi.org/10.1515/zcph.2012.008

Probert, P. (2008). *Ancient Greek accentuation: Synchronic patterns, frequency effects, and prehistory* (Vol. 1). Oxford: Oxford University Press.

Ringe, D. (2006). *Linguistic history of English, Vol 1. From Proto-Indo-European to Proto-Germanic*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199284139.001.0001

Rovers, J. (2016). Case merger in Indo-European and the independent datives in Old Irish, 1–65. (Unpublished Master's Thesis). University of Utrecht Archives, Utrecht, Netherlands.

Schwarzschild, L. A. (1956). Notes on the declension of feminine nouns in Middle Indo-Aryan. *Journal of the Royal Asiatic Society of Great Britain & Ireland*, 88(3–4), 181–190. https://doi.org/10.1017/S0035869X00115035

Sihler, A. L. (1995). *New comparative grammar of Greek and Latin*. Oxford: Oxford University Press.

Stuart-Smith, J. (2004). *Phonetics and philology: Sound change in Italic*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199257737.001.0001

Thurneysen. (1946). *A grammar of Old Irish (Revised)*. Dublin: Dublin Institute for Advanced Studies.

Wodtko, D. S. (2003). *An outline of Celtiberian grammar*. Freiburg: Albert-Ludwigs-Universität Freiburg.

Dative possessor in ditransitive Spanish predication, in diachronic perspective

Rosa María Ortiz Ciscomani
Universidad de Sonora

This paper analyses the diachronic behavior (12th to 20th centuries) of the ditransitive predications, with Direct Object (DO) and Indirect Object (IO), coded by Noun Phrases in Spanish. The analysis accounts for the properties of those participants, possessor and possessed respectively, and the verbs in the constructions. The analytic approach assumes that transitivity is a scalar phenomenon, without distinction between actantial and non-actantial dative. The *possessive dative* is an IO that refers to the possessor of the DO, while the ditransitive construction is an extension of the prototypical ditransitive. Our claim is that the construction (i) gives relevance to the possessor over the possessed, and (ii) shows that diachronically the DO has lexically diversified from concrete to non-concrete and animate entities, especially from the 16th century onwards.

Keywords: Spanish, possessive dative, ditransitive construction, diachronic perspective, IO

1.   Introduction

Like many other languages, Spanish expresses possession through various devices or strategies. Typologically, several types of possession exist, among them adnominal possession, where both possessor and possessed are coded in the same Noun Phrase (1); possessive predication, characterized by the presence of the possessive verb *tener* 'to have' (2); and external possession, in which the relation is coded without a possessive verb and without the concurrence of possessor and possessed in a Noun Phrase (3a–c):

(1)   [El auto *de María*] es nuevo.
      [*María's*-Possessor] [car-Possessed] is new.

(2)   [*María*] tiene [una casa nueva]
      [*María*-Possessor] has [a new house-Possessed]

(3)   a.   *Le*$_i$ lastimaron [la mano] *a María*$_i$
            Her$_i$ (they) hurt [the hand-Possessed] [*to Mary*$_i$-Possessor] = *Mary's hand*
      b.   Los soldados *le*$_i$ mataron a[l hijo] [*a María*$_i$]
            The soldiers her$_i$ killed [the son-Possessed] [*to Mary*$_i$- Possessor]
      c.   Los ladrones *le*$_i$ destruyeron [la casa] [*a María*$_i$]
            The thieves her$_i$ destroyed [the house-Possessed] [*to Mary*$_i$-Possessor]

In this paper I analyze the diachronic evolution of the ditransitive predications in which both objects, direct and indirect, are coded by Noun Phrases like the ones exemplified in (3).

These are examples of the so called *external possession* (Payne y Barshi, 1999, p. 3), structures with a *possessive dative*, that is, a constituent, not required by the argument structure (Payne y Barshi, 1999, p. 3) that expresses the possessor of the direct object (Alcina y Blecua, 2001/1975, § 7.2.13; RAE-Asale, 2009: § 35.7f, 2010: §35.4.1b), an entity conceived of as inalienable. In this work I assume that the possessor *to Mary* 'a María', co-referential with the clitic *le*$_i$ in (3) is codified as an argument of the construction, whereas the possessed is codified by another argument, the direct objects *the hand*, *the son*, and *the house* in (3a), (3b), and (3c), respectively.

The corpus is made up of data from 12th- to 20th-century Spanish. It includes all the ditransitive structures with possessive dative occurring in a quantitatively homogeneous sample of the texts (same number of words per text cited in *References*: §a.). The analysis shows that the ditransitive construction with possessive dative is documented in all centuries, providing information about the properties of predication, verbs, and possessor and possessed entities.

The analytic approach assumes that transitivity is a scalar phenomenon (Hopper and Thompson, 1980), that is, it recognizes high transitivity with prototypically transitive structures or those that represent the category well, alongside low transitivity structures with prototypical extensions. I assume, too, that the distinction between the actantial dative and the non-actantial dative is irrelevant. From my perspective, the ditransitive construction with a possessive dative is an extension of the Spanish Ditransitive Construction (Ortiz Ciscomani, 2006 and 2011), assuming that a *construction* is a pattern with form-meaning correspondence that exists independently of particular verbs (Goldberg, 1995, p. 1). In other words, according to my approach, the ditransitive structure with a possessive dative is an extension of the ditransitive skeletal or abstract pattern (Goldberg, 1995) that shares not only the syntactic configuration but also the basic meaning of the pattern ('to cause to move an entity to another location, a human location'). This syntactic and semantic correspondence is extended to other cause-motion structures (even those of other, distinct verb classes), forming a family of related senses (Goldberg, 1995, p. 66).

The aim of the paper is twofold: I hope to demonstrate that the possessive dative ditransitive structure:

i.   has existed since the 12th century, and that the possessed entity has diversified diachronically from concrete to non-concrete and animate entities, and
ii. has accommodated or adapted to the basic Ditransitive Construction to mirror the relevance of the possessor over the possessed (Company, 2001, Ortiz, 2006, 2011), fulfilling a specific communicative need, that of giving relevance to the possessor (Company, 2001, Ortiz, 2006, 2011).

2.   Possessive Dative in Latin

In Latin, possession can be expressed syntactically by means of a dative, as in (4), or a genitive, as in (5).

(4)   *Mihi*   *est*   *liber*
      DAT.1SG   be.PRS.3SG book.NOM.SG
      to/for me   be   the book
      'The book is mine'

(5)   *Dom-us*   *est*   *reg-is*
      house-NOM be.PRS.3SG king-GEN
      'The house is the king's'

According to Ernout and Thomas (1951/1964, § 9 and 53), possessive dative and genitive could appear in sentences with the verb *esse*, 'to be', as in (6a) and (6b), respectively:

(6)   a.   *est*   *part-i*   *me-o*   *dom-us*
           be.PRS.3SG father-DAT   mine-DAT house-NOM
           be   to my father   a house
           'My father has a house'
      b.   *hac*   *dom-us*   *patr-is*   *me-i*   *est*
           this.NOM.SG house-NOM.SG father-GEN mine-GEN be.PRS.3SG
           'This is my father's house'

Ernout & Thomas (1951/1964, § 93), as well as Bennett (1895/1927), said that these constructional options were also conceptual alternatives. For Ernout & Thomas, the possessive dative displays something from the possessor's perspective ('My father has a house') while the genitive uses the possessed entity as the reference point ('The house is my father's house'). For Bennett (1895/1927, § 359), the genitive construction emphasizes the possessor, as it is the entity from which the predicate is stated, whereas the dative emphasizes the possession itself.

Some authors (Havers, 1911, p. 211, n. 2; Kühner & Stegman, 1971/1912, p. 308; Löfstedt, 1963, p. 65; Oudot, 1964, p. 147: Woodcock, 1959: 46) asserted from a pragmatic-informative perspective that the difference between possessive dative and genitive was that the former highlights the subject (*domus*) as an entity not previously known, while the latter, with a known subject, emphasizes the possessor.

More recently, Bolkestein (2001, p. 274) has claimed that the possessive dative is one of the two participants involved in a possession situation and suggested that the dative is an Experiencer, the person for whom the situation is relevant, while the genitive is a possessive.

According to Bassols (1976), in Latin there was another device for possession, the dynamic or sympathetic dative, with similar meaning to the genitive, but used more expressively. This dative, very frequent in both Archaic Latin and Late Latin, could appear with non-copulative verbs. The Latin grammatical tradition, including Spanish grammars, use the term sympathetic dative to cover both the sympathetic and the possessive dative (Van Hoecke, 1996, p. 13). According to Van Hoecke (1996, p. 15), this dative can be interpreted as the possession link between an interested person and an affected object.

The following is an example of this kind of possessive interpretation with an intransitive verb, *advolavit*:

(7) Nesc-io   qua vox   ad aures   *mihi*   advola-vit
     not know-PRS.1SG ADV sound.NOM to ear.ACC-PL DAT.1SG fly.PRF. 3SG
     not know from where sound to ears of mine flew
     'I do not know from where the sound flew to my ears'

3.   The Ditransitive Construction in Spanish

In Spanish grammar, the possessive dative is considered or interpreted as a possessive IO (Alcina y Blecua, 2001/1975, § 7.2.13; RAE 2010, § 35.4.1b; RAE, 2009: § 35.7f) that denotes the possessor of a body part (inalienable possession) – a notion understandable as inclusion – or the possessor of another object linked to "the sphere of the personal" (alienable property) (Di Tullio, 1997, p. 128).

The possessive dative predication in my approach is an instance of the Ditransitive Construction, that is, an example or extension of an abstract pattern or conventionalized pairing of form and meaning, inferred from use but existing independently from it (Goldberg, 1995). The prototypical instance of this construction is the structure with the verb *dar*, 'to give', exemplified in (8),[1] whose

---

1.   V=italics; DO=bold; IO =italics and bold.

semantics is that of the ditransitive construction: an act of a perfective transference, (see Figure 1).

| - - - - - - V - - - - | - - - - - - - - - - | → | - - - - - - V - - - - - - - - - - - - - |

SBJ   +din   DO   IO   SBJ   +din   DO   at   IO

       +perf   +perf

Figure 1.   Initial and final moment with the verb *dar*, 'to give' (Ortiz Ciscomani, 2011, p. 110)

(8) a. É   *dieron*   ***al Rey***   **las cartas**   **que traian**
         and give-PRF.3PL to.the king.DAT the letter.ACC.PL that bring.PST.PL
         del   **Rey Don Carlos de Navarra**   [*Juan I*, 134]
         from the King Don Carlos de Navarra-SOURCE
         'And (they) gave to the King the letters they brought from King Don Carlos
         from Navarra'
     b. Los   tres   afortunados españoles   que *dieron*   **diez**
         the.M.PL three lucky   Spaniard.SBJ.PL who give-PAST.3PL ten
         mil   **pesos**   ***a Delgado***   para que organizara
         thousand peso.ACC.PL   to Delgado.DAT   to   organize.SBJ.PST.3PL
         una compañía   de ópera   italiana,   están   de malas
         a   company.ACC   opera.GEN italiana.GEN be.PRS.3PL in bad shape
         'The three lucky Spaniards who gave ten thousand pesos to Delgado to
         organize an Italian opera company are in bad shape'   [Obras *IV*, 90]

In both example the objects are asymmetrical, an inanimate direct object (DO) and an animate–human IO. The DO – *las cartas que traian del Rey Don Carlos de Navarra* and *diez mil pesos* – is a theme, an entity which was moved to a new location, an active location: the animate human recipient. *Al Rey* (8a) and *a Delgado* (8b) are the targets of the event. Both are dynamic situations in which the subjects cause the DO to move to the IO, who is interpretable as the possessor of DO entity.

The objects share a property: they are both affected entities, that is, modified or impacted positively or negatively by the energy of the subject entity. In other words, the ditransitive construction involves a split transitivity situation (Wierzbicka, 1988).

The change in both objects is also contrastive due to the lexical asymmetry between them. The inanimate DO undergoes a location change, whereas the animate IO, a human, active entity target of the transference situation (Van Belle and Langendonck, 1996, p. 23), is positively (benefactive) or negatively (malefactive) affected.

In this perspective, the IO of the ditransitive construction can be conceptualized as an a*ctive experiencer in the target dominion* (Langacker, 1991, p. 327). This

IO is a location, final point, or destiny of the causal energy chain, a final *locus* of a dynamic or movement situation. This idea is not a new one. Lyons (1967, p. 390) stated that location is an ambit conceptually more basic than possession and existence. The affinity between possession and notions such as location, existence, and perfect aspect, although somewhat elusive (Langacker, 1995: 51), is based on a diachronic relationship in which both represent different stages of the same chain of grammaticalization, in which Location precedes possession and existence (Heine, 1997, p. 208) and possession is related to aspect. Even though it has been recognized that the IO-possessor in possessive constructions is an animated place, a space in which the DO-possessed object is located just as it happens with an object designated in existential and locative sentences (Seiler, 1983, p. 56), possession is a complex phenomenon that cannot be reduced to location (Seiler, 1983, p. 1–3 and 57).[2]

The possessive dative structures with non-transference verbs use the same ditransitive productive pattern, but they include a DO distanced from that of the basic construction in some degree and a human IO, the so called Possessor Dative, so they are low ditransitive (not prototypical) structures. But they are ditransitive constructions, as Harley (2000: 201) claims: "whenever the recipient argument tends to have wide scope over the theme argument, the construction is a ditransitive" (Harley in Goldberg, 2006, p. 201). Attention to the contribution of the construction is essential (Goldberg, 2006, p. 126).

In sum, my approach is *constructional*. I assume that Constructions in general and the Ditransitive Construction in particular are primed in production. Saying or hearing instances of grammatical pattern primes speakers to produce other instances of the same (Goldberg, 2006, p. 120–121), that is, there are caused-motion constructions as CAUSE-MOVE that prime caused-motion expressions.

The data presented in the next section support the conceptualization of the IO as an entity that behaves essentially as a location, and also as a potential possessor since it is typically a human entity. Conceptually, this IO is constructed as a Reference Point in whose domain the DO is located (Langacker, 1995, p. 76) and from which the conceptualizer establishes contact with this DO entity. The

---

2. The hypothesis that in many, and perhaps in all, languages existential and possessive constructions derive (both synchronically and diachronically) from locatives is developed in Lyons (1967). Stassen (2013) recognizes four encoding strategies for predicative possession. One of them is the Oblique Possessive that has two subtypes. In one, the oblique marking on the possessor NP has as its basic meaning the specification of a locational relation, but it would be possible to further subcategorize this type into Locative Possessive (with the possessor NP being marked by some item meaning 'at', 'on' or 'in') and Dative Possessive (with a marker 'to' or 'for' on the possessor NP).

Reference Point has implied or contextually determined prominence and its own dynamics (Langacker, 1993: p. 6). In a Ditransitive structure where two objects compete for pragmatic and grammatical prominence, the struggle favors the IO, a higher ranking participant than the DO both in the hierarchy of individuation and in language topicality (Givón, 1984, p. 139, Hopper and Thompson, 1984; Silverstein, 1976; Timberlake, 1977, p. 162). This conception is consistent with the function that has been attributed to the possessive dative in Latin (§ 2).

4.   Possessive Dative: Diachronic data

The documented data are structurally homogeneous, that is, all of them include a DO which is moved by a subject to a new animate location, a recipient. The DO entities are characterized by the particular inalienable or alienable relation they establish with the IO, interpretable as its possessor.

Although the DOs have the same syntactic function, they are semantically heterogeneous, which brings out a diverse relationship with the IO, its possessor and reference point.

Three groups can be distinguished with regard to these relationships:

a.   In the first group, the main different feature between the objects of the basic pattern is that, although they are asymmetrical, the inanimate DO is related to the animate IO entity as a part, often a body part, whereas the IO is a whole. In my analysis, both DO and IO are independent participants.
b.   The second group is made up of structures whose objects are also lexically asymmetrical, but the possessed DO noun refers to a property or an ownership of an animate possessor, the IO; and
c.   The third group is made up of structures in which the possessed DO noun refers to an important or very valuable entity for the IO, which is conceptualized as the abstract or virtual possessor. I have only two tokens of this strong affective relation in the corpus.

Although the objects are asymmetrical in all the groups, the data suggest that there is a continuum of loss of independence in the DO, and so loss of cognitive prominence, stronger in the case of kinship relations than in that of body parts. However, the kinship relation is based on the existence of the IO. Examples of each of these groups are provided below.

The first group, focusing on the part–whole relationships, especially of body parts, is exemplified in (9):

(9) a. *Derribaron los moros   con un búzano   el brazo*
fall.PRF.3PL the Moor.SBJ with a búzano.INSTR[3] the arm.ACC.SG
al   alférez   de una batalla de las   de cardenal, que se
to.the lieutenant.DAT of a battle of those of cardinal, that REFL
llamaba   Juan de Perea, sobrino del   adelantado Rodrigo Perea
call.PST.3PSG Juan de Perea, nephew of the adelantado Rodrigo Perea
'the Moors cut off with a búzano the arm of the lieutenant of a battle of those of the cardinal, who was called Juan de Perea, nephew of the adelantado Rodrigo Perea'   (*CRCII*, 277.19)

b. Ensoberbecido éste (…), ***les***$_i$ *cortó*   **las manos**
proud.ADJ   this.SBJ, they.DAT.PL cut.off.PRF.3SG the hand.ACC.PL
a dos caballeros   *portugueses*   *que allí asistían*$_i$
to two gentlemen.DAT.PL Portuguese.DAT.PL who there be.PST.3PL
'This one, with a high degree of pride, cut off the hands *of two Portuguese gentlemen who were there*'   (*Infortunios*, 25)

c. la obsidiana con la que los incas   hacían   sus
the obsidian with the which the Inca.SBJ.P make.PST.PL their
espejos   y los aztecas   los cuchillos con los que
mirror.ACC.PL and the Aztec.SBJ.PL the knife.ACC.PL with the which
sacaban   el corazón   a sus víctimas
extract.PST.3PL. the heart.ACC.SG to their victim.DAT.PL
'the obsidian with which the Incas made their mirrors and the Aztecs the knives with which they extracted the heart of their victims

                                                        (*Noticias*, 203)

d. a la primera carga del   enemigo, *le*$_i$ *quebraron* **una**
at the first   charge of the enemy,   him.DAT$_i$ break.PRF.3PL a
pierna   ***a uno de los lanceros***$_i$ (*Trofeo*, 150)
leg.ACC.SG to one of the spearmen.DAT.SG$_i$
'at the first charge of the enemy, they broke a leg of one of the spearmen'

In these examples, the objects are lexically asymmetrical as in the Ditransitive Construction pattern. In (9a) to (9d), the DO is inanimate (*the arm*, *the hands*, *the heart and a leg*), while the IO is animate (*al alférez…*, *a dos caballeros portugueses …*, *a sus víctimas*, *a uno de los lanceros*), and more prominent than the DO in topic hierarchy (Givón, 1984). Generally, the animate IO conveys new information, similar to what was claimed for the Latin Possessive Dative by some authors. Both entities, the DO and the IO, are affected, but while the DO is moved from its original place or only touched by an agentive subject, the IO is affected in a different

---

3. A 'búzano' is an ancient artillery piece.

sense. It is perceived as being physically or psychologically affected by the event (Kliffer, 1987).[4]

Table 1 includes all the DO nouns of this group occurring in the corpus. As we can see, between the 12th and 20th century the DO has been codified as a concrete noun referring especially to a part of a human entity, and it diversified and extended to other nouns types – i.e., mass, abstract – from the 16th century onwards.

Table 1. Possessed entities: Part–whole relation

Lexical type		12th	13th	14th	15th	16th	17th	18th	19th	20th
Concrete, part of a human entity	*ojo(s)* 'eyes'	x	x			x				
	boca(s) 'mouths'	x	x			x			x	
	pies 'feet'	x				x	x			
	manos 'hands'	x				x	x		x	
	brazo(s) 'arm(s)'			x		x				
	cabeça 'head'			x		x				
	dedos 'fingers'					x				
	corazón 'heart'					x				x
	cuerpo(s) 'body(ies)'					x				
	heridas 'wounds'					x				
	entrañas 'entrails'						x			
	pierna(s) 'leg(s)'						x		x	x
	cara 'face'						x			x
	llagas 'sores'						x			
	oreja 'ear'								x	x
	muelas 'teeth'								x	
	pecho 'chest'									x
	hombro 'shoulder'									x
	nuca 'nape'									x
	tetillas 'nipple'									x
Concrete, part of an animate entity	*agallas* 'guts'									x
	cogote 'nape'									x
	ala 'wing'									x
	hocico 'snout'									x
Mass	*suspiro* 'sigh'		x							
	lágrimas 'tears'		x							
	hiel 'gall'						x			
	babas 'slime'							x		
	sangre 'blood'								x	

(*continued*)

---

4. Kliffer (1987) affirms that if the possessor is not as affected by the event, then the Genitive, not the Dative, is used.

Table 1. (*continued*)

Lexical type		12th	13th	14th	15th	16th	17th	18th	19th	20th
Concrete	*nombre* 'name'		x							
Personal	*freno* 'bridle'		x							
properties	*hábito* 'robe'						x			
entity	*riendas* 'reins'								x	
	sandalias 'sandals'									x
	una t 'a t'									x
	blusa 'blouse'									x
Abstract	*armonía* 'harmony'						x			
	muerte 'death'						x			

As we can see in Table 1, DO nouns in ditransitive structures diversified and in-creased from 4 items in the 13th century, to ten in the 16th century and to twenty in the 19th–20th centuries.

As for the verbs in this possessive dative construction referring to the part–whole possessive relation, they are touch, modification, or affect verbs generally with a malefactive IO-possessor, as we have seen in (9). In the early centuries I found touch verbs like *besar*, 'to kiss', or perception verbs like *ver*, 'to see', in these structures, and they remained in subsequent centuries (10a-b).

(10) a. *besó*　　**al rey**　　**las manos**　　　　(*Juan I*, 130)
　　　　kiss-3SG.PRF to.the king-DAT the hands-ACC
　　　　'(he) kissed king's hands'

　　b. ***ha Anna Petra***　*le*　　*vi*　　　**las llagas** en la sintura
　　　　to Anna Petra_i.DAT her_i.DAT see.1SG.PRF the sores at the waist-LOC
　　　　'I saw Anna Petra's sores at the waist'　　　　(*DLNE*, 632)

In the examples above, the Possessive Dative is incorporated as a Participant in the event, a metaphorical Receptor negatively or positively affected by the event.

The change in the argument structure of the verb *ver*, 'to see' (10b), from transitive to ditransitive can be understood as an accommodation (Goldberg, 1995) or coercion (Bergs & Diewald, 2008) to the basic Ditransitive Construction resulting in a co-composition of the lexical meaning of the verb and the structural meaning of the construction (Lawers and Willems, 2011, p. 1222).

The construction with possessive dative can also refer to abstract possession. For example, in the case below (11a–b), the DOs *harmony* and *death* are abstract nouns that refer to an essential property or component of the animated entity in the dative (*the sky* and *many children, brothers or parents of yours*, respectively), and for that reason, those DOs are conceived as located in the IOs:

(11) a.   Víbora   de vapores espantosa, / cuyo   silbo   es   el
viper.voc of vapor.pl frightening.voc/whose whistle.nom be.prs the
trueno, /   Que *al cielo*   *descompone*   **la armonía**
thunder.nom,/ that to heaven.dat break.prs.3sg the harmony.do.sg
'Oh, frightening viper of vapors, / whose whistle is thunder, / That *to heaven*
breaks the harmony'   [*Trofeo*, 191]

b.   *A muchos hijos,   hermanos o padres vuestros*,   siendo
to many   children, brothers   or parent of yours.dat.pl, being
dignísimos de larga vida,   *aceleraron*   **la muerte**
worthy   of long life.dat.pl, accelerate.prf.3pl the death.acc.sg
'To many children, brothers or parents of yours, being worthy of long life,
accelerated the death'   [*Trofeo*, 144]

Regarding the second group, Table 2 illustrates the DO nouns documented in
the structures which refer to properties or ownership of the IO-possessive da-
tive. As we can see, although with fewer tokens in the corpus, they also diversified
diachronically:

Table 2. Nouns referring to ownership of the IOs

	12th	13th	14th	15th	16th	17th	18th	19th	20th
palacios 'palaces'		x							
caballos(s) 'horse(s)'				x					
papeles 'papers'						x			
fortuna 'fortune'						x			
frutos y chorizos 'fruits and sausages'							x		
soldado(s) 'soldier(s)'								x	
cigarros 'cigarettes'									x
radio 'radio'					x				x
poesía 'poetry'								x	
colaboración 'collaboration'								x	

See an example in (12a–b):

(12) a.   e   *quemo*   *a mi*   **los**   **mios palacios**   (*GEII*, 22)
and burn.3sg.prf to me.dat the.acc.p my palace.acc.pl
'And (he) burned my palaces'

b.   *Robó*   *al   primero* **su poesía**   **soñadora**
steal.prf.3sg to.the first.dat.sg his poetry.acc.sg dreamy.acc.sg
'(he) stole the first his dreamy poetry'   (Obras IV, 19)

We occasionally found a possessive adjective instead of an article in the DO noun phrase only with a human referent, like in (13a) (and in (12b) above), where possessor and possessed have a property or an ownership relation (i.e., the possessed entities are controlled by the possessor). The possessive adjective in (13a), in contrast with (13b), is an optional device to emphasize the effect on the human dative participant, here in first position in the clause.

(13) a. Y   *a ellos y   a él*$_i$ les$_i$ herían   **sus**
           and to them$_i$ and to him$_i$.DAT them$_i$.DAT wound.PST.3PL their
           soldados   [*Hist. Verd.*, 896]
           soldiers.ACC.PL
           'And (they) wounded them their soldiers'
      b. Les$_i$ hirieron   **los caballos**   *a ellos*$_i$
           them$_i$.DAT wound.PRF.3PL the horse.ACC.PL to them$_i$.DAT
           'And (they) wounded them their horses'   [*Hist. Verd.*, 896]

In the third group, Possessive Dative is associated with a kinship link as one of the three kinds of possessive relation. The corpus only registered two of these tokens with a human DO, *varones e mujeres* in (14):

(14) *A los   de Benjamin*   *matamos* **varones** e   **mujeres**
      to those of Benjamin.DAT kill.PRF.1PL men   and women.ACC
      'We killed men and women to those of Benjamin'   [*GEII*, 2.191b.26]

5. Summary/Concluding remarks

In this work, I have shown that:

- The Possessive Dative has been an alternative possessive construction since Early Latin, and its function is to codify the person for whom the situation is relevant.
- The approach revealed the motivation for the use of Possessive Dative: to encode the participant more prominent in the possession relation, the cognitive reference point, the entity frequently affected negatively by the situation.
- The Possessive Dative in ditransitive structures in Spanish instantiate the Ditransitive Construction as a pair of form-meaning correspondences. The Possessive Dative structure can be considered as coerced or accommodated to the Ditransitive Construction in order to give relevance to possessor over possessed.
- The DO in the corpus refers basically to concrete entities, but it tends to diversify lexically, and to extend to mass, abstract and animate entities, especially from the 16th century onwards.

Abbreviations

ACC	accusative		PL	plural
ADJ	adjective		PRF	perfect
ADV	adverb(ial)		PRS	present
DAT	dative		PST	past
DO	Direct Object		RFL	reflexive
GEN	genitive		SG	singular
IO	Indirect Object		SBJ	subject
LOC	locative		VOC	vocative
M	masculine		1	first person
NOM	nominative		3	third person
P	patient-like argument of canonical transitive verb			

References

a. Corpus

[*GEII*] = Alfonso el Sabio (1966). *General estoria*, segunda parte, II. A. G. Solalinde, L. A. Kasten & V. R. B. Oelschläger (Eds.). Madrid: Consejo Superior de Investigaciones Científicas, Instituto Miguel de Cervantes, MCMLXI.

[*CRCII*] = F. del Pulgar (Ed.) & J. Mata Carriazo (1943). *Crónica de los Reyes Católicos*, vol. II, Guerra de Granada. Madrid: Espasa Calpe.

[*Juan I*] = Rosel, C. (Colection). (1877). Crónica de Juan I. In *Crónica de los Reyes de Castilla II*, BAE, Madrid: M. Rivadeneyra.

[*Hist. Verd.*] = Díaz del Castillo, B. (1991). *Historia verdadera de la conquista de la Nueva España*, C. Sáenz de Santa María (Edition, index & prologue). Madrid: Alianza Editorial.

[*DLNE*] = Company Company, C. (Ed.). (1994). *Documentos Lingüísticos de la Nueva España, Altiplano-Central*. México: UNAM.

[*Infortunios*] = De Sigüenza y Góngora, Carlos. (1972). Infortunios de Alonso Ramírez. In M. Romero de Terreros (selection, prologue & notes), *Relaciones históricas* (pp. 1–71). México: UNAM.

[*Trofeo*] = De Sigüenza y Góngora, Carlos. (1960). Trofeo de la justicia española en el castigo de la alevosía francesa. In Carlos de Sigüenza y Góngora, J. Rojas Garcidueñas (selection, prologue & notes), *Obras históricas* (pp. 109–204). Second Edition. México: Editorial Porrúa.

[*Obras IV*] = Gutiérrez Nájera, M. (1984). *Obras IV/Crónicas y artículos sobre teatro, II (1881–1882)*, Y. Bache Cortés & A. E. Díaz Alejo Eds.). México: UNAM.

[*Noticias*] = Del Paso, Fernando. (1997), *Noticias del Imperio*, Second Ed., Ninth Reprint. México: Editorial Diana.

b. Bibliography

Alcina Franch, J. & Blecua, J. M. (2001/1975). *Gramática española*. 11th edition. Barcelona: Ariel.

Bassols de Climent, M. (1976). *Sintaxis latina*. Madrid: CSIC.

Bennett, Ch. E. (1895/1927). *New Latin Grammar*. Boston/ New York/ Chicago/ Atlanta/ San Francisco: Allyn and Bacon.

Bergs, A. & Diewald, G. (2008). *Constructions and language change*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110211757

Bolkestein, M. (2001). Possessors and experiencers in Classical Latin. In I. Baron, M. Herslund & F. Sorensen (Eds.), *Dimensions of possession* (pp. 269–283). Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/tsl.47.15bol

Company Company, C. (2001). Multiple dative-marking grammaticalization. Spanish as a special kind of primary object language, *Studies in Language*, 25, no. 1, 1–47. https://doi.org/10.1075/sl.25.1.02com

Di Tullio, Á. (1997). *Manual de gramática del español*. Buenos Aires: La isla de la luna.

Ernout, A. & Thomas, F. (1951/1964). *Syntaxe latine*. Paris: Éditions Klincksieck.

Givón, T. (1984). *Syntax. A functional-typological introduction*. Vol I. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/z.17

Goldberg, A. E. (1995). *A Construction Grammar Approach to Argument Structure*. Chicago/ London: The University of Chicago Press.

Goldberg, A. E. (2006). *Constructions at Work. The Nature of Generalization in Language*. Oxford: Oxford University Press.

Harley, H. (2002). Possession and the double object construction. *Linguistic Variation Yearbook* 2, 29–68.

Havers, W. (1911). *Untersuchungen zur Kasussyntax der indogermanischen Sprachen*. Straßburg: Trübner. https://doi.org/10.1515/9783111384566

Heine, B. (1997). *Possession. Cognitive sources, forces, and grammaticalization*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511581908

Hopper, P. J. & Thompson, S. A. (1980). Transitivity in grammar and discourse. *Language* 56, 251–99. https://doi.org/10.1353/lan.1980.0017

Hopper, P. J. & Thompson, S. A. (1984). The Discourse Basis for Lexical Categories in Universal Grammar. *Language*, 60, 703–52. https://doi.org/10.1353/lan.1984.0020

Kliffer, M. (1987). Los sustantivos intrínsecamente relacionales: un enfoque multinovelístico. *RSEL*, 17, 283–299.

Kühner, R. & Stegmann, C. (1971/1912). *Ausführliche Grammatik der Lateinischen Sprache*, Zweiter Teil: Satzlehre. Hannover: Hahnsche Buchhandlung.

Langacker, R. (1991). *Foundations of cognitive grammar. Descriptive application. Vol. II*. Stanford: Stanford University Press.

Langacker, R. (1993). Reference point constructions, *Cognitive Linguistics*, 4, 1, 1–38. https://doi.org/10.1515/cogl.1993.4.1.1

Langacker, R. (1995). Possession and possessive constructions. In J. Taylor & R. Mac Laury (Eds.), *Language and the cognitive construal of the world* (pp. 51–79). Berlin-New York: Mouton de Gruyter. https://doi.org/10.1515/9783110809305.51

Lawers, P. & Willems, D. (2011). Coercion: Definition and Challenges, Current Approaches, and New Trends. *Linguistics* 49, 6, 1219–1235.

Lyons, J. (1967). A note on possessive, existential and locative sentences. *Foundations of language*, 3, 390–396.

Löfstedt, B. (1963). Zum lateinischen possessive Dativ. *Zeitschrift für vergleichende Sprachwissenschaft*, 78, 64–83.

Ortiz Ciscomani, R. M. (2006). La Bitransitividad, Chap. 7. In C. Company (Dir.), *Sintaxis Histórica de la lengua española, Primera parte: la frase verbal*, Vol. I, (pp. 575–668). México: Universidad Nacional Autónoma de México /Fondo de Cultura Económica.

Ortiz Ciscomani, R. M. (2011). *Construcciones bitransitivas en la historia del español*, México: Universidad Nacional Autónoma de México/Universidad de Sonora.

Oudot, J. (1964). *Syntaxe latine*. París: Les Belles Lettres.

Payne, D. L. & Barshi, I. (1999). External possession. What, where, how, and why. In D. L. Payne & I. Barshi (Eds.), *External Possession* (pp. 3–29). Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/tsl.39.03pay

[RAE-Asale]= Real Academia Española-Asociación de Academias de la Lengua Española. (2009). *Nueva gramática de la lengua española*. Madrid: Espasa-Libros.

[RAE-ASALE] = Real Academia Española y Asociación de Academias de la Lengua Española. (2010). *Nueva gramática de la lengua española*. Manual. Madrid: Espasa.

Seiler, H. J. (1983). *Possession as an operational dimension of language*. Language Universals Series, 2. Tübingen: Gunter Narr.

Silverstein, M. (1976). Hierarchy of features and ergativity. In R. M. W. Dixon (Ed.), *Grammatical categories in Australian languages* (pp. 112–171). Canberra: Australian Institute of Aboriginal Studies.

Stassen, Leon. (2013). Predicative Possession. In: Dryer, Matthew S. & Haspelmath, Martin (eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://wals.info/chapter/117, Accessed on 2018-09-28.)

Timberlake, A. (1977). Reanalysis and actualization in syntactic change. In Ch. Li (Ed.), *Mechanisms of syntactic change* (pp. 141–77). Austin: Texas University Press.

Van Velle, W. & Van Langendonck, W. (1996). *The dative. Descriptive studies*. Vol. I. Amsterdam Philadelphia: John Benjamins. https://doi.org/10.1075/cagral.2

Wierzbicka, A. (1988). The semantics of internal dative in English. In Anna Wierzbicka *The semantics of grammar* (pp. 359–387). Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/slcs.18

Woodcock, E. Ch. (1959). *A new Latin syntax*. London: Methuen.

'Liking' constructions in Spanish

The role of frequency and syntactic stimulus in constructional change

Andrea Mojedano Batel
Indiana University

This study examines the diachronic change in argument structure in Spanish psych-verbs of 'liking', with emphasis on the change from nominative-experiencer *gustar* 'to like' to dative-experiencer *gustar*. While previous studies have looked at factors pertaining to frequency and semantics, this change must also be studied taking into account certain syntactic factors, and especially the evolution of prepositional finite clauses introduced by functional prepositions. Results suggest that the subcategorization properties of the preposition are grammatically relevant in determining the linguistic encoding of the arguments of Spanish 'liking' constructions. This study offers an extensive corpus study of 'liking' verbs, spanning the 13th to the 17th centuries, and adopts a constructionist usage-based view of syntactic productivity.

Keywords: argument structure, complementizers, morpho-syntax, psych-verbs, construction grammar, Spanish language

This study offers a survey of psychological verbs and constructions that convey the idea of 'liking' from medieval to Golden Age Peninsular Spanish (13th to 17th centuries), with special emphasis on the change in argument structure of the verb *gustar* 'to like,' leaving aside their posterior development. It also provides a comprehensive description of all the relevant argument structures characterized in semantic terms. The study's main aim is to establish if the subcategorization properties of the preposition are grammatically relevant in determining the linguistic encoding of the arguments of 'liking' constructions. I argue that these properties aided in determining the morphosyntactic realization of a given construction and in triggering processes of variation and change in argument structure. The current study thus combines two paths of research in the study of Spanish psychological verbs of liking: (i) research on change in argument structure across languages, and (ii) the syntactic development of complementizers in Spanish.

In this analysis, I adopt a constructionist usage-based view of syntactic productivity, along the lines of Bybee (1995), Goldberg (1995, 2003), and Barðdal (2008). Frequency is expected to affect how new verbs are classified (Goldberg, 1995). Bybee (1995) and Barðdal (2008) argue that type frequency of a particular construction plays a crucial role in determining the likelihood that a process will be extended to new roles: the higher the type frequency, the higher the productivity, which is understood as the ability of a pattern to be extended to new items.[1] Barðdal (2008) further hypothesizes that low-type frequency constructions, which tend to be semantically more restricted, will disappear unless they are high in token frequency and/or if they attract new items that are semantically similar.[2] In this work, I use terminology that is compatible with a constructionist model (Goldberg, 1995), although this paper is descriptive in nature, and as such, theory independent. Construction grammar assigns meaning directly to various abstract argument structure types, recognizing these patterns as linguistic units in their own right (Bencini & Goldberg, 2000). Goldberg (2003) further defines constructions as "basic units of language" that are highly routinized but can be readily extended to new contexts in principled ways.

1. Previous accounts of argument structure change in Spanish psychological verbs

The present study aims to shed further light on argument structure change within verbs of 'liking' in Spanish, with emphasis on the change from nominative-experiencer *gustar* 'to like' with a stimulus introduced by a preposition (henceforth, NOMexp-PPstim) (1) to dative-experiencer *gustar*, whose stimulus is the grammatical subject (henceforth, DATexp-NOMstim) (2).

(1) *Gust-o   de la   poesía*
    like-PRS.1SG of DET.F.SG poetry.NOM.SG
    'I like poetry'

(2) *Me   gust-a   la   poesía*
    me.DAT like.PRS.3SG DET.F.SG poetry.NOM.SG
    'I like poetry'

---

1. Type frequency refers to the number of distinct words that occur in a particular construction (Goldberg, 1995).

2. Token (or 'raw') frequency refers to the frequency of occurrence of a unit (e.g. a word, morpheme, or construction) used in running text (Goldberg, 1995).

The NOMexp-PPstim *gustar* construction (i.e. Arg[nom-experiencer] *gustar de* + Arg[nom-stimulus], lit. 'to like of something') spans a period of about two hundred years (16th–18th c.). Before the eighteenth century, only a few sporadic examples of the DATexp-NOMstim construction (i.e. Arg[dat-experiencer] + *gustar* 'to like' + Arg[nom-stimulus]) can be found, such as (3); the clause does not contain an overt stimulus, although I provide it within brackets for clarity.[3]

(3) *[Un charlatán que hacía magia] gust-ó   mucho a   Andrenio*
     [a   charlatan that did   magic] like-PST.3SG much   to.DAT Andrenio
     'Andrenio liked him [a charlatan that performed magic] very much'

                                                                    (17th c., *Criticón*)

The direction of this change in argument structure proves interesting because it runs counter to the diachronic tendency in other European languages for dative or oblique experiencers to be recoded as subject experiencers, due to the subject experiencer construction being understood as more "regular" from the perspective of a transitive language system (Haspelmath, 2001). An example of this latter process can be found in Old English *lician* 'to like,' which occurred with a preverbal dative experiencer NP and a postverbal nominative stimulus NP, as in (4) below. Allen (1995) notes that the recoding from a dative experiencer (i.e., *him liketh*) to a nominative experiencer (i.e., *he likes*) took place in Early Modern English.

(4) *ac gode   ne licode na lieora geleafleast* […]
     but God.DAT not liked not their faithlessness.NOM[…]
     'But God liked neither their lack of faith […]'

                                                        (AEHom 21 68, qtd. in Allen, 1995)

By contrast, subject experiencers being recoded as dative or oblique experiencers remain less common, although this recoding has been attested in earlier stages of Germanic (Allen, 1995; Barðdal, 2009) and in Romance (Bauer, 2000). This recoding of arguments encompasses an attraction of regular transitive verbs into a low-type frequency construction that is semantically coherent, meaning that it displays internal consistency (Barðdal, 2008). Many researchers argue that DATexp-NOMstim *gustar* developed by analogy with dative-experiencer medieval psych-verbs *placer* 'to please, to like' and *pesar* 'to cause regret or pain' (Melis, 1998; Vázquez Rozas & Rivas, 2007; Melis & Flores, 2013; a.o.), which were the most frequent psych-verbs in medieval Spanish (Melis, 1997). However, it seems implausible that the infrequent,

---

3. Baltasar Gracián, author of *El Criticón*, was from Zaragoza. Most authors of the texts analyzed in this study, however, were Castilian: Miguel de Cervantes, Fernando de Rojas, Alfonso Martínez de Toledo, Lope de Vega, a.o. See Appendix 1 for full list of authors, when available.

fossilized verb-construction pairing DATexp-NOMstim *placer*, which only shows thirteen tokens in my data for the seventeenth century, could have exerted any analogical pressure – at least by itself – on the frequent Golden Age NOMexp-PPstim construction, which shows 103 tokens, all with *gustar*, for the same century; an analysis of how the 'disliking' verb *pesar* affected the change in argument structure of *gustar* is left for future studies.

To sum up, most researchers working on this phenomenon have looked at factors pertaining to frequency and semantics. Yet, the change in argument structure of Spanish verbs of 'liking' must also be studied taking into account certain syntactic factors, namely, the context of the creation of subordinating patterns and complementizers in Spanish (Tarr, 1922; Serradilla Castaño, 1995, 1996, 1997; Barra Jover, 2002; Amaral & Delicado Cantero, 2018), especially the evolution of prepositional finite clauses introduced by functional prepositions, that is, the evolution of the syntactic pattern [P [fin-CP que]] (Delicado Cantero, 2013). This syntactic pattern was not yet fully acceptable in the Spanish Golden Age period, when NOMexp-PPstim *gustar* was the most frequent construction used to convey the idea of 'liking.' My goal is to provide a syntactic account for the change of argument structure of *gustar*, taking into consideration the evolution of subordinate finite clauses. The diachrony of variation and change in the argument structure of 'liking' verbs is quite complex, as it involves some types of constructions that used *gustar* as a noun (such as *dar gusto* 'to give pleasure' and *ser gusto* 'to be of pleasure') but coded certain semantic roles differently. It also involves other dative-experiencer 'liking' verbs besides *placer*. It is beyond the scope of the present paper to discuss such matters in detail here, and while I examine them briefly, my study centers on the syntactic categories of stimuli of 'liking' argument structure constructions and how these might have aided in the change in the argument structure of *gustar*.

2.   Data collection and methodology

The evidence presented in this study comes from a corpus comprised of 34 digitized Peninsular Spanish texts ranging from the 13th century to the 17th century: 18 are narrative texts, and 16 are comedies written in verse from the Spanish Golden Age.[4] The data are divided into three chronological periods: early medieval Spanish (EMS, 13th–14th centuries); late medieval Spanish (LMS, 15th century);

---

4.   Please refer to Appendix A for a full list of the works compiled in the corpus, of their title abbreviations, and of their online source.

and Golden Age Spanish (GAS, 16th–17th centuries). In order to compile an exhaustive list of psychological verbs of 'liking' in the corpus, I made a list of all verbs that conveyed the idea of 'liking' found in a sample of fifty pages in each text and subsequently uploaded the texts that form the corpus onto the corpus software tool *AntConc* (Anthony, 2017), in order to extract all instances of constructions containing 'liking' verbs, that is, in order to extract the precise tokens. I looked for all possible morphological variations in tense, aspect, mood, person and number of each verb. Table 1 provides the list of 'liking' verb forms found in the corpus.

Table 1. List of 'liking' verb forms found in the corpus

Agradar (and the lexically connected *ser agradable*)
Encantar
Gustar (and the lexically connected *dar gusto* and *ser gusto*)
Pagarse (and the lexically connected *ser pagado* and *tenerse por pagado*)
Placer (and the lexically connected *haber placer*, *hacer placer*, and *tomar placer*)
Preciar

Table 2 provides raw and relative frequencies of the verb forms provided in Table 1, along with their relative frequency.

Table 2. Raw and relative frequencies by verb*

Verb	Early medieval Spanish		Late medieval Spanish		Golden Age Spanish	
	Raw frequency	Relative frequency	Raw frequency	Relative frequency	Raw frequency	Relative frequency
agradar	4	1%	25	8%	38	10%
encantar	0	0%	0	0%	2	1%
gustar	0	0%	0	0%	287	74%
pagarse	147	30%	13	4%	4	1%
placer	318	66%	269	84%	58	15%
preciar	14	3%	14	4%	0	0%
Total	**483**		**321**		**389**	

* In Table 2, all lexically connected verb forms are subsumed under the one that is a lexical verb, e.g., data for the verb *gustar* also includes frequencies for the lexically connected verb forms *ser gusto* and *dar gusto*.

Additionally, Table 3 shows the total number of tokens extracted for analysis per time period – divided by argument structure construction – along with their relative frequency.

Table 3. Raw and relative frequencies by argument structure construction*

Construction	Early medieval Spanish		Late medieval Spanish		Golden Age Spanish	
	Raw frequency	Relative frequency	Raw frequency	Relative frequency	Raw frequency	Relative frequency
DATexp-No stimulus	94	19%	79	25%	13	3%
DATexp-NOMstim	110	23%	123	38%	165	42%
DATexp-PPstim	74	15%	25	8%	1	0%
Middle voice	74	15%	15	5%	3	1%
Passive voice	57	12%	19	6%	5	1%
Verb-Object compound	57	12%	44	14%	43	11%
NOMexp-ACCstim	14	3%	16	5%	28	7%
NOMexp-PPstim	3	1%	0	0%	121	31%
NOMexp-No stimulus	0	0%	0	0%	10	3%
Total	**483**		**321**		**389**	

* The construction DATexp-No stimulus appears with a dative experiencer and no overt stimulus within the phrase (see example (a) below). Because the stimulus is not grammatically represented in the phrase, it is not possible to know if the argument structure contains a nominative stimulus or a stimulus introduced by a preposition.

(a) *E   díx-o=les   ella: Pláz-e=me.*
     and pst\say-3sg=dat.3pl nom.3sg please-prs.3sg= dat.1sg
     'And she said to them: "This pleases me"'   (13th c., *Sendebar*)

I have also explored the nature of the argument structures that the resulting verbs instantiate in the corpus. The identification of the number of verb forms realizing each argument structure has allowed me to determine the type frequency of that construction. Table 4 offers the type and token frequencies of the most relevant constructions by time period, along with the verb forms that appeared in said constructions.

Table 4. Type and token frequencies of most relevant constructions

Argument structure	Early medieval Spanish			Late medieval Spanish			Golden Age Spanish		
	Raw frequency	Type frequency	Verb forms	Raw frequency	Type frequency	Verb forms	Raw frequency	Type frequency	Verb forms
DATexp-NOMstim	110	2	placer, hacer placer	123	4	agradar, placer, hacer placer, causar agrado	165	5	dar gusto, ser gusto, ser agradable, agradar, placer
DATexp-No stim	94	1	placer	79	1	placer	13	2	placer, gustar
DATexp-PPstim	74	1	placer agradar,	25	1	placer	1	1	placer
Middle voice	74	3	pagarse, preciar pagarse agradar,	15	2	agradar, pagarse	3	1	agradar
Passive voice	57	1	placer	19	1	pagarse	5	1	agradar
NOMexp-PPstim	3	2		0	0	n/a	121	1	gustar

3. A survey of liking verbs and constructions
in the history of Spanish (13th–17th c.)

The two most frequent verbs of liking in EMS are *placer* (*N* = 318, 66%) and *pagarse* 'to be content, to like' (*N* = 147, 30%). *Placer* occurred in two possible constructions, both with a dative experiencer. Its stimulus, however, could either appear as a grammatical subject (5a), that is, in a DATexp-NOMstim construction, or as an oblique introduced by a preposition (typically *de* 'of', but also *con* 'with') (5b), that is, in a DATexp-PPstim construction. Both DATexp-NOMstim and DATexp-PPstim constructions were inherited in Old Spanish directly from Latin (see Elvira, 2006, for a thorough discussion on this issue).

(5) a. *Et   plog-o   a   la   fembra,   como plac-ía   a*
          and PST\please-3SG to.DAT the woman.DAT as   please-PST.3SG to.DAT
          su marido,   la   morada   del   zarapico con ellos
          her husband.DAT DET.NOM lodging.NOM of.the curlew   with them
          'And the woman liked that the curlew lived with them, as did her husband'
          (13th c., *CD*)

     b. *Plaz-e=me   d'esta presentaja*
          please-PRS.3SG=DAT.1SG of.this present
          'I am pleased with this present'   (13th c., *CMC*)

Furthermore, all throughout the Middle Ages, subordinate finite clause stimuli of dative-experiencer *placer* were introduced by the complementizer *que* 'that' ([fin-CP que]) directly after the verb; in this verb-construction pairing, a preposition never subcategorized for ([fin-CP que]) (6).[5]

(6) *plaz-e-r=me   ía   que   sopié-sse-des   lo   que*
      please-TV-INF=DAT.1SG COND.3SG COMPL PST\know-SBJV-2SG DET.N REL
      dix-o   una vez   el   conde Ferrant Gonsáles a   Muño Lainez.
      PST\say-3SG one time DET count Ferrant Gonsáles to.DAT Muño Lainez
      'It would please me that you knew what count Ferrant Gonsáles said once to Muño Lainez'   (14th c., *CL*)

The verb *pagarse*, on the other hand, just like its Latin antecedent PĀCĀRE, displayed either a passive construction (7a), or a middle voice construction with a nominative experiencer, a middle voice marker *se*, and an oblique stimulus

---

5. The terms 'finite subordinate clause' and 'complementizer clause/phrase' (CP) refer in the present study to phrases headed by *que* 'that' [fin-CP que]. The term 'non-finite subordinate clause' refers to an infinitival phrase (IP).

introduced by the preposition *de* 'of' (7b).[6] More rarely, *pagarse* occurred in transitive causative constructions (7c).

(7) a. *Mio Çid   de lo   que ve-ie   mucho era*
    Mio Cid.NOM of DET.N REL see-PST.3SG much be.
    paga-do
    PST.3SG please-PTCP
    'Mio Cid was very pleased with what he was seeing'   (13th c., *CMC*)

    b. *Et la garza,   pagá-ndo=se   mucho de aquella morada,*
    and the heron.NOM like-GERUND=REFL much of that lodging
    entristec-ió
    become.sad-PST.3SG
    'And the heron, quite liking that lodging, became sad'   (13th c., *CD*)

    c. *De tal   guisa   l-o-s   pag-a   Mio Cid.*
    of such manner ACC-M-3PL please-PRS.3SG Mio Cid.NOM
    'In such manner Mio Cid pleases them'   (13th c., *CMC*)

Placer continued to be the most frequent verb of the two in LMS ($N = 269$, 84%), still appearing in both the DATexp-NOMstim and the DATexp-PPstim constructions. Meanwhile, this period showed the almost complete demise of *pagarse* ($N = 13$, 4%) and the upsurge of *agradar* 'to like' ($N = 25$, 8%), which appeared predominantly in a DATexp-NOMstim construction ($N = 18$, 72%), as in Example (8a) below (see

---

6. *Pagarse* finds its origins in the Latin transitive verb PĀCĀRE 'to pacify' (< PĀCŌ, -ĀRE, -ĀUĪ, -ĀTUM). PĀCĀRE was mostly used in the passive form, usually with a location theme, i.e., *provincia pācāta est* [the province was pacified] (liv. 23,24,12, in *Thesaurus Linguae Latinae*). Yet, Du Cange's (1610–1688) *Glossarium mediae et infimae Latinitatis* offers an example of PĀCĀRE in the middle voice, with a meaning of 'contentum esse [to be happy, content]', provided in (a) below. This Late Latin example displays an argument structure which is the constructional antecedent of *pagarse (de)*. The construction includes a nominative experiencer, an ablative stimulus introduced by the preposition *de* 'of', and the reflexive pronoun *se* (i.e. Arg[nom.] + *pacare se de* + Arg[abl.] 'nominative argument + be paid of + ablative argument') (b). The ablative stimulus seems to be an ablative of cause, marking the reason why the experiencer feels a certain way.

    (a) **Pacare se**, (Du C.) contentum esse; *megnyugodni, megelégedni*. Ljub. Mon. Sl. I. 36. an. 1225: obligavit se … dare talem *pacatorem* Antolino … quod possit **pacare se** Ant. de centum libris.

    (b) *quod   poss-it   **pacare**   se Ant. **de***
    because be.able-PRS.SBJV.3SG become.happy.INF REFL NOM of
    centum   libris
    one.hundred books.ABL
    'Because Ant. would be able to be happy/content with a hundred books.'

Graph 1 for data regarding verb frequency per time period). *Agradar* also shows two instances of the middle voice construction (8b).

(8) a. *Bien me   agrad-a-n   tus   palabra-s*
         well DAT.1SG please-PRS-3PL your word-NOM.PL
         'I am quite pleased with your words / I like your words very much'

         (15th c., *LC*)

     b. *Ojos   hay   que de lagaña se   agrad-a-n*
         eyes.ACC there.are REL of rheum REFL please-PRS-3PL
         'There are eyes that are pleased with rheum'   (15th c., *Corbacho*)
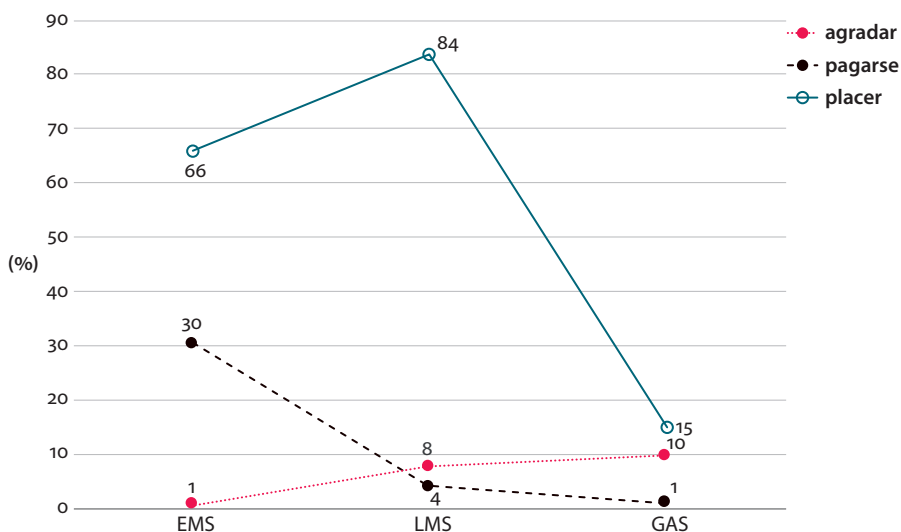
Graph 1. Relative frequencies of *agradar*, *pagarse*, and *placer* over time

With regard to the verb *gustar*, before the sixteenth century, this verb generally conveyed the idea of trying or tasting something (from Lat. GUSTARE 'to taste, sip; have some experience of; enjoy,' Arg[nom-experiencer] + Verb + Arg[acc-stimulus]). *Gustar* was construed with an animate or metaphorical subject, where the internal argument could either be a direct object (9a), especially for physical taste, or a prepositional object introduced by *de* (9b), particularly when it referred to a nominal thing or situation (Batllori, 2012; Melis & Flores, forthcoming).[7]

---

7. In Government and Binding Theory (Chomsky, 1981), the verb's external argument is structurally realized outside of the maximal projection headed by the verb (i.e., the VP) and is expressed by the subject. The remaining arguments are internal arguments generated inside the VP (Williams, 1981).

(9) a. *et   por buena fe   yo   ov-e   gosta-do   la*
         and by   good faith NOM.1SG have.PST1SG taste-PTCP the
         dulçor   della
         sweetness.ACC of.it
         'And in good faith I tasted the sweetness of it'
                                    (*Calila*: 157, as cited in Batllori, 2012)

     b. *El   que gostó   del   dulçor   de alguna obra,*
         NOM.3SG REL taste.PST.3SG of.the sweetness of some   deed,
         sufre   el   amargor de las sus carreras
         suffer.PRS.3SG the bitterness of the its endeavors
         'One who experienced the sweetness of a deed, suffers the bitterness of its
         endeavors'   (*Bocados de Oro* (Bonn, 111); Cuervo, DCRLC:
                                    sv. *Gustar*, as cited in Batllori, 2012)

Gustar entered the sphere of emotions via metaphorical extension between the
fourteenth and sixteenth centuries, when it acquired the meaning of 'to take pleas-
ure in x,' with no significant structural syntactic change (Corominas, 1954; Melis,
1998; Elvira, 2006; Vázquez Rozas & Rivas, 2007), namely, the *gustar* construction
appeared in a NOMexp-PPstim construction (10). This schema generally subcat-
egorized for nouns, infinitival clauses and relative clauses. By the Spanish Golden
Age, the verb *gustar* was the most frequent 'liking' verb form.

(10) *Gust-é   mucho – ponder-a-ba   Andrenio – de ver=l-a-s*
      like-PST.1SG much   ponder-TV-PST.3SG Andrenio of see.INF=ACC-F-PL
      tan bizarr-a-s
      so bizarre-F-PL
      'I liked it very much – pondered Andrenio – seeing them [acting] so bizarrely'
                                                      (17th c., *Criticón*)

It is important to note that nominative-experiencer *gustar* usually appeared without
a preposition, (i.e., in a NOMexp-ACCstim construction) when the stimulus was
a subordinate finite clause, as in (11).

(11) *Si gust-áis,   señores, que os   dig-a   en breves*
      if like-PRS.2PL gentlemen COMPL DAT.2PL PRS.SBJV\say-1SG in brief
      razones la   inmensidad de mis desventuras
      reasons the immensity of my misadventures
      'If you like/if you want, gentlemen, I will briefly tell you the enormity of my
      misadventures'   (17th c., *DQ*)

Although *gustar* was the most frequent verb form in the Golden Age (overwhelm-
ingly found in a NOMexp-PPstim construction), the DATexp-NOMstim con-
struction continued to be the most frequent schema with verbs of 'liking' during

the Spanish Golden Age (N = 165, 42%). Thus, the higher type frequency of the DATexp-NOMstim construction (which appeared with five verbal forms in this period) over the NOMexp-PPstim construction (which only appeared with the verb *gustar* in the same period) would make the DATexp-NOMstim construction the more productive one of the two, in terms of how likely it was to occur with a novel item (Bybee, 1995, Bybee & Thompson, 2000; Barðdal, 2008). Taking these assertions into account, I argue that there was constructional [analogical] extension from the DATexp-NOMstim construction to the NOMexp-PPstim *gustar* construction. That is, analogical extension did not take place solely from DATexp-NOMstim *placer* to NOMexp-PPstim *gustar*, since the former verb-construction pairing was very infrequent by the Golden Age, but from all 'liking' verb forms instantiated by the DATexp-NOMstim construction together. Also, importantly, and as I will discuss in detail in the following section, the DATexp-NOMstim construction was the only construction of the two that seemed to easily accept finite subordinate clauses as stimuli. I argue that this is due to the fact that the use of prepositions before a subordinate finite clause headed by *que* 'that' [P [fin-CP que]] is a phenomenon that only took place at the end of the Golden Age (Serradilla Castaño, 1995, 1996, 1997; Barra Jover, 2002).

The data reveal that during the Middle Ages, there were two frequent 'liking' verbs that differed in their argument structures from each other. In the thirteenth and fourteenth centuries, *placer* and *pagarse* were the two most frequent 'liking' verbs. In the fifteenth century, *pagarse* decreased drastically in frequency, while the use of *agradar* increased; meanwhile, *placer* was the most frequent 'liking' verb of all during this time period. In the Golden Age, *gustar* emerged as a verb of 'liking,' gaining ground on the other two frequent 'liking' verbs, *agradar* and *placer*. The following section will delve into the analysis of the syntactic properties of stimuli in 'liking' constructions, in order to establish if the subcategorization properties of the preposition were grammatically relevant in determining the linguistic encoding of the arguments of 'liking' constructions.

4. Stimuli of 'liking' constructions: An analysis of their syntactic properties

A fine-grained analysis of the syntactic properties of the stimulus will illuminate the path of change in 'liking' argument structure constructions, especially as that path of change pertains to finite and non-finite subordinate clauses. A focus on subordinate clauses is vital to the present study because they add up to a sizable amount of all overt stimuli within the clause, as illustrated in Table 5 below.[8] Furthermore, the

---

8. Total amounts in Table 5 differ from total amounts in Table 3. This is due to tokens with no overt stimuli within the clause not being considered for Table 5.

present data shows that different argument structures tend to appear with different types of syntactic stimuli, as will be explained in detail in the present section. Thus, the diachronic path of subordinate clauses might have contributed to the choice of one 'liking' argument structure construction over another and as such, must be studied vis-à-vis the development of 'liking' argument structures.

Table 5. Raw and relative frequencies of all syntactic types of stimuli for all argument structures combined*

Stimulus	Early medieval Spanish		Late medieval Spanish		Golden Age Spanish	
	Raw frequency	Relative frequency	Raw frequency	Relative frequency	Raw frequency	Relative frequency
NP	242	68%	102	45%	174	51%
RC	29	8%	21	9%	13	4%
IP	17	5%	40	18%	98	28%
CP	66	19%	62	28%	59	17%
Total	**354**		**225**		**344**	

* *NPs* refer to noun phrases, *RCs* to relative clauses, *IPs* to infinitival phrases, and *CPs* to complementizer phrases.

4.1   A diachronic account of clausal complementation in Spanish

4.1.1   *The diachrony of infinitival clauses*

Infinitives in historical Spanish display a typically nominal distribution: they can appear as subject, object and prepositional object; they can also be introduced by articles and participate in clausal nominalization (Delicado Cantero, 2013). Infinitives, while having nominal characteristics, are traditionally described as mixed categories, because they can still take objects (Beardsley, 1921). The following example shows coordination between two PPs, the first one with a noun and the second one with an infinitive:

(12) *sobrepui-a-ua   en piedad et   en dar   limosnas*
      excel-TV-PST.3SG in piety   and in give.INF alms
      'He excelled at piety and at giving alms'   (13th c., *Primera Crónica General*,
                                Alfonso X, as cited in Delicado Cantero, 2013, p. 76)

Infinitival clauses already appeared with prepositions in the earliest preserved Spanish (Schulte, 2007b). Additionally, Schulte (2007b) notes that plain infinitives in medieval Spanish (that is, non-prepositional infinitives) were more frequent than prepositional infinitives by a factor of nearly three, indicating a fundamental shift in favor of the [PP + infinitive] construction, which became more frequent around the sixteenth century, coinciding with an increase in the use of subordinate finite and infinitival clauses (Schulte, 2004). This extreme divergence, however, was only

temporary, as both constructions presented almost equal frequencies by the end of the seventeenth century; this equal share continues today in present-day Spanish (Schulte, 2007a).

4.1.2   *The diachrony of subordinate finite clauses*

The use of prepositions before a subordinate finite clause headed by *que* 'that' [P [fin-CP que]], as in (13), is a phenomenon that only took place late in the history of Spanish.

(13) *yo   no podr-é   afirm-a-r   si la dulce mi*
      PRON.NOM.1SG NEG be.able.to-FUT.1SG claim-TV-INF if the sweet my
      enemiga gust-a   o no de que   el   mundo sep-a
      enemy like-PRS.3SG or NEG of COMPL the world know-PRS.SBJV.3SG
      que   yo   l-a   sirv-o
      COMPL NOM.1SG ACC-F.3SG serve-PRS.1SG
      "I cannot say positively whether my sweet enemy is pleased/likes or not [of] that the world should know I serve her"   (17th c., *DQ*)

Serradilla Castaño (1995) argues that in medieval Spanish, nouns and infinitives could be preceded by a preposition when they appeared as internal arguments of psychological or speech verbs, as in (5b); however, when the verb's internal argument was a complementizer phrase (CP), then, as a general rule, this CP was joined directly to the verb, without a preposition, as in (6). Tarr (1922), Serradilla Castaño (1995, 1996, 1997) and Bogard and Company (1989) demonstrate that subordinate finite clauses underwent analogical extension and a regularization process through the extension of the preposition used with nouns and infinitives to the subordinate finite clause.

While in medieval Spanish, the dependent infinitive could appear in the same nominal contexts as a noun (as in example [12] above), the nominalization of the subordinate finite clause was gradual, not being able to appear in the same contexts as nouns and infinitives until the end of the Golden Age (Serradilla Castaño, 1997; Barra Jover, 2002). Barra Jover argues that the co-appearance of a subordinate finite clause with a determiner is proof of nominalization, and as such, he states that it is only in the late seventeenth century when we find [fin-CP que] preceded by the determiner *el* 'the'.

4.2   'Liking' constructions in Spanish and their subcategorization properties

The present section analyzes the subcategorization properties of 'liking' constructions over time, in terms of the syntactic type of stimuli they appear with, paying special attention to the appearance of finite subordinate clauses as stimuli in these constructions.

4.2.1   *'Liking' constructions in early medieval Spanish*

In EMS, the DATexp-NOMstim construction exhibited a clear preference for [fin-CP que] as stimuli, as in Example (6). The DATexp-PPstim construction, on the contrary, never appeared with [fin-CP que] as stimuli in the same time period, showing a preference for stimuli that contained more nominal features, i.e., NPs, IPs, and RCs, in that order, as can be gauged from the information portrayed in Charts 1 and 2, and in Example (5b). All charts in this study show two types of data: raw and relative frequencies. Relative frequencies are presented as percentages in parentheses. As an example, in Chart 1, DATexp-NOMstim constructions with NPs as stimuli show a raw frequency of 20 tokens, which is 23% of all DATexp-NOMstim constructions for EMS data.
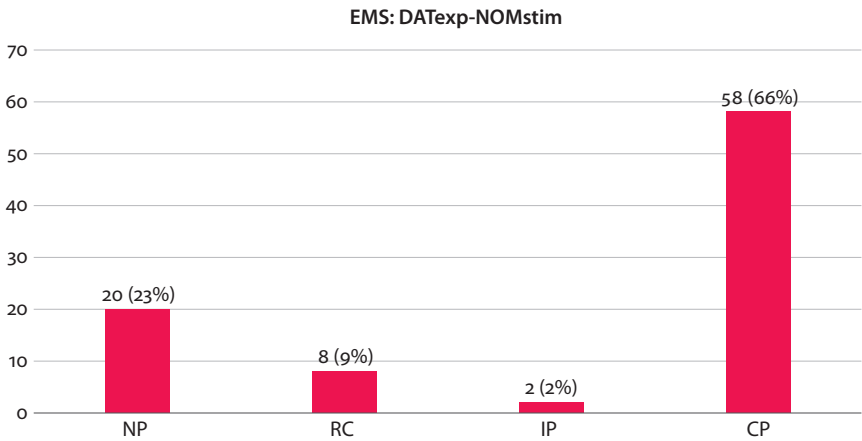
EMS: DATexp-NOMstim

Chart 1. Types of stimuli with DATexp-NOMstim constructions in EMS
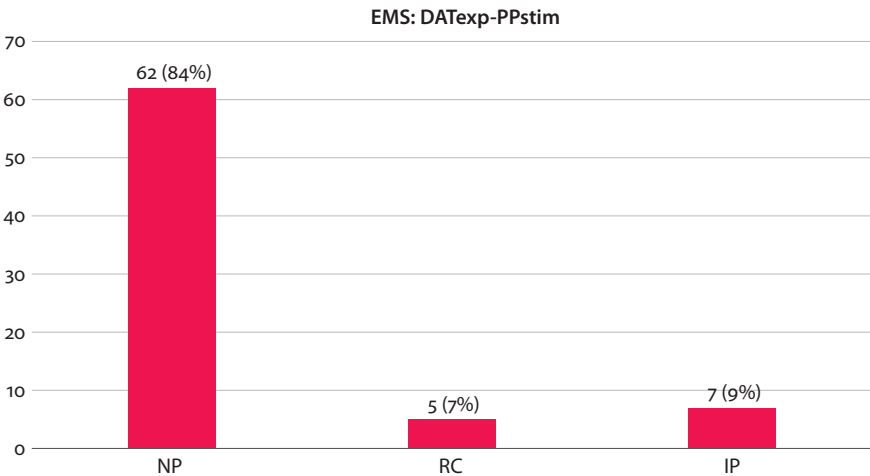
EMS: DATexp-PPstim

Chart 2. Types of stimuli with DATexp-PPstim constructions in EMS

It must be noted that the DATexp-PPstim construction already appeared with RCs as stimuli in some of the earliest known Spanish texts, such as *El cantar de Mio Cid* and *Calila e Dimna* (14a-b). This is because the relative clause, headed by a determiner, already had the necessary nominal features to be subcategorized by the preposition.

(14)   a.   *De lo   que el   moro   dix-o,   a   l-o-s*
            of   DET.N.SG REL   DET moorish.man PST\say-1SG to.DAT DET-M-PL
            infante-s   no   plaz-e.
            prince-PL   NEG please-PRS.3SG
            'The princes were not pleased with [of] what the moorish man said'
                                                                            (13th c., *CMC*)

       b.   *plác-e=le   con   lo   que   Dios fac-e*
            please-PRS.3SG=DAT.3SG with   DET.N.SG REL   God   do-PRS.3SG
            'He is pleased with what God does'   (13th c., *CD*)

In sum, EMS constructions showed a preference for different types of stimuli depending on their argument structure, strengthening Bencini & Goldberg's (2000) proposal of abstract argument structure types containing meaning in their own right. The DATexp-NOMstim construction allowed the speaker to introduce a less definite, more abstract stimulus such as a CP, while the DATexp-PPstim construction subcategorized for stimuli with more nominal features. Additionally, the verb *pagarse*, which introduced its stimulus through a preposition, behaved similarly to the DATexp-PPstim construction, by appearing mostly with NPs as stimuli ($N = 104$, 71%), as illustrated in (15).

(15)   *vid-o   una muger muy fermosa e   pag-ó=se   mucho*
       PST\see-3SG DET woman very beautiful and like-PST.3SG=REFL much
       de ella
       of   NOM.F.SG
       'He saw a very beautiful woman and liked her very much' (13th c., *Sendebar*)

4.2.2   *'Liking' constructions in late medieval Spanish*

Barðdal's (2008) predictions regarding low type frequency constructions – that is, that these constructions will disappear unless they are high in token frequency, and/ or if they attract new items that are semantically similar – are borne out for LMS in terms of the competition between the DATexp-NOMstim and DATexp-PPstim constructions: the DATexp-NOMstim construction was more frequent and showed a higher degree of constructional persistence, as Latin verbs that previously realized this construction continued to do so in Old Italian (Fedriani, 2014) and in Old and Modern Spanish. Thus, by the fifteenth century, the DATexp-NOMstim construction attracted the similar, albeit less frequent, DATexp-PPstim construction, which showed a less productive pattern. Furthermore, CPs and RCs typically appeared

in the DATᴇxᴘ-NOMꜱᴛɪᴍ schema (*N* = 47, 76% and *N* = 21, 100%, respectively) and CPs never displayed a DATᴇxᴘ-PPꜱᴛɪᴍ schema. Interestingly, most IP stimuli (*N* = 22, 55%) and NP stimuli (*N* = 33, 32%) also appeared in the DATᴇxᴘ-NOMꜱᴛɪᴍ construction, showing a reverse tendency to that of EMS, where IP and NP stimuli displayed a preference for DATᴇxᴘ-PPꜱᴛɪᴍ construction. This points to a low-type but semantically coherent schema – the DATᴇxᴘ-NOMꜱᴛɪᴍ construction – spreading to the same verb in a different construction (i.e. *placer*), as well as to synonymous or semantically similar verbs, such as *agradar* 'to like, be pleasing'. Charts 3 and 4 display the total occurrences of DATᴇxᴘ-NOMꜱᴛɪᴍ and DATᴇxᴘ-PPꜱᴛɪᴍ constructions in LMS, along with the types of stimuli with which they appeared.
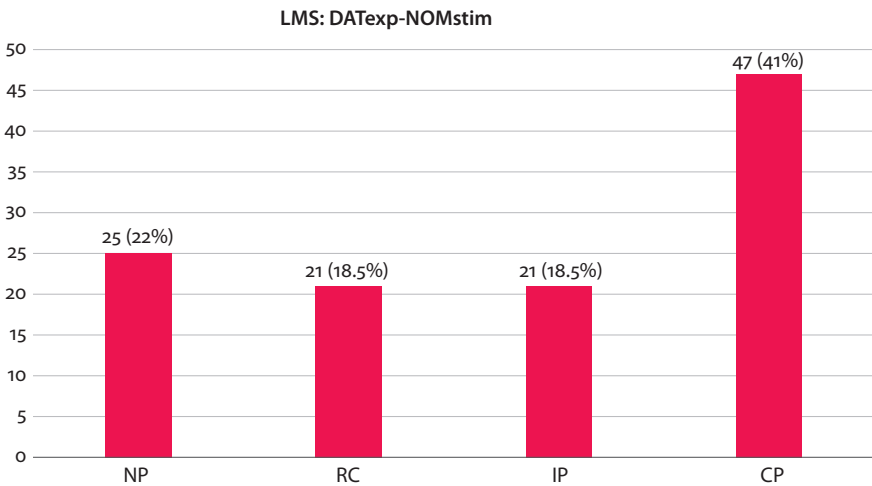
Chart 3. Syntactic types of stimuli with DATᴇxᴘ-NOMꜱᴛɪᴍ constructions in LMS
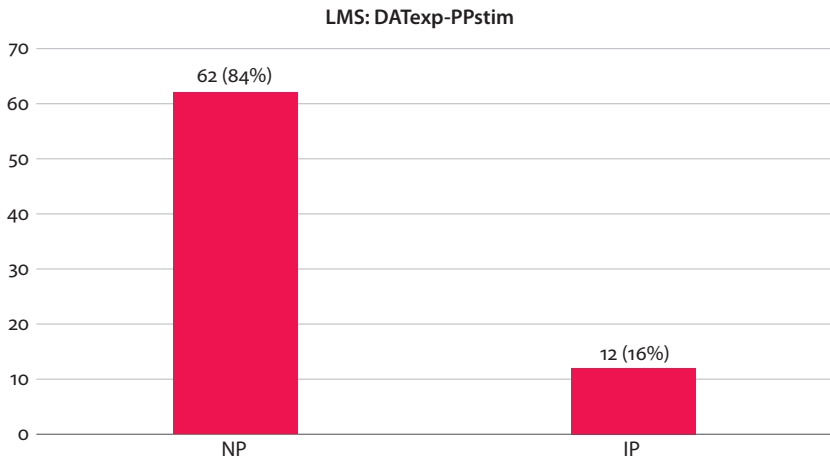
Chart 4. Syntactic types of stimuli with DATᴇxᴘ-PPꜱᴛɪᴍ constructions in LMS

4.2.3   Experiential liking constructions in Spanish: The Golden Age

In GAS, subordinate finite clauses headed by *que* mostly appear as stimuli in the DATExp-NOMstim construction (*N* = 40, 68%). The second most frequent construction in which these clauses appear is the NOMexp-ACCstim construction (*N* = 11, 19%), as in (16).

(16)   *Pues nos   const-a   que   gust-a   Dios que   como*
       as   DAT.1SG be.known-PRS.3SG COMPL like-PRS.3SG God COMPL as
       proprios cada uno sient-a   los trabajos de su   prójimo
       own   each one feel-PRS.3SG the labors   of POSS.SG neighbor
       'It is known to us that God likes that each one of us understands the labors of
       others as if they were one's own'   (16th c., *GdA*)

While the transitive construction (that is, the NOMexp-ACCstim construction) is more frequent than the DATExp-NOMstim construction in Spanish, it is less semantically coherent, being instantiated by many more verb classes. Conversely, the DATExp-NOMstim construction displays a very high degree of semantic consistency as well as diachronic stability. This enabled it to resist analogical leveling to the transitive schema and also aided in the expansion of the construction to the semantically related verb *gustar*.

The data reveal that the degree of definiteness (in terms of nominal features) of the stimulus may account for the alternation in argument structure. Subordinate finite clauses headed by *que* hardly ever appeared in a NOMexp-PPstim schema in Golden Age Spanish, meaning that the subcategorization properties of the preposition (usually *de* 'of') did not easily allow these clauses because they lacked nominal features. Consequently, the NOMexp-PPstim construction mainly appeared with nouns and infinitives as stimuli, as in (10). Charts 5 and 6 display the total occurrences of DATExp-NOMstim and NOMexp-PPstim constructions in GAS, along with the types of stimuli with which they appeared.

In the GAS data, NOMexp-PPstim *gustar* appears with a finite subordinate clause as stimulus only on four occasions (3%), while the DATExp-NOMstim construction appears with this same type of syntactic stimulus on 43 occasions (26%). These data are in line with the diachrony of infinitives and subordinate finite clauses, discussed in Section 4.1, as both Serradillo (1995, 1996, 1997) and Barra Jover (2002) argue that it was only at the end of the seventeenth century when the syntactic pattern [P [fin-CP que]] became fully acceptable, but that the [PP + infinitive] pattern was acceptable from the earliest Spanish texts.

In this same vein, it is not irrelevant that the DATExp-NOMstim verb *placer*, quite infrequent in GAS, overwhelmingly appeared with [fin-CP que] as stimuli (*N* = 37, 92.3%). The similar DATExp-No stimulus construction appeared 12 times (92%) with the verb *placer*. It seems that *placer* survived just to be able to
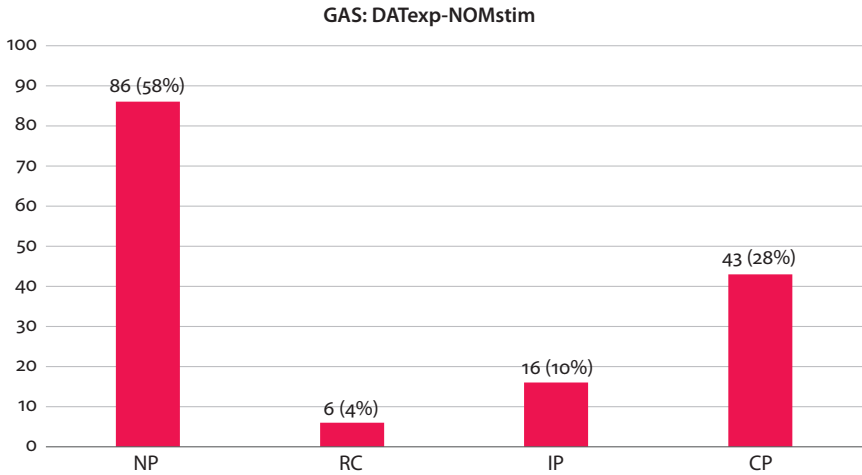
GAS: DATexp-NOMstim

Chart 5. Syntactic types of stimuli with DATexp-NOMstim constructions in GAS
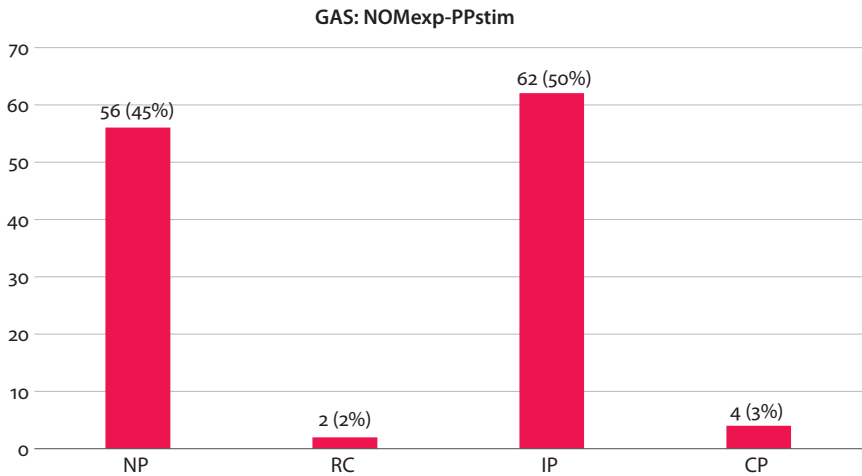
GAS: NOMexp-PPstim

Chart 6. Syntactic types of stimuli with NOMexp-PPstim constructions in GAS

express the idea of liking certain situations, and, because these situations often appeared as finite subordinate clauses, the DATexp-NOMstim schema had to be employed.

Additionally, nominal stimuli usually co-appeared in the DATexp-NOMstim construction during the Golden Age. The verb forms that instantiated this construction included the verb-object compounds *dar gusto* (which displays a causative structure) and *ser gusto/ser agradable*, and the lexical verb *agradar*, as illustrated in (17).

(17)   a.   *Procur-ó   Doña Leonor   dar   gusto   a   su   amante*
               try-PST.3SG Miss Leonor.NOM give.INF pleasure to.DAT her lover
               'Miss Leonor tried to please her lover'   (17th c., *NAE*)

    b.   *Anselmo le   replic-ó   que   aquél era*
               Anselmo PRON.DAT.3SG answer-PST.3SG COMPL that   be.PST.3SG
               su   gusto
               his.POSS liking
               'Anselmo answered that that was his liking'   (17th c., *DQ*)

    c.   *Parec-e   que   no   le   agrad-a-n tus   palabras*
               seem-PRS.3SG COMPL NEG PRON.DAT.3SG like-PRS-3PL your words.NOM
               'It seems that your words do not please him'   (16*th* c., *La Diana*)

The following five verb forms may have also aided, via constructional [analogical] extension, in the change in argument structure from nominative experiencer *gustar* to dative experiencer *gustar*, as I have mentioned in previous sections: *dar gusto* and *ser gusto* through lexical and semantic similarity, and *placer*, *agradar*, and *ser agradable* through semantic similarity.

5.   Conclusion

In medieval Spanish, the DATexp-NOMstim construction had a higher token and type frequency than the less productive DATexp-PPstim, and thus, the former attracted the latter. These two constructions, all throughout the Middle Ages, were in complementary distribution: while the DATexp-NOMstim schema could appear with all types of stimuli (noun phrases, infinitival clauses, relative clauses, and subordinate finite clauses), the DATexp-PPstim schema was precluded from appearing with subordinate finite clauses as stimuli. This shows that the subcategorization properties of prepositions were grammatically relevant, determining the use of one argument structure over another.

A similar pattern arose with experiential constructions of 'liking' in Golden Age Spanish. Two main schemas were used, namely, the DATexp-NOMstim and the NOMexp-PPstim constructions. At this time, however, prepositions still could not easily subcategorize subordinate finite clauses headed by *que*. Because the NOMexp-PPstim construction hindered the presence of [fin-CP que] and had a lesser degree of internal consistency and a lower type and token frequency than the DATexp-NOMstim schema, it was eventually dispreferred.

To conclude, the present usage-based view of productivity has brought to light the fact that the DATexp-NOMstim construction, due to its high token frequency and higher type frequency than other 'liking' constructions, and ampler

subcategorizing properties with regard to the type of syntactic stimuli it can appear with, has attracted other less productive schemas to it, such as the DATexp-PPstim or the NOMexp-PPstim construction. Studying the change in argument structure while taking into account the syntactic development of complementizers in Spanish has highlighted the fact that not only semantic and frequency factors but also syntactic factors determined the morphosyntactic realization of a given construction and triggered processes of variation and change in argument structure.

Abbreviations

ABL	ablative		NOM	nominative
ACC	accusative		NP	noun phrase
COMPL	completive		P	patient-like argument of canonical transitive verb
COND	conditional		PL	plural
CP	complimentiser phrase		POSS	possessive
DAT	dative		PP	past participle
DET	determiner		PRON	pronoun
EMS	early medieval Spanish		PRS	present
F	feminine		PST	past
FUT	future		PTCP	participle
GAS	golden age Spanish		RC	relative clause
GERUND	gerundive		REL	relative
INF	infinitive		RFL	reflexive
IP	infinitival phrase		SG	singular
LMS	late medieval Spanish		SBJV	subjunctive
M	masculine		TV	transitive verb
NEG	negation, negative		1	first person
			3	third person

References

Allen, C. (1995). *Case marking and reanalysis: Grammatical relations from Old to Early Modern English*. Oxford: Clarendon Press.

Amaral, P. & Delicado Cantero, M. (2018). Subcategorization and change: A diachronic analysis of *sin embargo* (*de que*). In J. E. Macdonald (Ed.), *Contemporary trends in Hispanic and Lusophone linguistics. Selected papers from the Hispanic Linguistics Symposium* 15. (pp. 31–48). Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/ihll.15.03ama

Barðdal, J. (2008). *Productivity: Evidence from case and argument structure in Icelandic*. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/cal.8

Barðdal, J. (2009). The development of case in Germanic. In J. Bardðal & S. Chelliah (Eds.), *The role of semantic, pragmatic, and discourse factors in the development of case* (pp. 123–159). Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/slcs.108.09bar

Barra Jover, M. (2002). *Propiedades léxicas y evolución sintáctica. El desarrollo de los mecanismos de subordinación en español*. La Coruña: Toxosoutos.

Batllori, M. (2012). Diacronía de los verbos psicológicos: una propuesta de entrada léxica. In G. Clavería, M. Freixas, M. Prat, & J. Torruella (Eds.), *Historia del léxico: perspectivas de investigación* (Vol. 47, pp. 341–374). Madrid/Frankfurt: Iberoamericana/Vervuert.

Bauer, B. (2000). *Archaic syntax in Indo-European. The spread of transitivity in Latin and French*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110825992

Beardsley, W. (1921). *Infinitive constructions in Old Spanish*. New York: Columbia University Press.

Bencini, G. & Goldberg, A. (2000). The contribution of argument structure constructions to sentence meaning. *Journal of Memory and Language*, 43(4), 640–651. https://doi.org/10.1006/jmla.2000.2757

Bogard, S. & Company Company, C. (1989). Estructura y evolución de las oraciones completivas de sustantivo en el español. *Romance Philology*, 43(2), 258–274.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, 10(5), 425–455. https://doi.org/10.1080/01690969508407111

Chomsky, N. (1981). *Lectures on Government and Binding: The Pisa lectures*. Berlin: Mouton de Gruyter.

Corominas, J. (1954). *Diccionario crítico etimológico de la lengua castellana*. Madrid: Gredos.

Delicado Cantero, M. (2013). *Prepositional clauses in Spanish. A diachronic and comparative syntactic study*. Berlin: De Gruyter. https://doi.org/10.1515/9781614510628

Du Cange, C., Henschel, G., Carpentier, P., Adelung, J. & Favre, L. (1610–1688 [1937]). *Glossarium mediae et infimae latinitatis*. Paris: Librairie des sciences et des arts.

Elvira, J. (2006). El desarrollo de la construcción biactancial estativa en español. *Revista de Historia de la Lengua Española*, 1, 45–66.

Fedriani, C. (2014). *Experiential constructions in Latin*. Leiden: Brill.

Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Goldberg, A. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences*, 7(5), 219–224. https://doi.org/10.1016/S1364-6613(03)00080-9

Haspelmath, M. (2001). Non-canonical marking of core arguments in European languages. In A. Aikhenvald, R. Dixon, & M. Onishi (Eds.), *Non-canonical marking of subjects and objects* [Typological Studies in Language, 46] (pp. 53–84). Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.46.04has

Melis, C. (1997). Las emociones, la transitividad y el aspecto. *Anuario de Letras. Lingüística y Filología*, 35, 383–415.

Melis, C. (1998). Sobre la historia sintáctica de gustar. In C. García, F. González, & J. Mangado (Eds.), *Actas del IV Congreso Internacional de Historia de la Lengua Española*: La Rioja, 1–5 de abril de 1997 (pp. 295–306).

Melis, C. & Flores, M. (2013). On the historical expansion of non-canonically marked 'subjects' in Spanish. *The Diachronic Typology of Non-canonical Subjects. SLCS*, 140, 163–184. https://doi.org/10.1075/slcs.140.08mel

Melis, C. & Flores, M. (forthcoming). *The dative experiencer of Spanish gustar*. Mexico City: Universidad Nacional Autónoma de México.

Serradilla Castaño, A. (1995). Sobre las primeras apariciones de construcciones preposicionales ante *que completivo* en español medieval. Factores determinantes. *EPOS*, XI, 147–163.

Serradillla Castaño, A. (1996). *Diccionario sintáctico del español medieval. Verbos de enten-dimiento y lengua*. Madrid: Gredos.

Serradilla Castaño, A. (1997). *El régimen de los verbos de entendimiento y lengua en español me-dieval*. Madrid: Ediciones de la Universidad Autónoma de Madrid.

Tarr, F. C. (1922). Prepositional complementary clauses in Spanish with special reference to the works of Pérez Galdós. *Revue Hispanique*, 56, 1–264.

Schulte, K. (2004). Pragmatic causation in the rise of the Romance prepositional infinitive: a statistically-based study with special reference to Spanish, Portuguese and Romanian. Ph.D. Dissertation, University of Cambridge.

Schulte, K. (2007a). *Prepositional Infinitives in Romance: A Usage-Based Approach to Syntactic Change*. Frankfurt: Peter Lang.

Schulte, K. (2007b). What causes adverbial infinitives to spread? Evidence from Romance. *Language Sciences*, 29, 512–537. https://doi.org/10.1016/j.langsci.2007.05.004

Vázquez Rozas, V. & Rivas, E. (2007). Un análisis construccionista de la diacronía de *gustar*. In I. Ibarretxe-Antuñano, C. Inchaurralde & J. Sánchez-García (eds.), *Language, mind, and the lexicon* (pp. 143–164). Frankfurt, Peter Lang.

Williams, E. (1981). Argument structure and morphology. *Linguistic Review* 1, 81–114.

Appendix 1. Corpus of Spanish texts (13th to 17th centuries)

Century	Year	Title	Author	Abbreviation	Number of words in text	Source
XIII	1100 [?]-1210 [?], copied in 1301 ca. – 1350 ca.	El Cantar de mio Cid	Anonymous	CMC	31930	www.cervantesvirtual.com
	1251, copied in 1401 ca. [?] – 1410 ca. [?]	Calila y Dimna	Anonymous	CD	78926	www.cervantesvirtual.com
	1253, copied in 1429-08-12 a quo – 1500	Sendebar	Anonymous	Sendebar	15887	www.cervantesvirtual.com
	1246 a quo – 1252 ad quem	Milagros de Nuestra Señora	Gonzalo de Berceo	MNS	27595	www.cervantesvirtual.com
XIV	1335	El Conde Lucanor	Don Juan Manuel	CL	77356	www.cervantesvirtual.com
	1300–1305	Libro del Cavallero Zifar	¿Ferrand Martínez?	CZ	54087	www.cervantesvirtual.com
XV	c.1400	Libro de don Tristán de Leonís	Anonymous	DT	112849	Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings; Prepared by John O'Neill. Madison & New York 1999
	1406	Embajada a Tamorlán	Ruy Fernández	ET	83437	www.cervantesvirtual.com Textos
	1418	Las siete edades del mundo	Pablo de Santa María	Siete edades	51203	Lemir. Edición y estudio a cargo de Juan Carlos Conde 1997

Century	Year	Title	Author	Abbreviation	Number of words in text	Source
	1438	El Corbacho	Alfonso Martínez de Toledo	Corbacho	76126	Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings; Prepared by John O'Neill. Madison & New York 1999, Escorial: Monasterio h.III.10. Transcribed by Eric W. Naylor
	1438–1445	Triunfo de las donas y cadira de onor	Juan Rodríguez del Padrón	Triunfo	23119	www.cervantesvirtual.com
	1497 a quo [?] – 1499 ad quem	La Celestina	Fernando de Rojas	LC	68506	www.cervantesvirtual.com
XVI	1552	Historia general de las Indias	Francisco López de Gómara	HGI	293868	www.cervantesvirtual.com
	1559	Los siete libros de la Diana	Jorge de Montemayor	Diana	73086	www.cervantesvirtual.com
XVII	1604	Segunda parte de la vida de Guzmán de Alfarache, atalaya de la vida humana	Mateo Alemán	GA	138132	www.cervantesvirtual.com
	1605	El ingenioso hidalgo don Quijote de la Mancha	Miguel de Cervantes	DQ	187087	www.cervantesvirtual.com
	1638	Novelas amorosas y ejemplares	María de Zayas	NAE	111651	www.cervantesvirtual.com
	1651–1657	El Criticón	Baltasar Gracián	Criticón	212511	www.cervantesvirtual.com
	c. 1605	Peribáñez y el comendador de Ocaña	Lope de Vega	Peribáñez	18205	www.cervantesvirtual.com
	1613	El perro del hortelano	Lope de Vega	El perro	20577	www.cervantesvirtual.com

Century	Year	Title	Author	Abbreviation	Number of words in text	Source
	1613	La dama boba	Lope de Vega	La dama	19155	www.cervantesvirtual.com
	1614	Fuenteovejuna	Lope de Vega	Fuenteovejuna	14486	www.cervantesvirtual.com
	c. 1620	El caballero de Olmedo	Lope de Vega	Olmedo	16093	www.cervantesvirtual.com
	1631	El castigo sin venganza	Lope de Vega	Castigo	18623	www.cervantesvirtual.com
	1634	Las bizarrías de Belisa	Lope de Vega	Belisa	15385	www.cervantesvirtual.com
	c. 1600	El conde Alarcos	Guillén de Castro	Alarcos	16283	www.cervantesvirtual.com
	c. 1605	Las mocedades del Cid	Guillén de Castro	Mocedades	16969	www.cervantesvirtual.com
	c. 1606	El curioso impertinente	Guillén de Castro	Curioso	18304	www.cervantesvirtual.com
	c. 1612	El Narciso en su opinión	Guillén de Castro	Narciso	16499	www.cervantesvirtual.com
	1629	La dama duende	Calderón de la Barca	Dama	16502	www.cervantesvirtual.com
	1634	Mañanas de abril y mayo	Calderón de la Barca	Mañanas	15533	www.cervantesvirtual.com
	1636	El alcalde de Zalamea	Calderón de la Barca	Alcalde	16344	www.cervantesvirtual.com
	Published in 1657	Darlo todo y no dar nada	Calderón de la Barca	Darlo	21088	www.cervantesvirtual.com
	Published in 1691	Amado y aborrecido	Calderón de la Barca	Amado	20452	www.cervantesvirtual.com

Alignment & Diathesis

The actualization of new voice patterns in Romance

Persistence in diversity

For Romano Lazzeroni

Michela Cennamo

University of Naples Federico II

This article discusses some aspects of the reorganization of voice distinctions in the transition from Latin to Romance, namely the grammaticalization of activity (DO / MAKE) and change of state (BECOME) verbs as markers of the passive voice, and the reanalysis of the reflexive morpheme SE as a voice modulator, focussing on patterns of invariance (i.e., persistence) of Latin inheritance and principled differences (i.e., divergence) in the type and extent of variation and further developments in this area of Romance morphosyntax.

Keywords: voice, passive, reflexive, light verb, auxiliary

1. Introduction

This article addresses the actualization of a number of new structures which became available for the encoding of voice in the transition from Latin to (Italo-)Romance, exploring the various stages in their development and implementation, in relation to (i) the role played by the paramenters of animacy, aspect, the continuum of control,[1] and the person hierarchy in the reorganization of the voice domain and the rise of different passive and impersonal strategies, including passive auxiliaries (e.g., DO / MAKE, BECOME), the reflexive passive, and the impersonal / indefinite SE, (ii) the direction(ality) of the changes investigated.

The discussion is organized as follows. Section 2 discusses the major changes affecting the encoding of voice and of the argument structure of the clause in the passage from Latin to Romance. Section 3 focuses on the Latin antecedents of

---

1. The term refers to the degree of 'primary responsibility' (Lakoff, 1977) of a core argument of the clause over the verbal process, involving various transitivity parameters, such as agency, volitionality, and the aspectual nature of the predicate (Lehmann, 1988: 56–61; Comrie, 1989: 61–62).

nascent auxiliaries *fieri* 'become' and *facere* 'do / make', while Section 4 tracks the early Italo-Romance outcomes of these developments, witnessing the persistence of constructions and parameters of Latin inheritance and principled differences in the distribution of *fieri* and *facere* as imperfective and perfective passive markers, respectively. Section 5 explores the Italo-Romance developments of the reflexive as a voice marker, showing that its impersonal / indefinite usage reflects the widening of its referential scope and its degree of grammaticalization. Finally, Section 6 draws the conclusions.

2.   Voice in the transition to Romance

Four major changes take place in the voice system in the transition from Latin to Romance:

1. the demise of the synthetic passive (i.e., the *-R* form);
2. the spread of the analytic passive, consisting of a form of *esse* 'be' + the P(ast) P(article) of the lexical verb, from perfective to imperfective function, typically attested in the present indicative / subjunctive (Winters, 1984, int. al.), with rare and controversial early examples (Herzog, 1910: 102–106; Bassols de Climent, 1948: § 40; Adams, 2013: 719–2). This structure ultimately replaces the synthetic passive, the *-R* form, that disappears in Romance (e.g., *laud-a-tur* praise.MPASS.PRS.IND.3SG 'he is praised' > *laud-a-tus est* praise.PP.M.SG.NOM be.PRS.IND.3SG 'he is praised').[2]
3. the grammaticalization of lexical verbs denoting activity (e.g., *facere* 'do, make'), change of state (e.g., *fieri* 'become', *evenire* 'befall'), and change of location (e.g., *venire* 'come') as passive markers (auxiliaries) / light verbs (Reichenkron, 1933; Green, 1982; Michaelis, 1998: 73–76; Adams, 2013: 721–724; Cennamo, 2019);
4. the proliferation of the reflexive morpheme in all voice domains, the rise of the reflexive passive and changes in the referential domain of SE (Cennamo, 1991b, 1993b, 1998, 1999).

The present study focuses on some aspects of the latter two of these changes, the development of the passive periphrases *fieri* 'become' / *facere* 'do/make' + past participle (§ 3–4) and the passive and impersonal/indefinite patterns with SE (§ 5). These

---

2.   The perfective-resultative meaning of the sequence *esse*+PP, conversely, comes to be conveyed in Romance by double compound forms which did not exist in Latin (**sum status laudatus*, lit. '(I) am been praised'), supplied by the verb *stare* 'stand' in Italian, French, and Catalan, *sedere* 'sit' in Spanish and Portuguese (Green, 1982; Pountain, 1982: 147; 151; De Melo, 2012; Cennamo, 2016, 2020; Danckaert 2016, further references therein and § 2.1).

constructions will be investigated in light of the *temporary loss of the grammatical dimension of voice* resulting from the reorganization of the marking of voice distinctions in the passage to Romance (§ 2.1) and in relation to *the restructuring of argument linking and marking* in the transition between Late Latin and Early Romance (§ 2.2). The two phenomena came to intersect with the aspectually-determined cleavage existing in the Latin verbal system (both in the active and the passive voice) between forms expressing imperfective aspect (denoting ongoing, continuous, repetitive actions) and forms expressing perfective aspect (referring to a completed action), so-called *infectum* and *perfectum* in Latin grammars (Vincent, 1988; Pinkster, 2015: 231–305, int.al.), their interplay causing a reshuffling leading to the rise of new constructions for the encoding of voice, aspect, and the argument structure of the clause (Cennamo, 2005, 2008, 2009 and discussion below).

In the passive voice in Early and Classical Latin the aspectual split is instantiated by the use of an inflectional ending, the *-R* form – added to the verb stem – for the tenses of the *infectum* (e.g., *laud-a-tur* 'he is (being) praised), and by a syntactic construction – consisting of a form of *esse* 'be' + PP – for the tenses of the *perfectum*, as shown in Table 1 for the present/imperfect and the perfect/pluperfect indicative, respectively.

Table 1.

SYNTHETIC (*infectum*)	ANALYTIC (*perfectum*)
amatur (present) 'he is (being) loved'	*amatus est* (perfect) 'he was loved/ 'he is beloved (adj.)'/ 'he has been loved'
amabatur (imperfect) 'he was (being) loved'	*amatus erat* (pluperfect) 'he had been loved/ he was beloved (adj.)'

Thus, a perfect passive pattern such as *amatus est* love.PP.M.SG.NOM be.PRS.IND.3SG could refer both to a past event ('he was loved', past perfective reading) and to the current relevance of a past event ('he has been loved', perfective-resultative meaning), as well as having a stative-adjectival reading ('he is a loved man'). As illustrated in Table 1, in the analytic tenses, the present auxiliary is used for the perfect, the imperfect for the pluperfect, the tense of the pattern being marked by the construction as a whole (the past participle signalling anteriority), unlike in Romance, where the tense of the structure is conveyed by the auxiliary (Woodcock, 1959: 79; Winters, 1984: 446; 451 and more recent discussion in De Melo, 2012; Burton, 2016; Dankaert, 2016, 2017, among others).

As shown in the next sections, in Late Latin and early Italo-Romance the voice system is radically restructured, but the aspectual distinctions originally shaping its encoding at earlier stages in the language – both compositional, reflecting the

divide between imperfective and perfective aspect, and lexical, with the inherent temporal properties of verbs (i.e., *Aktionsart*) determining the distribution of voice forms in the anticausative alternation[3] – continue to play a crucial role (together with the notions of control and animacy) in molding the use and distribution of the new strategies which become available, among which the reflexive pattern and various periphrases conveying different tense-aspectual nuances within the aspectual spectrum originally covered by the *esse* + PP construction. In addition, a new parameter, the person hierarchy, comes to impinge on the coding of voice alternations (e.g., the reflexive/middle construction) (§ 5) (Cennamo, 2005, 2009, 2016).

2.1 Restructuring of the voice system and changes in the referential domain of SE in Late Latin

By the end of the fourth century AD, the uncertainties in the use of voice forms, already existing in Early and Classical Latin, gradually spread, affecting not only the periphery of voice categories, where there had always been areas of fluctuation among voice forms (active / reflexive / -*R* form / *esse* 'be'+ PP), but also their core (Cennamo, 1998). In later centuries, additional constructions began to emerge, alternating with the canonical markers of voice (Cennamo, 1998, 2006; Herman, 2002 and references therein). The following sections will explore the nature of this restructuring, beginning with the extension of the reflexive and the active into the realm of the passive (§§ 2.1.1–2.1.2), the passive into that of the active (§ 2.1.3), the concomitant re-organization of argument structure (§ 2.2), as well as non-voice related changes involving the referential domain of the reflexive. As we shall see in subsequent discussion (§§ 3–5), the cluster of phenomena investigated all conspire, in the course of time, towards co-opting new tools for the marking of voice, such as the verbs *fieri* and *facio* as new auxiliaries in passive constructions (§ 3–4) and the reflexive as a passive and impersonal(/indefinite) marker (§ 5).

2.1.1 *Reflexives*
In late Latin, the continuum of control, affecting the distribution of voice forms in Early and Classical Latin (Ronconi, 1968: 20; Cennamo, 1998; Cennamo, Eythórsson & Barðdal 2015), no longer differentiated the use of the -*R* form vs. the reflexive pattern with animate subjects, as shown in (1a-b) from Classical Latin. *Se* + the active voice, therefore, could now also occur with inactive / non-agentive

---

3. The term "anticausative" refers to intransitive patterns derived from corresponding transitive structures, with the original inanimate object occurring as subject and the verbal process presented as taking place spontaneously (Haspelmath, 1987: 7; Koontz-Garboden, 2009, int.al.).

subjects, in non-canonical middle / endo-reflexive patterns, structures where the verb's core arguments have identical reference and the action 'remains' within the actor (the A argument)[4] (Haspelmath, 1987: 27), undergoing the verbal activity, rather than performing it. This new flexibility is illustrated in (2), where the reflexive structure, *ustulant se* (lit. 'burn themselves'), co-occurs with the canonical *-R* form, *ustulantur* (lit. 'burn/get burnt') (Cennamo, 1998: 79; 88–89; Cennamo, Eythórsson & Barðdal 2015: 699).

(1) a. *interea* **nos** *delectabimus*
        meantime we.ACC enjoy-ourselves.MPASS.PRS.IND.1PL
        'Meantime we shall organize our own pleasure'
                                              (Cic. De Finibus I, 3 (+ Control)

     b. *si* **delectamur** *cum scribimus*
         if enjoy-ourselves.MPASS.PRES.IND.1PL when write.PRS.IND.1PL
         'If we like writing (lit. we find enjoyment when we write)'
                                              (Cic. Ad Attic. II, 4, 2) (– Control)

(2) *cum male sibi senserint,* **ustulant** *se foco in*
     when ill RFL feel.PRF.SBJV.3PL burn.PRS.IND.3PL RFL fire.ABL in
     stomacho quomodo caballi furiosi **ustulantur**
     stomach.ABL like horses.NOM mad.PL.NOM burn.MPASS.PRS.IND.3PL
     'When they fall ill, they burn with fire in their stomach like mad horses burn'
                                              (Anthim. 3, 6–8)

By this time, with early examples in texts from the Imperial age (first to early second century AD), the reflexive pattern (*se* + active verb) is also found in anticausative function with verbs that in Classical Latin only admitted the *-R* form and/or the active intransitive, such as degree achievements (Hay et al. 1999) / gradual completion verbs (Bertinetto & Squartini 1995), verbs denoting the gradual approximation to a telos that may not be achieved, characterized by a lower degree of telicity (e.g., *minuere* 'decrease', *ampliare* 'enlarge', *lenire* 'soothe', *sedare* 'calm down') (3a) vs. (3b) (Feltenius, 1977; Cennamo, Eythórsson & Barðdal 2015). By contrast, achievements (e.g., *scindere* 'crack', *frangere* 'break') admitted instead both the *-R* form and the reflexive strategy, as shown in (3c–d) (see Feltenius, 1977; Adams, 2013 686-692; Cennamo, Eythórsson & Barðdal 2015: 685–687 for further discussion and data).

(3) a. **minuente** *se morbo*   (Plin. Nat. 23, 50) (I AD)
        decrease.PRS.PTCP.ABL RFL desease.ABL
        'When the disease is on the decline'

---

4. S, A, O are syntactico-semantic primitives, referring to the sencence core arguments, following a well established terminology (Haspelmath, 2011 and references therein).

    b.   *memoria*   **minuitur**   (Cic. *Sen.* 7, 21) (Classical Latin)
        memory.NOM decrease.MPASS.PRS.IND.3SG
        'Memory is impaired'

    c.   **frangitur**   *aestus*   (Lucr. *De Rer. Nat.* 6, 121)
        break.MPASS.PRS.IND.3SG tide.NOM
        'The rolling tide breaks'

    d.   *dum* **se** *calor*   **frangat**   (Cic. *De Or.*, 1, 62, 265)
        until RFL heat.NOM break.PRS.IND.3SG
        'Until the heat abates' (lit. 'breaks')

The reflexive pattern also occurs, albeit rarely and late, in passive function, initially attested with inanimate subjects, as in (4a), from the second half of the fourth century AD, subsequently with animate ones, as in (4b), from the sixth century AD.

(4) a.   *stercora si* **se** post   *ex*   *aggravatione stercoris*
        faeces.PL if RFL afterwards from.ABL oppression.abl faeces.GEN
        provocaverint   (*Mul. Ch.* 230)
        cause-3PL-PRF.SBJV.3PL
        'Faeces, if afterwards they are induced from their weight'

    b.   *per ista sunnis* …*homo* …**excusare se**   **poterit**
        for this amount man   excuse-INF RFL can-FUT.3SG
        'For this amount of money a man can be discharged' (lit. 'excused / justified')
                                          (*Child.reg.cap.* 8, 32–33)

The trigger of the passive reinterpretation of the sequence *se*+active verb lies in a change in the aspectual classes of verbs allowing the *se*+active V in the anticausative alternation, which spread from achievements and accomplishments, i.e., inherently telic and / or punctual verbs denoting change of state (i.e., situations that can be brought about spontaneously) (e.g., *mutare* 'change', *frangere* 'break', *scindere* 'crack'), to activity verbs, i.e., atelic verbs (e.g., *vocare* 'call', *excusare* 'excuse') (4b), with and without a telic alternant (e.g., active accomplishments[5] such as *vendere* 'sell', *provocare* 'cause'). With these verb classes in Classical Latin the reflexive pattern could only occur in a truly reflexive function (*se vendere* 'sell oneself', *se vocare* = 'call oneself') (see full discussion in Cennamo, 1998, 2001b, 2001c). Thus, *se* + active becomes a marker of induced actions / processes, in which the agent argument is unexpressed (see, however, Adams, 2013: 683 for the anticausative interpretation of (4a).

---

5. The term, adopted from the Role and Reference Grammar classification of predicates (Van Valin, 2005: 31–50), refers to the telic use of activity verbs with definite, referential objects (e.g., Lat. *tres epistuas scripsit* three letters write.PRF.3SG 'He wrote three letters'). Unlike accomplishments (e.g., *mutare* 'change', they denote the sequence of two events rather than a relationship of causation between two events.

2.1.2   *The extension of the active voice into the realm of the passive*

The blurring of voice distinctions, owing to the functional opacity of their morphological and syntactic exponents can also be clearly seen in the gradual extension of the use of the active voice to structures where most typically either the *-R* form or the reflexive occurred in early and Classical Latin, such as the anticausative alternation. Thus, in Late Latin the active intransitive may be found with verbs lexically encoding a final state, like achievements (e.g., *rumpere* 'break') (5a) and accomplishments (e.g., *citare* 'cause') (5b) (Cennamo, Eythórsson & Barðdal 2015: 695). The active is also found with non-canonical middles (5c), patterns with an animate, affected subject. These structures may alternate with the corresponding reflexive forms, often in the same text, as in (6a–b). At times a construction is ambiguous between an anticausative and a passive interpretation, which is only resolved by the context, as in (5d–e) and (6c) (Svennung, 1935: 462; Hofmann, Leumann & Szantyr, 1965: 296; Feltenius, 1977; Cennamo, 1998):

(5)   a.   *collectiones… quae **rumpunt***   (*Mul. Chir.* 42)
        abscesses   that break.PRS.IND.3PL
        'Abscesses that break'

   b.   *aut marmur si ei   **citaverit***   (Mul. Chir. 606)
        or   cancer if it.DAT cause.PRF.SBJV.3SG
        'Or if it developed cancer of the joints'

   c.   *cum   … non **potest**   **sustinere***   (id. 910)
        when … not can.PRS.3SG stand-up.INF
        'When … it cannot stand up'

   d.   si *vulnus **elimpidaverit***   (id 645)
        if wound cleanse-PRF.SBJV.3SG
        'If the wound cleansed / was cleansed'

   e.   ***equa…   si non purgavit***   (id. 764)
        mare-NOM if not purge-PRF.IND.3SG
        'The mare, if it didn't purge itself / get purged'

(6)   a.   *si iumento   genua   marmora **se   citaverint***
        if beast-of-burden.DAT knee.PL cancer-PL RFL cause.PRF.SBJV.3PL
        'If the donkey develops cancer of the joints'   (Mul. Chir. 592)

   b.   *si* (sc. *iumentum*) ***se sustinere   potuerit***   (Mul Chir. 986)
        if beast-of-burden RFL stand-up.INF can.PERF.SBJV.3SG
        'If it (sc. the beast of burden) can stand up'

   c.   *mala… toto   anno   **servare**   **se   possunt*** (Pall. 3, 25, 18)
        apples whole.ABL year-ABL keep-PRS.INF RFL may.PRS.IND.3PL
        'The apples may keep / be kept for the whole year'

In late, 6th–8th century texts the active is also found in passive function, as in (7a–b), with an early second-century AD Example (7c) (Bonnet, 1890: 628–30; Haag, 1898: 57; Svennung, 1935: 568).

(7) a. *petens*   *ut   per eius   auxilium **liberaret (=liberaretur)***
         ask.PRS.PTCP that by of-him help   set-free-IMPF.SBJV.3SG
                                          (Chron. Fred. IVc, 183, 17; VII AD)
         'Asking to be set free with his help' (lit. 'to set free')
     b. *consummatio **dabit**   (= **dabitur**)   *super desolationem*
         report.NOM   give.FUT.3SG (= give.PASS.FUT.3SG) over   desert.ACC
         'A report will be given (lit. will give) over the desert'
                                                      (Chron. Fred. I 56.16)
     c. *quomodo aliis   facitis,   sic et   **faciet (=fiet)** vobis*
         how   to-others do-PRES.IND.2PL so also do-3s-FUT   you-2PL.DAT
         'What you do to others *will be done / will happen* to you (lit. the same way
         you do to others so it will do (= it will be done / happen) to you'
                     (*Clem. Epist. Ad Cor.* 13, 2; Svennung, 1935: 568) (early II AD)

Thus, in some Late Latin texts, although continuing to occur in their canonical functions, voice forms may be used interchangeably, as shown in (7c) for the verb *facere* 'do / make' and its lexical passive *fieri* 'be done/be made' and in the examples in (8), where the verb *fieri*, in its expression of indefinite change 'become, arise' (8a), *tumor fit* 'a swelling arises', can be replaced by the reflexive *tumor facit se* (lit 'a swelling makes itself) (8b) and by the active transitive form *tumor facit* (lit. 'a swelling makes') (8c), under the same meaning 'a swelling arises', while also functioning as the pro-passive of *facere*, as in (8d) (cf. Pirson, 1906; Cennamo, 1998; 2001; 2005: 178–179; Herman, 2002):

(8) a. **fieri** 'become / be done / made'   = b. **facere se** 'become'   =
         *tumor **fit***   =   *tumor **facit se***   =
         swelling-NOM become.PRS.IND.3SG   make.PRS.IND.3SG RFL
     c. **facere** 'do / make'
         *tumor **facit***
         make. PRS.IND.3SG
         'A swelling occurs'
     d. *sarda   ita   **fit***   (Apic. 80, 3–5)
         pilchard.NOM   thus make.PASS.PRS.IND.3SG
         'The pilchard is prepared in this way'

In Late Latin, therefore, the dividing lines among the active, the reflexive, and the passive voice are no longer clearcut: the reflexive pattern and the active intransitive may replace the passive form in its non-canonical middle (*iumentum se exulcerat* 'the donkey exulcerates' / *non potest sustinere* 'the donkey cannot stand up'), anticausative

(*tumores se faciunt / faciunt* 'swellings appear'), and passive function (*faciet vobis* 'will be done to you' / *ut liberaret* ('(so that) he was set free') (Cennamo, 1998).

2.1.3   *Passivization and deponentization*

In the same period and often in the same texts displaying 'quirky' uses of the reflexive and active patterns, the passive voice is attested in active function, both in the *infectum* and the *perfectum*, a phenomenon known as "deponentization" (Flobert, 1975, int. al.). This phenomenon is attested with both transitive and intransitive verbs throughout the history of the language, as seen in Classical usage in (9a–b), increasing in Late Latin (10a–c).

(9) a. *qui **punitur** (= **punit**)   aliquem*
         who punish.MPASS.PRS.IND.3SG someone.ACC
         'He who punishes someone'   (Cic. *off.* 1, 25, 88) (Classical Latin)

     b. *epistulam quam   **eram**   **elucubratus***
         letter.ACC that.ACC be.IMPF.IND.3SG write with great labour.PP.M.SG
         'The letter that I had carefully written (lit. 'that I was written')'
                                                 (= **elucubraveram**) (id. *Att.* 7, 19)

(10) a. *et   sabbato non **ieiunantur** (= **ieiunant**)*
         and Saturday not   fast-MPASS.PRS.IND.3PL
         'And they do not fast on Saturdays' (lit. 'they are not fasted')
                                                 (*Per. Aeth.* 27, 1) (Late Latin)

     b. *iste   (servus) postulatus **est** (= **postulavit**) vestram clementiam*
         he.NOM (servant) ask.PP.M.SG be-PRS.IND.3SG   your.ACC mercy.ACC
         'He (the servant) asked for your mercy'(lit. 'he is asked for your mercy')
                                                 (*Agnell.* 165)

     c. *ille   eam   **dugatur** (= **ducat**)   uxorem*
         he.NOM she.ACC take.MPASS.PRS.IND.3SG spouse.ACC
         'He will marry her' (lit. 'he is taken her as his spouse')
                                                 (*Codex Vercell.* Cap. 192)

In some late texts the passive as a strategy at times may be abandoned in favour of the active (Svennung, 1935: 460; Herman, 2002; Hageman, 2006), as witnessed in the *Liber Historia Francorum* (eigth century AD), a chronicle whose anonymous author in some passages replaces the synthetic passive of the original passages taken from Gregory of Tours (sixth century AD). Thus, the passive *celebrari* 'be celebrated' in the original passage from Gregory of Tours (11) is changed into the active *celebrare* 'celebrate' (12) with overt expression of the Agent (*sanctus Mamertus*), which is provided in the text even though it is missing in the original passage (Herman, 2002; also Svennung, 1935: 460; recently Adams, 2013: 718, n. 6 for a critical discussion of the issue and Cennamo, 2016: 968).

(11) (original passage with passive)
    dum missarum **celebrantur** *sollemnia* (Greg.. II, 34 (98, 2)
    while Mass.GEN.PL celebrate.IMPF.PASS.3PL solemnity.PL.NEUT
    'While Masses were celebrated' (lit. 'While the (customary) liturgy of the Mass
    was celebrated')

(12) (new passage with active and overt Agent)
    dum missarum sacrificia ….**celebraret** (sc. *sanctus Mamertus*)
    while Mass.GEN.PL sacrifice.PL.NEUT celebrate.IMPF.ACT.SUBJ.3SG
    'While he (sc. saint Mamertus) celebrated Masses'
                                                (*LHF* 16 (266, 20)) (Herman, 2002)

The Late Latin data, therefore, point to the functional disruption of voice forms
(Cennamo, 1998, 2001a, 2001b; Herman, 2002), with uncertainties in their use
revealing a deep restructuring in the encoding of voice in the spoken language.
However, the functions of the passive forms are only occasionally replaced by pas-
sive verbal periphrases and the *se*+active pattern, since the new voice systems are
not yet organized into coherent paradigms. Whereas the active-as-passive is subse-
quently lost and only testifies to the temporary loss of the grammatical dimension
of voice, the reflexive-as-passive function will evolve into a new voice system in the
transition to Romance (§ 5).

2.2 Equivalences among voice forms and changes in argument linking / marking

Thus, as a result of the equivalences and interchangeability among voice forms,
formally (i.e., morphologically / syntactically), passive structures came to have an
ambiguous interpretation, both in the *infectum* and in the *perfectum*, being sus-
ceptible to a canonical passive reading as well as to an active reading. Conversely,
formally active structures came to occur with a passive function in some texts.

The use of the *passive-as-active* function and of the *active-as-passive* function
indeed signals a violation of the canonical rules for the assignment of grammatical
functions to the arguments of a verb, so-called linking rules: the active morphol-
ogy no longer consistently signals an A argument in subject function, and, vice
versa, the passive morphology no longer always signals an O argument in subject
function, as in canonical active and passive sentences, respectively (Cennamo
2001a, 2001b).

In Late Latin, the ambiguity of interpretation of a passive form (e.g., *ama-
tus est* 'he was loved / he has been loved/ he is a loved man'), therefore, was no
longer confined to the *perfectum*, and it no longer involved only temporal-aspectual

distinctions, as in Early and Classical Latin. Since also the active may be found in passive function, ambiguity now also came to concern the assignment of grammatical functions to the arguments of verbs (O in subject function with active morphology, as in (7a) *petens ut … liberaret* (= *liberaretur*) 'asking to **be set free**', and A/S in subject function with passive morphology, as illustrated in (10c) *ille eam dugatur* (= *ducat*)… *uxorem* ''he **will take** her as his spouse'. When the coding of verbal arguments began to follow an active-inactive and subsequently 'neutral' alignment, as witnessed by the use of the accusative in 'subject' function (to mark initially So (13a), subsequently and more rarely, A participants (13b) (Plank, 1985; Cennamo 2001a, 2001b, int. al.), the ambiguity of the constructions came to affect the identification of the role of verbal arguments (A / O status) (Cennamo 2001a, 2001b, 2009), as witnessed in (13c), from a sixth century Gallic inscription (see Pirson, 1901: 189; and discussion in Cennamo 2011a: 185; 2011b: 304).

(13) a. ***crepitavit***   *panem*   *in furno*   (*Agnell*. 175)
crackle.PRF.IND.3SG bread.ACC in oven.ABL
'Bread crackled in the oven'

  b. *fontem*   *vero*   *… quater* *in anno*   *colorem*
spring.ACC in-fact four-times in year.ABL colour.ACC
mutat
change.PRES.IND.3SG
'The colour of the spring-water changes in fact … four times a year'
(Per. Arth. Excerpta; Rovai 2005: 81)

  c. *Theodovaldo*   *lapide(m)*   *non*
Theodovaldus.DAT / ACC / ABL the-gravestone.ACC not
revolvatur   (Pirson, 1901: 189)
turn-over.MPASS.SUBJV.3SG
'One should not turn Theodovaldus' gravestone (O) over / Theodovaldus' gravestone (S$_O$) should not be turned over')

2.3   Changes in the referential domain of SE

In Late and Medieval Latin, with attestations from as early as the second century AD, the reflexive morpheme was no longer restricted to third person sg./pl. participants only, but it was also found with first person (14a) and second person plural (more rarely also the second person singular) reference, sometimes owing to Greek influence, as in (14b). It was even extended to first person singular reference (14c), with earlier (third to fourth century AD) examples from the inscriptions for this use (Svennung, 1935: 315ff; Cennamo 1993b: 58).

(14) a. ***inter   se   singuli dissimiles invenimur***
          between themselves each   different find.MPASS.PRS.IND.1PL
          'We find ourselves different from each other'
                                              (Min. Fel. 18.15; II AD) (*se = nos*)

 b. ***Attendite   se   ipsis***
         pay-attention.IMPER.2PL RFL yourselves.DAT.PL
         'Pay attention to yourselves'   (Itala, act. 20, 28 [cod. e])

 c. ***Ego*** … *Adhalardus* … ***recogitans   se***
         I   Adhalardus   think.PRS.PTCP RFL
         'I Adhalhardus … thinking …'
                                          (Tardif, N: R. 59; VIII AD, France) (*se = me*)

Thus, in Late Latin SE comes to refer to the Speech Act Participants (SAPs) which
appear to be part of its referential domain already by the third to fourth century
AD, as witnessed by the above examples (see Cennamo, 1991b: 16–17 and 1993b
for the early Italo-Romance continuants of these structures).

In the same period, SE is also found in non-reflexive, (anaphoric) pronominal
function (15), whereby it becomes equivalent to the third person anaphoric-deictic
pronoun *is* 'he' (Cennamo, 1991b: 17; 1993b: 60).

(15) … ***Maura*** *coniux Bonifaci* …,   *quae* … ***fui*** (= *fuit*) ***secum***
          Maura wife   Bonifacius.GEN who   be.PST   RFL.with
       (= cum *eo*) *annos XX*
                        years twenty
       '… Maura … who lived with him for twenty years' (lit. 'was with him')
                                                                          (ILCV 1326)

The above innovations, involving the widening and lexicalization of the referential
domain of the reflexive morpheme as well as its pronominal function, will pave the
way for the acquisition of the impersonal / indefinite function of the continuators
of Latin SE in Italo-Romance, with SE becoming an indefinite pronoun in some
varieties (Cennamo 1993a, 1993b; 2014), a change that is not attested in Late Latin
and that does not fall under the domain of voice and the parameters affecting it (see
also Cennamo, 2014). I intend to explore this issue from a diachronic pan-Romance
perspective in future work (see Cennamo, 2016 for a synchronic overview).

3. Steps in the auxiliarization of lexical verbs in the transition to Romance: *Fieri* and *facere*

One of the outcomes of the deep restructuring in the encoding of voice distinctions and its interaction with the rise of active-inactive / neutral alignment patterns, illustrated in § 2, is the development of new constructions to convey the different temporal-aspectual nuances within the imperfective-perfective spectrum, especially *fieri* 'become'+PP (in imperfective passive function), and *facere* 'do / make' + PP (in perfective passive function) (Cennamo, 2006, 2018). The former is a Late Latin phenomenon, the latter witnesses instead a Romance innovation, only attested in eleventh to thirteenth-century Logudorese Sardinian texts and disappearing in subsequent centuries (Cennamo, 2003, 2018). Both periphrases exemplify the different outcomes of the persistence of the semantic parameters of animacy and aspect, both compositional and lexical, in shaping the voice domain in the passage to Romance.

3.1   *Fieri* 'become' as a voice marker

The passive function of the verb *fieri* 'become' (also used as the lexical passive of the verb *facere* 'do, make' in early and Classical Latin) + the past participle of a lexical verb is well attested in late Latin in fourth-century texts. It is found only in imperfective tenses (present indicative / subjunctive) (16a), alternating with the canonical synthetic (imperfective) passive, the –R form. (16b). The analytic passive BE (*esse*) + PP) is also occasionally found in imperfective function (16c), apparently confined to the present indicative / subjunctive (Svennung, 1935: 457–58; Winters, 1984 and recent discussion in Danckaert 2016, int. al.), sometimes in one and the same text, as shown in (16a-c), from the *Peregrinatio Aetheriae* (second half of the fourth century AD). *Fieri* 'become' as a passive auxiliary is found initially with [−An] subjects (16a), later also with [+An] ones (16d) (ninth century AD) (Svennung, 1935: 459; Hofmann, Leumann & Szantyr, 1965: § 211; Cennamo, 2005, 2006).

(16) a. ***interpositae   orationes   fiunt***   (*Per. Aeth.* 35, 6)
           intersperse.PP.F.PL.NOM prayer.PL.NOM become.PRS.IND.3PL
           'Prayers are being / get interspersed' (lit. become interspersed)
       b. ***interponuntur   orationes***   (*Per. Aeth.* 37, 6)
           interperse-MPASS.PRS.IND.3PL prayer.PL.NOM
           'Prayers are being / get interspersed'

    c. *ut   forsitan… gemitus   populi   omnis **auditus***
      that perhaps the-moan crowd.GEN all.GEN hear.PP.M.SG
      sit   (*Per. Aeth.* 36, 3)
      be.PRS.SBJV.3SG
      'That perhaps the moan of the whole crowd is heard'

    d. ***fiat   reconpensatus.***   (Lex Cur. 442, 31)
      become.PRS.SBJV.3SG reward.PP.M.SG.NOM
      'That he be rewarded'

In the auxiliarization process of the verb *fieri* the structure of the clause changes and the two predicates, originally two different constituents, merge into one: the lexical verb becomes a tense-aspect marker, and the original complement (e.g., a non-finite verbal complement) becomes the main verb (i.e., the lexical verb), as illustrated in (17) (Heine, 1993: 53; Harris & Campbell, 1995: 172–173).

(17) *carnes* [***coctae***]   [***fiunt***]>   [***coctae   fiunt***]
       meat-PL cook.PP.PL.F become.3PL cook.PP.PL.F become.3PL
       [complement] [lexical verb] > [lexical verb aux]
       'Meat becomes / is / gets cooked'

Two different and at some point converging paths or 'chains' can be identified: an initial stage in which the verb *fieri* and the copula *esse* can be used interchangeably (Svennung, 1935: 460), with *fieri* coming to denote a quality that is not the outcome of a process, as shown in (18), and a subsequent step, where the copula *fieri* gets integrated into the verbal paradigm, a process referred to in the literature as 'extension' (Andersen, 2001: 230) or 'expansion' (Heine & Reh 1982: 38–39), becoming a TAM marker, with O orientation (i.e., a passive auxiliary), so-called *copula auxiliarization* (Dik, 1987: 57; Cennamo, 2019 for a recent discussion).

    The change is well attested in fourth-century texts (e.g., *Mulomedicina Chironis*) and even more so in later texts, as shown in (18), from the tenth century AD (Cennamo, 2005, 2006), continuing in early Romance vernaculars (Cennamo, 2005, 2006 and § 4):

(18) ***fiat***   *ei   fermum et   stabilem*   (*Cod. Cajet. 906*)
       become-PRS.SBJV.3SG he.DAT firm-ACC and stable-ACC
       'That it be firm and stable to him   (De Bartholomaeis, 1902: § 73)

At a subsequent stage, a change occurs in the nature of the complement, from adjective / noun (19a) to verbal adjective (19b) and then verbal participle (19c):

(19) a. *tumor durus fit* ('the swelling becomes hard') > b. *caro cocta fit* ('meat becomes / is / gets cooked') > c. *provincia gubernata fit* 'the territory gets governed'

Stage (b), where the past participle is formed from an accomplishment verb (e.g., *coquere* 'cook') is ambiguous between a two-constituent and a one-constituent analysis, owing to the possible ambiguity of the past participle (adjectival vs verbal). This ambiguity arises from the aspectual nature of the lexical verb from which the past participle is formed (e.g, accomplishment *coquere* 'cook'), and is resolved by the syntactic context. In contrast, no ambiguity arises when the past participle is formed from an activity verb (e.g., *gubernare* 'control, govern', which only has a verbal function) (see also Cennamo, 2005, 2019).

3.2   Latin antecedents of *facere* 'do, make' passive

In late Latin the verb *facere* 'do, make' also occurs in copular function, as shown in (20a–b):

(20) a. *remedium   optime   facit…   si*
          remedy.NEUT excellent.NEUT make.PRS.IND.3SG if
          'It makes (=is) an excellent remedy if…'   (Colum.60,15,1) (I AD)

      b. *bonum   aerem   facit*
          good.M.ACC weather.M.ACC make.PRS.IND.3SG
              (*Vitae Patrum* 5, 11, 51; Hofmann, Leumann & Szantyr, 1965: § 416)
          'It is good weather' (lit. '(it) makes good weather')

Facere is also found in patterns where it is equivalent to the verb *fieri*, in all its uses, as illustrated in (21a–b), exemplifying, respectively, the fientive and intransive usage (see also Meyer-Lübke, 1902: 51–52).

(21) a. *lacrimosum oculum faciet…   (= fiet)   et*
          tearful.ACC eye.ACC make.FUT.3SG become.FUT.3SG and
          extumidior (fientive) fit
          swollen.COMP.NOM   become.PRS.3SG
          'Its eye will become (lit. will-make) tearful… and becomes more swollen'
                                                          (*Mul. Ch.* 70)

      b. *tumores   in capite   faciunt   (= fiunt)*
          swelling.NOM.PL on head.ABL make.PRS.3PL become/arise.PRS.3PL
          'Swellings appear (lit. make) on its head'

An analogous equivalence *fieri/facere* can also be hypothesized for the auxiliary use of *fieri* in anticausative / passive function (22a), with *facere* replacing *fieri* in a pattern such as *coctus fit* 'it cooks/is cooked', as shown in (22b) *coctus fieret*, (already found in texts from the Imperial age, e.g., Petronius, *Satyricon*, first century AD), and the corresponding intransitive form of the type *facere* + predicative past

participle, exemplified in (22c) *vitulos … coctos solent facere*, modelled on a fairly common use of the past participle in predicative function in early Latin. These usages resemble light verb patterns, similar to those formed with transitive (causative) verbs such as *facere* 'do, make', *curare* 'care for, attend to' (e.g., *aliquem missum facere* someone.ACC send.PP.ACC.M.SG make.INF 'let someone go' (lit. 'someone sent let')) (Hofmann, Leumann & Szantyr, 1963: § 209, c; see also Svennung, 1935: 459–460, int.al., and recent discussion in Cennamo, 2003, 2018).

(22) a. *?*gallus*   *coctus*   *facit* (=*fit*)
           hen.M.SG.NOM cook.PP.M.PL.NOM make.PRS.IND.3PL
           (<*gallum coctum facit* = *coquit* 'he cooks the hen')
           'The hen cooks/is cooked'

      b. *iussit*   *ut   aeno*   *coctus*
           order.PRF.IND.3SG that bronze-vessel.ABL cook.PP.M.SG.NOM
           fieret (sc. *gallus*)   (Per. 74, 4) (I AD)
           become.SBJV.IMPF.3SG
           'He ordered that it (=the hen) cooked / was (lit. became) cooked in a bronze
           cauldron'   (anticausative / passive)

      c. *mei   coci   etiam vitulos   aeno*
           my.PL.NOM cook.PL.NOM also   veal.PL.ACC bronze-couldron.ABL
           coctos   solent   facere
           cook.PP.PL.M.ACC use.PRS.IND.3PL make.INF
           (*coctos … facere* = *coquere* 'cook')
           (= cook)
           'My cooks cook veal (lit. make veal cooked) in a bronze cauldron'
                                                             (Petr. 47, 25)

A subsequent step would involve the use of *facere* replacing *fieri* with activities, i.e., atelic verbs (e.g., *gubernare* 'govern') as in ?**provincia gubernata facit* province govern.PP.F.SG make.PRS.3SG 'the province is governed (lit. '(the) province governed makes')) (see also Cennamo, 2006: 328–330), in a truly passive function of the sequence *facere* + past participle. This hypothesis, which appears to be confirmed by Old Logudorese Sardinian (§ 4.2), needs further confirmation through a corpus-based investigation of relevant texts of late Latin and early Medieval Latin.

3.3   Grammaticalization paths in the rise of passive periphrases in Late Latin

The Late Latin data investigated here provide evidence for the auxiliarization of the verbs *fieri* and *facere* – the former already completed in the fourth century AD, the latter only incipient by the same period –, as resulting from the interplay of radical changes in the domains of voice and grammatical relations due to the aspectual

cleavage in the Latin verbal system between imperfective and perfective verbal forms (see also Cennamo, 2006).

Both changes involve a stage where *fieri* and *facere* became functionally equivalent to the canonical copula, *esse*. A later, subsequent step, where the copula acquired further grammatical meanings, becoming fully integrated into the tense-aspect-mood and person marking system of the language, the so-called *copula auxiliarization* (Dik, 1987: 57; Cennamo, 2005, 2006, 2019 for late Latin), is well documented for *fieri*, and involves a change in the complement of the verb, from noun / adjective to past participle. The passive interpretation of the pattern was triggered by a change in the aspectual classes of verbs in the participial form, from accomplishments to activities. Unlike *fieri*, the verb *facere* does not occur as a passive auxiliary in Late Latin, but shows desemanticized uses in which *fieri* is equivalent to the copula *esse* 'be' and to the corresponding lexical passive / intransitive *fieri* 'be done/become, arise'.

4. Auxiliaries and voice in Old Italo-Romance

Continuators of late Latin *fieri* 'become' and *facere* 'do / make' in early Italo-Romance (e.g., Old Venetian (*fir*), Old Lombard (*fi*), Old Tuscan (*fir*)) and in Old Logudorese Sardinian (*facere*), exhibit the persistence of the parameters affecting the use of these verbs as voice markers and / or copulas from late Latin.

4.1   **Fire* 'become'+ PP

The construction **fire* + PP (and its variants) in several northern Italo-Romance vernaculars displays the same function as its late Latin antecedent *fieri* + PP, with the verb **fire* occurring as a marker of imperfective passive, attested only as a subjunctive / future in Old Florentine (23a) and a present / imperfect indicative in Old Venetian (23b). In Old Lombard, on the other hand, **fire* is the most common passive auxiliary, alternating with *essere*-passives, and it occurs in all tenses except the past perfect and the gerundive (23c) (Cennamo, 2003).

(23) a. *non ne   **fia**   mai nessuno **ingannato***
          not of-them become.PRS.FUT.3SG ever nobody deceive.PP.M.SG
          'None of them will ever be deceived'   (Old Florentine)
                    (Sacchetti, *Trecentonovelle*, XIV, Pernicone, 1946: LXVII. 85)
      b. *el qual **fi**   **dito**   esser stado   santo homo*
          the which become.PRS.3SG say.PP.M.SG be.INF be.PP.M.SG holy man
          'Who is said to have been a holy man'   (Old Venetian)
                              (*Cronica*, XIV, Ceruti, 1878: 228. 58b)

c. *Tu* *fi'* **metua** *sot* *pei* *e* **fi'**
you become.PRS.2SG put.PP.F.SG under feet and become.PRS.2SG
fagia *morir* (Old Lombard)
make.PP.F.SG die.INF
'You are trampled upon and are made to die'
(Bonvesin, *Disp.*,. XIII, Contini, 1937: 32. 98)

*Fire is also found in copular function in Old Venetian, Old Lombard (17a), Old Tuscan (17b), and other early Italo-Romance varieties.

(24) a. *plu_tost* **firav** *fastidio*
rather be.PRS.FUT.3SG bother
(Bonvesin, XIII, *De quinq. Curial.*, XIII, Contini, 1937: 59, 123)
'It will rather bother him' (lit. 'rather it will be a nuisance')
b. *le* *genti* **fiano** *ora* *tutte* *nuove*
the people be.PRS.3PL now all new
(*Novellino*, XIII, Conte, 2001: XX1. 28)
'The people are now different' (lit. 'the people are now all new')

4.2 *Faker* 'make / do'+PP

In eleventh to thirteenth-century Logudorese Sardinian texts (but not at a later time) the verb *faker* 'make / do' also appears as a passive auxiliary, although only in the third person singular / plural, the perfect (*fekit*: 3rd singular -*fekerun*: 3rd plural) (25a), and the pluperfect (*fekerat*:3rd singular) (25b). It only marks a perfective passive, occurring, most typically, with human subjects. This construction dies out in later centuries, replaced by BE+PP in its double compound participial, *stato*-form) (25c). It shows a type of grammaticalization involving the desemanticization and decategorialization of the lexical verb *faker*, as it turns into an auxiliary (i.e., a TAM marker) (Blasco Ferrer, 1995; Cennamo, 2003, 2007, 2018).

(25) a. *ki* **fekerun** **datos** *a Mariane de Capathennor*
that do.PRF.3PL give.PP.PL.M to Marianus of Capathennor
'that were given to Marianus of Capathennor' (*CSNT* 270, 1)
b. *ca* *non* **fekerat** **pettita** *s'ankilla* *de scu. Petru*
that not had-done ask.for.PP.S.F. the-servant of saint Peter
'Because S. Peter's servant had not been asked for' (*CSPS* 33, 5–6)
c. **est** **istadu** **dadu** *cumandamentu* (*RSPS* 3, 2)
be.PRS.IND.3SG be.PP.M.SG give.PP.M.SG order
'It has been ordered' (lit.'is been given order')

The verb *faker* also occurs as a copula, as in (26a), a function that is also attested in other early vernaculars, such as Old Campidanese (26b).

(26) a. (Gosantine de Thori) … ***fegi   malabitu***   (*CSNT* 218, 2)
          Gosantine of Thori   do.PRF.3SG ill.SG.M
          "Gosantine of Thori was ill"
      b. *Jurgia Cucu … aligando **muniaria non fegit***   (*CV* 13, 10)
          Jurgia Cucu   never   servant   not do.PRF.3SG
          "Jurgia Cucu was never a servant"

Two stages can be identified for the auxiliarization of the continuants of the Latin verb *facere*: (i) the establishment of equivalence between *facere* and the copula *esse*, as seen in other Late Latin passive periphrases (e.g., *fieri / venire* +PP) (Cennamo, 2005, 2019), (ii) the auxiliarization of this copula, becoming a TAM marker with O-orientation, its passive function stemming from a change in the nature of the verbal complement, from noun / adjective (as in its copular / fientive function) to past participle, as illustrated in (27):

(27) **a.** *tumor maturus facit /* **b.** **carnes coctae faciunt*   **c.** *ancilla fekit pettita*
          equus sanus facit   *(=fiunt) (=coquuntur)*   'the servant was
          (=fit=est) 'the swell-   'meat becomes / is / gets   asked for' (lit. 'the
          ing becomes / is soft   cooked' *(*carnes coctae*   servant did / made
          / the horse heals'   *fecerunt (= coctae sunt)) >*   asked for') (O.
                                                                     Logudorese)

Stage (a) results from the equivalence *facere / esse / fieri* and exemplifies the copular and fientive uses of *facere*; at stage (b) (hypothesized) (cf. § 3.2) the pattern would be ambiguous between an anticausative and a passive interpretation (i.e., between a spontaneous and an externally caused reading). The trigger of *faker* as a passive (stage c) appears to be a change in the classes of verbs occurring in the construction, initially confined to telic verbs denoting change of state (e.g., Latin *coquere* 'cook') (i.e., accomplishments) (stage b), with which the pattern may be interpreted either as a spontaneous or as an induced process, depending on the context. Finally, it extends to atelic / non-inherently telic verbs (e.g., activities / active accomplishments), where the spontaneous interpretation (i.e., anticausative function) is impossible, and an external human causer is implied (passive function), as in (27c).

Thus, although similar in their steps and paths of development, the continuators of Latin *fieri* and *facere* end up covering two opposite poles along the temporal-aspectual spectrum, the imperfective and perfective ones, respectively, showing how an original aspectual split continues to operate in Romance, conveyed by different structures, but performing the same functions.

5.   The reanalysis of SE as a voice modulator in early (Italo-)Romance and Old Logudorese Sardinian

Just as the renovated analytic passive structures of late Latin develop into new passive constructions in Romance by means of a process of persistence-through-divergence, so also does the extension of the reflexive morpheme SE to passive function occur in early Italo-Romance; by contrast, its impersonal / indefinite function represents innovative invigoration of ancient categories and developments.

Animacy and aspect, which model the encoding of voice in early Italo-Romance and Old Logudorese Sardinian passive auxiliary formation (§§ 3–4), also play an essential role in the employment of the reflexive strategy in passive and impersonal / indefinite function alongside a new reliance on the person hierarchy, a semantic parameter emerging in some Late Latin uses of the reflexive/middle patterns (§ 2.3), where SE occurs with first and second person reference, thereby coming to include both speaker and hearer (i.e., the Speech Act Participants) in its referential domain. The degree of grammaticalization of this feature manifests itself in the different distribution of impersonal / indefinite SE, as revealed by the investigation of eleventh to fifteenth-century texts from different geographical areas (see Cennamo, 2014 for a thorough discussion and further references). Evidence for this innovation from the early Italo-Romance varieties and from Old Logudorese Sardinian is presented in the following sections.

5.1   Old Venetian

In Old Venetian the reflexive-as-passive function is well attested most typically with inanimate subjects and the order S(ubject) *se* V(erb) (28a), which favours the passive, O-oriented interpretation. This is the only sequence attested in texts such as *Testi Veneziani* of the thirteenth to fourteenth century. Very few examples with animate S are found. The *se* V S sequence favours the impersonal interpretation (28b), the different distribution of S generally reflecting its given vs. new informational status. The use of impersonal/indefinite SE has a low occurrence and is confined to bivalent verbs with a latent object (28c) (see Cennamo, 2000, 2014: 84–85 for a full discussion).

(28)   a.   *voio*   *che le dite*   *chase se*
           want.PRS.IND.1SG that the afore.mentioned houses RFL
           afita[6]   *tute*
           rent.PRS.IND.3 ALL
           'I want all the aforementioned houses to be let' (lit. 'that they rent themselves')
           (*Testi Ven.* 29,v)

---

6.   Already in Old Venetian texts the third person singular / plural have identical endings, a phenomenon characteristic of contemporary Venetian and several Venetan dialects.

b. *avanti che se   paga   sta   dota*
before that RFL pay.PRS.IND.3 this dowry
                    (*Cedola di Marco Michel*, 112.22; Cennamo, 2014: 85)
'Before this dowry is paid/one pays/they pay this dowry' (lit. 'this dowry
pays itself')

c. *…   chom se   leze in la   ecclesiastica ystoria*
CONJ RFL   read.PRS.3SG/ PL in the ecclesiastical history
'… As one reads in the ecclesiastical history' (lit. 'as it reads itself')
                                                      (*Cronica* 234, 64)

In early texts, on the other hand, there are no examples of impersonal *se* with mono-
valent verbs. Such uses appear to be confined to fixed phrases, as with the verbs *con-
venire* 'be advisable', *contenere* 'contain, state', *deser* 'become', sometimes preceded
by expletive *el* 'it' and its variants in Old Venetian as well as in Old Lombard and
Old Florentine, often alternating with non-reflexive forms. In these patterns the
reflexive has a pleonastic function, as in (29) (Cennamo, 2014: 84).

(29) *eno   se coviene   a celar   senpre mai lo*
and-not RFL be-advisable.PRS.3SG/PL to hide.INF always ever the
secreto amor
secret love
'And it is never advisable to hide one's secret love'   (*Panfilo* 229, 753)

SE in impersonal/indefinite function only occurs in simplex tenses, in imperfective
verb forms (present, imperfect indicative / subjunctive / future). It is not attested
with monovalent verbs until the mid- to late- fourteenth century, as shown in (30),
and it always refers to an indefinite human participant which does not include the
speaker (cf. also Wehr, 1995: 106).

(30) *la calle che **se va**   in Riolto Novo*
the street that RFL goes to Riolto new
'The small street through which one goes / people go to Riolto Novo'
                                                  (*Capitolare* 124, 58.18–19)

5.2   Old Lombard

Also in Old Lombard SE is well attested in passive/impersonal function, with [± An]
subjects and both word orders, S *se* V/*se* V S (albeit the S *se* V structure with [−An]
subject is the most frequent one), with varying distribution, according to the area.
The reflexive morpheme may signal an indefinite participant, which may comprise
the speaker, as in (31), from Old Comasco. SE is also found with monovalent verbs,
both unergatives and unaccusatives, in simple (imperfective) tenses only, as in (32),
from an Old Pavese text:

(31) *anchor   quando nu voloma   provar una vigna… **guarderemo nu***
      furthermore when   we want.PRS.1PL assess a   vineyard see.FUT.1PL we
      *s'el' a   longhi filagni… Per lo   semegliante se **contempla***
      if-it have.PRS.3SG long   rows   for the resembling RFL contemplates
      e   guarda   l'oliva   e   gl'altri   arbori fruteveli
      and see.PRS.3SG the-olive-tree and the-others fruit-trees
      'Furthermore when we want to assess a vineyard …we shall see whether it
      has long rows (of vines)… Similarly, one looks at the olive trees and the other
      fruit-trees'   (*Passione* 5, 37)

(32) *ma per   le citae se çeva   pur peçorando*
      but through the city RFL go.IMPF.3SG also get-worse.GER
      'But in the town things were also getting worse'   (Grisostomo 40, 30)

5.3   Old Florentine

In Old Florentine passive/impersonal SE, occurring with [± An] subjects and in
both S *si* V/*si* V S orders, with a preference for the sequence *si* V S (favouring an
impersonal interpretation) with [+An] S) (Cennamo, 2014: 87), is also found in
structures with the overt expression of the Agent, through a prepositional phrase
(33), a construction unknown to Latin:

(33) *venne   al   porto   … lo quale si tenea   per lo*
      come.PST.3SG to.the harbour   the which RFL keep.IMPF.IND.3SG by the
      re   Carlo
      king Charles
      'He arrived at the harbour that had been conquered by King Charles'
                                                    (*Cronica fior.*, 146. rr.10–11)

SE in impersonal/indefinite function is well attested with bivalent verbs (34a), also
with examples with the verb in the 3rd person singular and a non-agreeing post-
verbal nominal, as in (34b):

(34) a. *e   cchosì s' ordinò   e   fece*
          and thus RFL order.PST.3SG and do.PST.3SG
          'And so they (indef.) ordered and did'
                                                    (*Distruzione di Troia* 157, rr. 24-25)
      b. *si dea   soldi xl (40) a' poveri*
          RFL give.SBJV.PRS.3SG denarius 40 to the-poor
          'They (indef.) / one should give forty denarius to the poor'
                                                    (*Compagnia di S. M. del Carmine*, 65, rr. 31–32)

In early thirteenth century Florentine texts, there are no examples of impersonal *si* with monovalent verbs; these become well attested, however, by the end of that century (e.g., Dante) and in the next century (e.g., Boccaccio), with both unaccusatives and unergatives, also with overt expression of the agent, as in (35):

(35)   *non vuol   che'n   sua città per me si   vegna*
       not want.PRS.3SG that-in his town by me RFL come.PRS.SUBJV.3SG
       'He does not want me to come to his town' (lit. 'by me one comes to his town')
                                                        (Dante, *Inf.* I, vv. 125ss)

In Old Tuscan, e.g., in the *Milione* (end of the thirteenth-early fourteenth century), impersonal/indefinite *si* also occurs in compound tenses, alternating with the MAN strategy, *l'uomo* 'the man' (36 a,b) (Wehr, 1995: 156):

(36)   a.   *quando si è   ito   uno die e   una notte, si*
            when   RFL be.PRS.3SG go.PP.M.SG one day and one night RFL
            truova   acqua
            find.PRS.IND.3SG water
            'When one has gone for one day and night, one finds water' *(Milione, 62)*
       b.   *quando l'uomo è   ito   tre   giorni*   (id. 168)
            when   the-man be.PRS.3SG go.PP.M.SG three   days
            'When one has gone for three days'

As for the interpretation of impersonal *si*, it has both a generic and an indeterminate / existential (both exclusive and inclusive) reading and it may refer to both speaker and hearer / interlocutor (37).

(37)   *Demoli …   a Iacopo Lanberti, ke   lisi*
       give.PST.IND-they.ACC to Iacopo Lanberti that they.ACC-RFL
       scontò   de' denari che ci   dee   dare
       deduct.PST.IND.3SG of-the denarius that we.DAT must.PRS.IND.3SG give.INF
       'We gave them (sc. the money) to Iacopo Lanberti, money that was deducted
       from the denarius that he owes us'   (Ricordi: 33 p. 223)

5.4   Old Neapolitan

In Old Neapolitan SE frequently occurs in passive/impersonal function with [± An] subjects and both word orders, S *se* V/*se* V S. It also appears with the overt expression of the agent, introduced by the prepositions *da/per* 'by' (38a), with a prevalence of simple tenses and imperfective verb forms, and a higher frequency with a preverbal S (Cennamo, 2014: 89).

(38) a. *azò   che da tuti li   Troyani se   potessero   (sc. li   cuorpi)*
      so.that that by all the Trojans RFL can.SBJV.3PL   the bodies
      resguardare   (*Destructione de Troya* 230, r. 2)
      watch.INF
      'So that their bodies could be seen by all the Trojans'

In impersonal function, with monovalent verbs SE is mainly found with unaccusatives denoting telic change of location (e.g., *gire* 'go', *andare* 'go', *scendere* 'go down') and, less frequently, unergatives (e.g., *parlare* 'talk').

(38) b. *a quista insula non se nce   poteva   gire   se non per mare*
      to this   isle   not RFL there can.PST.3SG go.INF if not by sea
      'One could only reach this island by sea'   (*Destructione de Troya* 51, r.2)

Impersonal / indefinite SE only occurs in imperfective tenses. As regards its reference, it refers to an unspecified human participant who is not the speaker / narrator; it never alternates with the 1st plural, but only with the 3rd plural, at times the 2nd person singular, displaying either a generic or an indeterminate/existential non-inclusive interpretation (Ledgeway, 2009: 672–674; Cennamo, 2014: 92).

5.5   Old Logudorese Sardinian

There are no examples of impersonal / indefinite reflexives with monovalent verbs in the eleventh to fourteenth-century texts of Old Logudorese Sardinian; these forms are only attested in fourteenth to fifteenth-century texts. There, passive / impersonal *si* also occurs with animate subjects (39a), occasionally in perfective tenses (39b), albeit only with bivalent verbs. With monovalent verbs impersonal /indefinite *si* occurs mainly with unaccusatives like *andare* 'go', in imperfective tenses, and always with a non-referential, non-inclusive interpretation (39c).

(39) a. *qui non si podiat   hobligare sensa licentia   desu*
      who not RFL can.PRS.SBJV.3SG force.INF without permission of.the
      p[re]ladu   (*RSPS* 81, 5)
      priest?
      'Who cannot be force/one cannot force without his priest's permission'
   b. *qui pius non si li   siat   dadu   termen …*
      that ever not RFL he.DAT be.SBJV.3SG give.PP.M.SG temporal.limit
      'That one has never given him the temporal limit …'   (id. 305, 5)
   c. *fina   assa uia per   issa quale se uaet   ad Osilo*
      as.far.as to.the path through it   which RFL go.PRS.IND.3SG to Osilo
      'As far as the path through which one goes to Osilo'   (*SRS*, 8, r.5)

In Old Logudorese Sardinian, therefore, impersonal / indefinite *si* only has a generic, non-inclusive interpretation.

5.6   Summary

Whereas passive / impersonal SE is well attested in all the early varieties investigated – most typically with [−An] subjects, more rarely and at a later stage with [+An] ones, with both S SE V and SE V S order, (according to the given vs new status of S) – impersonal / indefinite SE shows an uneven distribution within and across the varieties, which appears to reflect the gradual acquisition of the existential / indefinite inclusive interpretation of SE in some early Italo-Romance vernaculars, with different degrees of grammaticalization. This process was more advanced in Old Florentine and Old Lombard, where impersonal / indefinite SE could have both a generic and indeterminate/existential (inclusive/ exclusive) reading. By contrast, SE had only a non-inclusive interpretation in Old Neapolitan (both generic and existential) and Old Logudorese Sardinian (where impersonal / indefinite SE, in fact, has only a generic reference). In Old Venetian, on the other hand, the inclusive interpretation of the impersonal / indefinite use of the reflexive morpheme was a subsequent development.

5.7   The origin of impersonal / indefinite SE in Italo-Romance

The existence of a correlation between the use of the reflexive in impersonal / indefinite function and its (inclusive / non-inclusive) interpretation suggests that the spread of the reflexive morpheme in impersonal / indefinite patterns reflects the presence / absence of the 1st person plural participant in its referential domain and its degree of grammaticalization, a hypothesis also confirmed by the synchronic distribution of impersonal/indefinite SE in Italo-Romance. Florentine (and Tuscan dialects) used a different strategy, the third person plural, rather than impersonal *si*+active$_{3SG}$ – nowadays functioning as the first person plural ending (e.g., Fl. *si va* RFL go.PRS.IND.3SG = *andiamo* go.PRS.IND.1PL 'we go') – if the speaker was excluded from the universe of discourse (Cennamo, 2014: 82 and references therein).

We can identify at least two possible sources of the impersonal / indefinite reflexive in Italo-Romance, which led to its subsequent reanalysis as an indefinite pronoun with a variable referential status and generic / existential reference: (i) the gradual widening of its referential scope in reflexive / middle function, attested in Latin by the second to third century AD, whereby SE comes to include the first and second person participants (*se* = *nos* 'we') (40a), (ii) its anaphoric pronominal use (*se* = *is* 'he, it') (40b) (Cennamo 1991a; 1991b, 1993a: 81; 2014: 93–94 and § 2.3).

(40) a. (= 14a) *inter   se   (= nos)   singuli dissimiles invenimur*
                    between RFL (we.ACC) each   different find.MPASS.PRS.1PL
                 'We find ourselves different from each other'   (Min. Fel. 18,15) (II AD)

  b. *ipsi… sibi   (= ei)   crediderunt*
        they   RFL.DAT   (he.DAT) believe.PST.IND.3PL
        'They trusted him'   (*Trad. Frising.* 553; Cennamo 1991b: 8)

Thus, the impersonal / indefinite use of the reflexive morpheme with one-argument verbs and the various types of impersonal patterns in which SE occurs in the diachronic data investigated appear to be related to a series of changes where the Person hierarchy is one of the features involved.

6.   Conclusions

The analysis of the paths and patterns of actualization of some aspects of the new systems of voice arising in the transition from Latin to Romance reveals the persistence of the semantic parameters of aspect and animacy in the encoding of voice as well as a divergence with respect to Latin. With the rise of a new semantic parameter, the person hierarchy, signalled by the widening of the referential domain of SE, the reflexive morpheme came to refer to the SAPs in late Latin, also occurring in (anaphoric) pronominal function. These changes paved the way for future Romance developments and their differentiation as regards the use of SE as an impersonal / indefinite strategy, an innovation that appears to reflect the presence / lack of the SAPs in its referential domain. This trend will be investigated in future work for its wider Romance implications.

    The persistence of the notions of animacy and aspect is apparent in the rise and distribution of passive constructions, analysed for the changes affecting the lexical verbs BECOME, MAKE and for the use of the reflexive as a voice marker in the passage to from Latin to Romance. Indeed, the data investigated point to the existence of a multifaceted semantic and syntactic space in the voice domains investigated, characterized by different types and degrees of desemanticization and decategorialization, molded by the animacy of the verbal core arguments and by the internal temporal constituency of eventualities, i.e., lexical aspect, interacting with the completion / perfectivity of the situation described by the verb in a clause.

    More specifically, evidence has been provided for the following claims:

1.   The grammaticalization of change of state verbs such as *fieri* 'become' and activity verbs such as *facere* 'do / make', involves a stage where they equal the copula *esse* 'be';

2. The passive function of the constructions *fieri / facere* + PP appears to be a subsequent stage, involving copula auxiliarization, and triggered by a change of the aspectual classes of the lexical verbs occurring in participial forms in these periphrases. When they no longer belong to achievements / accomplishments, but are activity verbs, the pattern has a passive interpretation, signalling an externally-caused eventuality. When the main verb (i.e., the past participle) is an achievement / accomplishment, the structure is ambiguous, varying between an anticausative (i.e., spontaneous, internally-caused) and passive (i.e., induced, externally caused) reading of the past participle.

3. The reanalysis of the reflexive morpheme as a voice modulator, from anticausative to passive, reflects the parameters of animacy and aspect, whilst its impersonal / indefinite function is related to the non-reflexive, pronominal value of SE, the widening of its referential domain, and its degree of grammaticalization, as witnessed by the early Italo-Romance and Old Logudorese Sardinian data investigated.

Although the issue needs further investigation, with a detailed analysis of variational and diachronic data from other Romance varieties, a preliminary study of the diachronic distribution of the morpheme SE as a voice marker in early Italo-Romance and Old Logudorese Sardinian confirms the role played by the interplay of animacy, aspect, and the person hierarchy in this transitivity domain (Cennamo 1993a), revealing interesting recurrent patterns, throwing a light on the possible factors involved in the rise of impersonal / indefinite SE. Whereas reflexive / middle / anticausative and passive SE are attested in most Romance languages, impersonal / indefinite SE is not equally attested and not in the same range of constructions, depending on its referential scope and its degree of grammaticalization. The languages / varieties where SE does not have an existential / inclusive interpretation either lack impersonal / indefinite SE or show a more limited use of this morpheme in this function, with other strategies being preferred, as discussed above. The data also confirm the hypothesis that passive and impersonal SE result from different diachonic paths that at some point merge: whereas passive SE continues a Late Latin pattern, expanding it with the possibility of overtly expressing the agent (albeit with a different chronology), impersonal / indefinite SE is a Romance phenomenon, not equally attested in the Romance languages and appearing at a later stage.

Acknowledgements

Abbreviations

ABL	ABLATIVE (CASE)	NEUT	neuter
ACC	accusative	NOM	nominative (case)
ACT	active	PASS	passive
AGR	agreement	PL/pl.	plural
AN	animate	PLUPF	pluperfect
DAT	dative	PP	past participle
F	feminine	PRF	perfect
FUT	future	PRS	present
GEN	genitive	PST	past
GER	gerundive	PTCP	participle
IMPF	imperfect	RFL	reflexive
IMPER	imperative	SG/sg.	singular
IMPRS	impersonal	SBJV	subjunctive
IND	indicative	trans.	transitive
INF	infinitive	1	first person
intr.	intransitive	2	second person
Lat	Latin	3	third person
M	masculine		
MPASS	medio-passive marker – R (in middle, passive or impersonal function)		

References

Adams, J. (2013). *Social variation and the Latin language*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511843433

Adams, J. & Vincent, N. (Eds.). (2016). *Early and late Latin: Continuity or change?*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781316450826

Andersen, H. (Ed.). (2001). *Actualization: Linguistic change in progress*. Amsterdam Studies in the Theory amd History of Linguistic Science Series 4. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.219

Andersen, H. (2001). Actualization and the (uni) directionality of change. In Henning Andersen (Ed.) *Actualization: Linguistic change in progress*. (pp. 225–248). Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.219.11and

Andersen, H. (Ed.). (2011). *Actualization: Linguistic change in progress*. Amsterdam: John Benjamins.

Barðdal, J. & Chelliah, S. (Eds). (2009). *The role of semantic, pragmatic and discourse factors in the development of case*. Amsterdam / Philadelphia: John Benjamins. https://doi.org/10.1075/slcs.108

Bassols de Climent, M. (1948). *Sintaxis Històrica de la Lengua Latina*. vol 2. Barcelona: Escuela de Filología.

Benincà, P., Ledgeway, A. & Vincent, N. (Eds.). (2014). *Diachrony and dialects. Grammatical change in the dialects of Italy*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198701781.001.0001

Bertinetto, P. M. & Squartini, M. (1995). An attempt at defining the class of gradual completion verbs. In P. M. Bertinetto (Ed.). *Temporal reference, aspect and actionality. 1. Semantic and syntactic perspectives*. (pp. 11–26). Torino: Rosenberg and Sellier.

Bertinetto, P. M. (Ed.). (2005). *Temporal reference, aspect and actionality. 1. Semantic and syntactic perspectives*. Torino: Rosenberg and Sellier.

Blasco Ferrer, E. (1995). Un passivo smarrito: *fecit positum* 'venne assegnato. In C. Mastrelli (Ed.), *Studi Linguistici per i 50 anni del Circolo Linguistico Fiorentino e i secondi Mille Dibattiti 1970–1995*. (pp. 47–53). Firenze: Olschki.

Bonnet, M. (1890). *Le latin de Grégoire de Tours*. Paris: Hachette.

Burton, P. (2016). Analytic passives and deponents in classical and later Latin. In J. Adams & N. Vincent (Eds.). *Early and Late Latin: Continuity or change?*. (pp. 163–179). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781316450826.008

Cennamo, M. (1991a). La nascita di un nuovo sistema di voce in italiano antico. In Dieter Kremer (ed.), *Actes du VIIIe Congrès International de Linguistique et de Philologie Romanes*, (pp. 242–262). Tübingen: Niemeyer.

Cennamo, M. (1991b). *Se, sibi, suus* nelle *Inscriptiones Latinae Christianae Veteres* ed i successivi sviluppi romanzi. *Medioevo Romanzo* 16.1 / 2. 3–20.

Cennamo, M. (1993a). *The Reanalysis of reflexives: A diachronic perspective*. Naples: Liguori.

Cennamo, M. (1993b). L'estensione del dominio referenziale del riflessivo in testi italiani antichi. *Archivio Glottologico Italiano* 78.1. 53–62.

Cennamo, M. (1998). The loss of the voice dimension between Late Latin and early Romance. In M. Schmid, J. Austin & D. Stein (Eds.), *Historical linguistics 1997*. (pp. 81–100). Amsterdam/ Philadelphia: John Benjamins. https://doi.org/10.1075/cilt.164.06cen

Cennamo, M. (1999). Late Latin pleonastic reflexives and the Unaccusativity Hypothesis. *Transactions of the Philological Society* 97.1. 103–150.

Cennamo, M. (2000). Costruzioni passive e impersonali in Veneziano e in Napoletano antico. In A. Englebert, M. Perrard, L. Rosier, & D. Van Raemdonck (Eds.), *Actes du XXIIe Congrès International de Linguistique et Philologie Romanes* (Bruxelles, 23–29 July 1998), vol. 2: *Les Nouvelles Ambitions de la Linguistique Diacronique*, (pp. 91–103). Tübingen: Niemeyer. https://doi.org/10.1515/9783110933116.91

Cennamo, M. (2001a). L'extended accusative e le nozioni di voce e relazione grammaticale nel latino tardo e medievale. In V. Viparelli (Ed.), *Ricerche linguistiche tra antico e moderno*, (pp. 3–27). Naples: Liguori.

Cennamo, M. (2001b). On the reorganization of voice distinctions and grammatical relations between Late Latin and early Romance. In Claude Moussy (Ed.), *De Lingua Latina Novae Quaestiones, Actes du Xè Colloque International de Linguistique Latine*. (pp. 51–65). Paris: Peeters.

Cennamo, M. (2001c). Classi verbali e cambiamento sintattico: la reinterpretazione passiva del costrutto riflessivo. In Z. Fábián & G. Salvi (Eds.), *Atti del XXXII Congresso Internazionale della Società di Linguistica Italiana*. (pp. 225–242). Rome: Bulzoni.

Cennamo, M. (2003). Perifrasi passive in testi non toscani delle origini. In N. Maraschio & T. Poggi Salani (Eds.), *Italia linguistica anno mille, Italia linguistica anno Duemila, Atti del XXXIV Congresso della Società di Linguistica Italiana*, 105–127. Rome: Bulzoni.

Cennamo, M. (2005). Passive auxiliaries in Late Latin. In S. Kiss & G. Salvi (Eds.). *Études de Linguistique offertes à Jôzsef Herman à l'occasion de son 80ième anniversaire*. (pp. 179–196). Tübingen: Niemeyer. https://doi.org/10.1515/9783110944532.177

Cennamo, M. (2006). The rise and grammaticalization paths of Latin *fieri* and *facere* as passive Auxiliaries. In W. Abraham & L. Leisiö (Eds.). *Passivization and typology*. (p. 311–336). Amsterdam / Philadelphia: John Benjamins. https://doi.org/10.1075/tsl.68.18cen

Cennamo, M. (2007). Auxiliaries and serials between Late Latin and early Romance. In D. Bentley & A. Ledgeway (Eds.). *Sui Dialetti Italoromanzi. Saggi in onore di Nigel B. Vincent. The Italianist, special supplement* 1. 63–87.

Cennamo, M. (2008). The rise and development of analytic perfects in Italo-Romance. In T. Eythórsson (Ed.), *Grammatical Change and Linguistic Theory*. The Rosendal Papers, (pp.115–142). Amsterdam/Philadelphia: John Benjamins.

Cennamo, M. (2009). Argument structure and alignment variation and changes in Late Latin. In J. Barðdal & S. Chelliah (Eds). *The role of semantic, pragmatic and discourse factors in the development of case*. (pp. 307–346). Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/slcs.108.17cen

Cennamo, M. (2011a). Impersonal constructions and accusative subjects in Late Latin. In A. Malchukov & A. Siewierska (Eds.), *Impersonal constructions. A cross-linguistic perspective*. (pp. 169–188). Amsterdam / Philadelphia: John Benjamins. https://doi.org/10.1075/slcs.124.06cen

Cennamo, M. (2011b). Case marking of core arguments and alignment in Late Latin. In M. Fruyt, M. Mazoyer & D. Pardee (Eds). *Grammatical case in the languages of the Middle East and Europe. Studies in Ancient Oriental Civilization* 64, (pp. 297–314). Chicago: The Oriental Institure of the University of Chicago.

Cennamo, M. (2014). Passive and impersonal reflexives in the Italian dialects. Synchronic and diachronic aspects. In P. Benincà, A. Ledgeway & N. Vincent (Eds.), *Diachrony and dialects. Grammatical change in the dialects of Italy*. (pp. 71–95). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198701781.003.0004

Cennamo, M. (2016). Voice. In M. Maiden & A. Ledgeway (Eds.), *The Oxford Guide to the Romance Languages*, (pp. 967–980). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199677108.003.0060

Cennamo, M. (2018). Grammaticalization and changes in argument linking: A case-study from old Logudorese Sardinian. In M. Chini & P. Cuzzolin (Eds.), *Tipologia, acquisizione, grammaticalizzazione – Typology, acquisition, grammaticalization studies*, (pp. 96–117). Milan: FrancoAngeli.

Cennamo, M. (2019a). Aspects of grammaticalization and reanalysis in the voice domain in the transition from Latin to early Italo-Romance. In L. Heltoft, I. Igartua, B. Joseph, K. Kragh, & L. Schøsler (Eds.). *Perspectives on Language Structure and Language Change*. (pp. 205–232). Amsterdam / Philadelphia: John Benjamins. https://doi.org/10.1075/cilt.345.09cen

Cennamo, M. (2019b). Intransitive alternations and the semantics of predicates in Latin. In P. Cotticelli Kurras & S. Ziegler (Eds.), *Tra Semantica e Sintassi: il Ruolo della Linguistica Storica*, (pp. 1–31). Rome: Il Calamo.

Cennamo, M. (2020). Mechanisms and paths of grammaticalization and reanalysis in Romance. In W. Bisang & A. Malchukov (Eds.), *Grammaticalization scenarios. areal patterns and cross-linguistic variation. A comparative handbook*. Berlin: De Gruyter.

Cennamo, M., Eythórsson, T. & J. Barðdal (Eds.). (2015). Anticausatives in Latin and Old- Norse Icelandic: (Morpho)syntactic aspects and semantic constraints. *Linguistics* 53.4. 1–53. https://doi.org/10.1515/ling-2015-0015

Comrie, B. (1989). *Language Universals and Linguistic Typology*. 2nd ed. London: Blackwell.

Danckaert, L. (2016). Variation and change in Latin *be*-periphrases: Empirical and methodological considerations. In J. Adams & N. Vincent (Eds.). *Early and Late Latin: Continuity or change?* (pp. 132–162). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781316450826.007

Danckaert, L. (2017). The origins of the Romance analytic passive: Evidence from word order. In E. Mathieu & R. Truswell (Eds.). *Micro-change and macro-change in diachronic syntax*. (pp. 216–235). Oxford: Oxford University Press.

De Bartholomaeis, V. (1902). Codex diplomaticus Cajetanus. *Archivio glottologico italiano* 16: 9–27.

De Melo, W. (2012). Kuryłowicz's first "law of analogy" and the development of passiv periphrases in Latin. In P. Probert & A. Willi (Eds). *Laws and Rules in Indo-European*. (pp. 83–101). Oxford: Oxford University Press.

Dik, S. (1987). Copula auxiliarization: how and why?. In Martin Harris & Paolo Ramat (Eds.), *Historical development of auxiliaries*, (pp. 54–84). Berlin: Mouton De Gruyter. https://doi.org/10.1515/9783110856910.53

Feltenius, L. (1977). *Intransitivization in Latin*. Uppsala: Almkvist & Wiksell.

Flobert, P. (1975). *Les verbes déponents latins des origines à Charlemagne*. Paris: Les Belles Lettres.

Giacalone Ramat, A. & Sansò, A. (2011). From passive to impersonal. A case study from Italian and its implications. In: A. Malchukov & A. Siewierska (Eds.). *Impersonal constructions. A cross-linguistic perspective*, (pp. 189–228). Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/slcs.124.07ram

Green, J. (1982). The status of the Romance auxiliaries of voice. In N. Vincent & M. Harris (Eds.). *Studies in the Romance verb*. (pp. 97–138). London: Croom Helm.

Haag, O. (1898). *Die Latinität Fredegars*. Erlangen: Junge.

Hageman, K. (2006). The Latin synthetic passive in the Códice Emilianense. *Romansk Forum* 20.2. 91–102.

Harris, A. & Campbell, L. (1995). *Historical syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511620553

Haspelmath, M. (1987). *Transitivity alternations of the anticausative type*. Köln: Institut für Sprachwissenschaft. Universität zu Köln. (Arbeitspapiere N.F. 5).

Haspelmath, M. (2011). On S, A, P, T, and R as comparative concepts for alignment typology. *Linguistic Typology* 15.3. 535–689. https://doi.org/10.1515/LITY.2011.035

Hay, J., Kennedy, C. & Levin, B. (1999). Scale structure underlies telicity in 'degree achievements'. In T. Mathews and D. Strolovitch (eds.), *Semantics and Linguistic Theory* 9, 127–144. Ithaca, NY: CLC Publications.

Heine, B. (1993). *Auxiliaries: Cognitive forces and grammaticalization*. Oxford: Oxford *University Press*.

Heine, B. & Reh, M. (1982). *Grammaticalization and reanalysis in African languages*. Hamburg: Helmut Buske Verlag.

Herman, J. (2002). La disparition du passif synthétique latin: nouvel essai sur l'écrit et le parlé en latin Mérovingien. *Estudis romànics* 24. 31–46.

Herzog, E. (1910). Das –to Partizip im Altromanischen. *Zeitschrift für romanische Philologie* 26. 76–186.

Hofmann, J., Leumann, M. & Szantyr, A. (1963). *Lateinische Grammatik*, vol. 1. München: Beck.

Hofmann, J., Leumann, M. & Szantyr, A. (1965). *Lateinische Grammatik*, vol. 2. München: Beck.

Koontz-Garboden, A. (2009). Anticausativization. *Natural Language and Linguistic Theory* 27. 77–138. https://doi.org/10.1007/s11049-008-9058-9

Lakoff, G. (1977). Linguistic Gestalts. *CLS* 13. 238–87.

Lehmann, Christian. (1988). Predicate classes and participation. *Studies in General Comparative Linguistics*, 33–77. Köln: Akup.

Ledgeway, A. (2009). *Grammatica diacronica del napoletano. Beihefte zur Zeitschrift für romanische Philologie* 350. Tübingen: Niemeyer. https://doi.org/10.1515/9783484971288

Maiden, M. & Parry, M. (Eds.). (1997). *The dialects of Italy*. London / NY: Routledge.

Malchukov, A. & Siewierska, A. (Eds.). (2011). *Impersonal constructions. A cross-linguistic perspective*. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/slcs.124

Manzini, R. & Savoia, L. (2005). *I dialetti italiani e romanci*, vol. 2. Alessandria: Edizioni Dell'Orso.

Meyer-Lübke, W. (1902). *Zur Kenntniss der Altlogudoresischen*. Wien: Sitzungsberichte der Kais. Akademie der Wissenschaften. Band CXLV.

Michaelis, Susanne. (1998). Antikausativ als Brücke zum Passiv: *fieri*, *venire* und *se* im Vulgärlateinische und Altitalienischen. In W. Dahmen, G. Holtus & J. Kramer (Eds.). *Neuere Beschreibungsmethoden der Syntax romanischen Sprachen*. (pp. 69–98). Tübingen: Narr.

Pinkster, H. (2015). The Oxford Latin syntax. Vol. 1. *The simple clause*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199283613.001.0001

Pirson, J. (1901). *La langue des inscriptions latines de la Gaule. Bibliothèque de la Faculté de Philosophie et Lettres de l'Université de Liège, fasc.* XI.

Pirson, J. (1906). Mulomedichina Chironis. La syntaxe du verbe. *Festschrift zum Allgemeinen Deutschen Neuphilologentage*, Erlangen, 390–431.

Plank, F. (1985). *Relational typology*. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110848731

Pountain, C. (1982). **Essere / Stare* as a Romance phenomenon. In N. Vincent & M. Harris (Eds.). *Studies in the Romance verb*, (pp. 139–60). London: Croom Helm.

Probert, P. & Willi, A. (Eds). (2012). *Laws and rules in Indo-European*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199609925.001.0001

Reichenkron, G. (1933). *Passivum, Medium und Reflexivum in den romanischen Sprachen*, Jena/Leipzig: Gronau (Berliner Beiträge zur romanischen Philologie III.1).

Ronconi, A. (1968). *Il verbo latino. Problemi di sintassi storica*. Florence: Le Monnier.

Salvi, G. (1988). La frase semplice. In L. Renzi (Ed.), *Grande grammatica italiana di consultazione*, vol. 1, (pp. 29–113). Bologna: Il Mulino.

Salvi, G. (2010). La realizzazione sintattica della struttura argomentale. In G. Salvi & L. Renzi (Eds.). *Grammatica dell'Italiano Antico*, vol. I, (pp. 123–189). Bologna: Il Mulino.

Svennung, J. (1935). *Untersuchungen zu Palladius und zur Lateinischen Fach- und Volkssprache*. Uppsala: Almquist.

Van Valin, R. (2005). *Exploring the syntax-semantics interface*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511610578

Vincent, N. (1987). The interaction between periphrases and inflection: some Romance examples. In M. Harris & P. Ramat (Eds.). *Historical development of auxiliaries.* (pp. 237–256). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110856910.237

Wehr, Barbara. (1995). SE-Diathese im Italienischen. (Romanica Monacensia 37). Tübingen: Narr.

Winters, Margaret. (1984). Steps toward the Romance passive inferrable from the Itinerarium Egeriae. *Romance Philology* 37. 445–54.

Woodcock, Eric. (1959). *A new Latin syntax*. London: Methuen.

Old Italian sources

Bonvesin, *Disputatio = Disputatio musce cum formica*. In G. Contini (Ed.). (1937). *Cinque Volgari di Bonvesin da la Riva* (pp. 27–40). Modena: Società Tipografica Modenese.

Bonvesin, *De quinq. curial. = De quinquaginta curialitatibus ad mensam*. In G. Contini, (Ed.). (1937). *Cinque Volgari di Bonvesin da la Riva.* (pp. 41–44). Modena: Società Tipografica Modenese.

Capitolare = Princivalli, A. & Ortalli, G. (Eds.). (1993). *Il Capitolare degli Ufficiali sopra Rialto. Nei luoghi al centro del sistema economico veneziano (secoli XIII-XIV)*. Milan: La Storia.

Cedola di Marco Granello = Cedola di Marco Granello. In A. Stussi (Ed.). (1965). *Testi veneziani del Duecento e dei primi del Trecento*. (pp. 36–40). Pisa: Nistri-Lischi.

Cedola di Marco Michel = Cedola di Marco Michel. In A. Stussi (Ed.). (1965). *Testi veneziani del Duecento e dei primi del Trecento*. (pp. 111–22). Pisa: Nistri-Lischi.

Compagnia di S. M. del Carmine = *Libro degli ordinamenti della Compagnia di Santa Maria del Carmine*. In A. Schiaffini (Ed.). (1926). *Testi fiorentini del Dugento e dei primi del Trecento*. (pp. 55–72). Florence: Sansoni.

Cronica = Ceruti, A. (Ed.). (1878). Cronica degli Imperadori. *Archivio Glottologico Italiano* 3: 177–243.

Cronica fior. = Cronica Fiorentina. In A. Schiaffini (Ed.). (1926). *Testi fiorentini del Dugento e dei primi del Trecento*. (pp. 82–150). Florence: Sansoni.

CSNT = Merci, P. (Ed.). (1992). *Il Condaghe di San Nicola di Trullas*. Sassari: Delfino.

CSPS = Delogu, I. (Ed.). (1997). *Il Condaghe di San Pietro di Silki. Testo Logudorese Inedito dei Secoli XI-XIII*. Sassari: Dess.

CV = Solmi, A. (1905). Le carte volgari dell'Archivio Arcivescovile di Cagliari: Testi campidanesi dei secoli XI–XIII. *Archivio Storico* 35.238: 273–330.

Dante, *Inf.* = Petrocchi, G. (Ed.). (1966). *Dante Alighieri, La Commedia secondo l'antica vulgata*, vol. II *Inferno*. Milan: Mondadori.

Grisostomo = Forster, W. (Ed.). (1880-1883). Antica parafrasi lombarda del <<*Neminem laedi nisi a se ipso*>> di S. Giovanni Grisostomo', *Archivio Glottologico Italiano* 7: 1–120.

Destructione de Troya = De Blasi, N. (Ed.). (1986). *Libro de la Destructione de Troya. Volgarizzamento Napoletano Trecentesco da Guido delle Colonne*. Rome: Bonacci.

Distruzione di Troia = Libro della distruzione di Troia (Da un). In A. Schiaffini (Ed.). (1926). *Testi fiorentini del Dugento e dei primi del Trecento*. (pp. 151–184). Florence: Sansoni.

Milione = Bertolucci Pizzorusso, V. (Ed.). (1975). *Il Milione di Marco Polo (versione toscana del Trecento)*. Milan: Adelphi.

Novellino = Conte, A. (Ed.). (2001). *Il Novellino*. Roma: Salerno.

Panfilo = Tobler, A. (Ed.). (1886). Il Panfilo in antico veneziano. *Archivio Glottologico Italiano* 10: 179–243.

Passione = Salvioni, C. (Ed.). (1886). La Passione e altre antiche scritture Lombarde. *Archivio Glottologico Italiano* 9: 1–22.

Ricordi = Ricordi di compere e cambi di terre in Val di Streda e dintorni. In A. Castellani (Ed.). (1982). *La prosa italiana delle origini: I, Testi toscani di carattere pratico*. (pp. 215–254). Bologna: Pàtron.

RSPS = Piras, S. S. & Dessì, G. (Eds.). (2003). *Il Registro di S. Pietro di Sorres*. Cagliari: Cuec.

Sacchetti, *Trecentonovelle* = Pernicone, Vincenzo (ed.), 1946. Franco Sacchetti, *Il Trecentonovelle*, Firenze: Sansoni.

SRS = P. E. Guarnerio (Ed.). (1892/1894). Gli Statuti della Repubblica sassarese, testo logudorese del secolo XIV, nuovamente edito sul codice e annotato. *Archivio Glottologico Italiano* 13: 1–124.

Ergative from passive in Proto-Basque

Mikel Martínez-Areta

University of the Basque Country (UPV/EHU)

This paper is an abbreviated version, in English, of Martínez-Areta (2016), in which I argue that the ergative alignment of historical Basque came about as a result of an active reanalysis of passive predicates in Proto-Basque. This view has been previously defended by Trask (1977), who drew on the template of finite verbs. Since the first slot of intransitives (representing S) and of transitives (representing P) are identical, P must have been also a clause-level S at a previous stage prior to an active reanalysis of passive clauses. I attempt to solve some problems of Trask's proposal, and suggest that some finite verbs of the past/irrealis beginning with A actually come from active clauses.

Keywords: ergativity, passives, alignment, grammatical relations, reanalysis, verb morphology, mood-prominent, Basque, Proto-Basque, functional typology

1. Introduction

Basque is the only ergative language in Western Europe. While interpretations of its synchronic functioning have abounded since the 80s, there has hardly been any interest in the origin of the alignment itself. Trask (1977, 1997) proposed that this alignment arose as a result of an active reanalysis of passive clauses, which for some reason were generalized to express predicates with a semantic A and a semantic P. Unlike some of Trask's other contributions, this proposal has been ignored. This paper is an attempt to show that Trask was on the right track, suggesting some solutions for the difficulties his analysis encountered. In what follows, I synthesize a detailed version of my views already defended in Martínez-Areta (2016), written in Basque, into the extension allowed here.

The ergativity of Basque is a morphologically-based one, as shown in (1a, b, c). It has no syntactic ergativity, as shown by coordination tests like (2), where the S of the first clause does not maintain its grammatical role in the second clause by means of an antipassive, as would occur e.g. in Dyirbal.

(1) a. *Gaur gizon-a-Ø denda-n ikus-i du-t*
today man-ART-ABS shop-INES.SG see-PFV AUX.PRS-1SG.A
'Today I have seen the man in the shop'

b. *Gaur gizon-a-Ø Bilbo-tik etorr-i d-a*
today man-ART-ABS p.n.-ABL come-PFV AUX.PRS+3SGS-be
'Today the man has come from Bilbao'

c. *Gaur gizon-a-k liburu bat-Ø eros-i du-Ø*
today man-ART-ERG book one-ABS buy-PFV AUX.PRS-3SG.A
'Today the man has bought a book'

(2) *Gizon-ak etorr-i d-ir-a eta Ø ogi-a eros-i du-te*
man-ABS.PL$_i$ come-PFV PRS-3PL-be and A$_i$ bread-ART buy-PFV PRS-3PL.A$_i$
'The men have come and brought the bread'

In today's Basque, there are passives, but they are a calque of Romance passives, and not as productive as in the latter. These passives are unrelated to the (hypothesized) ones I will propose from Section 3 onwards. On the other hand, Basque has no grammaticalized anti-passives, in spite of being ergative.

It is important to note that the morphosyntactic alignment of today's Basque is not strictly ergative from a semantic point of view. A number of predicates which are semantically intransitive have a subject marked as A, perhaps the majority of them being "light verbs" of the structure *X egin* 'do X', such as *lo egin* 'sleep' or *korrika egin* 'run'. There are also lexically simple verbs of this unergative type, but in most of these the subject is marked as S in eastern dialects (compare west. *bazkaldu dot/det* '(I-ERG) have had lunch' vs east. *bazkaldu n(a)iz* '(I-ABS) have had lunch'). As Aldai (2009) has shown, although a few unergative verbs are common to all dialects, the unergative pattern of most of these is a western innovation; eastern dialects have instead preserved the original alignment based on valence. This fits with the rise of the ergative alignment proposed by Trask (1977) and in the following sections, based on valence features.

Before undertaking a description of the verbal morphology, I will provide a brief explanation about the chronology assumed in today's Basque diachronic research. The earliest substantial texts written in Basque date to the 16th century. Before this, we hardly have anything beyond sparse onomastic items within texts written in diplomatic Latin or Romance in the Low Middle Ages, and anthroponyms and theonyms in Roman inscriptions in antiquity. Since Mitxelena (1981/1987), it has commonly been assumed that Basque dialects started to develop in the 5th/6th centuries, and the Basque of this period is called Old Common Basque (henceforth OCB). Before OCB, Proto-Basque (henceforth PB) is usually defined as the phase prior to the contact with Latin-Romance influence (~ 1st c. BC), a definition conceived, however, on phonological grounds.

Methodologically, the lack of written evidence for such a long time obliges us to draw exclusively on internal reconstruction for research on PB morphology and syntax.

2.   Verbal morphology of Basque

Modern Standard Basque has finite and non-finite forms. The former are of two types:

a.   *Synthetic verbs*. These contain a lexical root, TAM morphemes which form present, past and irrealis series, and person markers for S, P, A, R (= Recipient), depending on the valence.
b.   *Auxiliary verbs*. Morphologically, these are like synthetic verbs, but their roots, which etymologically derive from 'be', 'have', etc., have lost their lexical meaning. Combined with one or more non-finite form(s), they make up analytic VPs. There are two sets of auxiliaries, one for indicative, one for marked moods (potential, subjunctive, imperative).

The main non-finite forms, in turn, are the following (the example is of 'bring'):[1]

a.   verbal stem (*ekar*)
b.   pfv. participle (*ekarr-i*)
c.   ipfv. participle (*ekar-tzen*)
d.   prospective participle (*ekarri-ko*)

By combinations of the previous elements, Basque forms a TAM system which is functionally quite similar to the verbal system of Standard Average European (= SAE) (see Trask, 1997; Mounole, 2014; Padilla-Moyano, 2013; Ariztimuño, 2013, for a concise description). VPs formed by the pfv. participle + auxiliaries with 'be' (for intransitives) and 'have' (for transitives) must have arisen through contact with Latin/Romance, or at least in some kind of convergence with Western SAE. VPs formed by the verbal stem plus the sets of auxiliaries used in marked moods are older than the rise of pfv. VPs, although it is unclear whether they are to be traced back to PB.

Be that as it may, the oldest morphological layer is, to all appearances, the one represented by one-part synthetic verbs. This strong conjugation can be formed only by a closed class of verbs which belong to the nuclear lexicon ('be', 'have',

---

1.   I leave aside modal verbs, verb locutions, as well as progressive periphrases and other verbal structures of more recent formation (see Trask, 1997, pp. 108–109).

'come', 'bring', 'walk', 'use'…). However, there were many of these verbs in 16th century Basque, and they must have been the productive type in PB. By combinations of TAM morphemes, historical Basque forms three groups of these.

a. *Present*. As in neighboring languages, the present covers stative presents, near or planned futures, gnomic statements, etc. There are grounds to believe that this form was originally a progressive, as the corresponding analytic VP of verbs without synthetic conjugation has been demonstrated to be a progressive periphrasis in origin by Mounole (2014).

b. *Past*. This is a past imperfect today, but some synthetic verbs were contextually aorist or narrative past (pfv. in the past) in 16th century Basque (see Mounole, 2014: 50–51). This latter meaning must have been closer to the original value.

c. *Irrealis*. Over the course of time, the irrealis has been restricted to an irrealis conditional, but in 16th century Basque it covered a wider range (like prescriptive and optative notions; see Mounole, 2014: 97–98).

A crucial point to consider is that the person markers of the past and the irrealis finite verbs share the same layout, as shown in Table 1 and 5 below. In finite forms in which either the A or the P – or both – is 1st or 2nd person, the only difference between the past forms and their corresponding irrealis forms is that the former have *-(a/e)n* suffixed, whereas the latter have no suffix. This fact points to a common origin. Bearing all that in mind, the structure of Basque finite verbs can be best represented by two synthetic verbs, intransitive *etorri* 'come' and transitive *ekarri* 'bring', in present and past/irrealis, as in Table 1.[2]

The 1st and 2nd person pronouns in Basque are: *ni* 'I', *hi* 'you', *gu* 'we', *zu* 'you (pl., originally)'. It is hence clear that the first slots representing 1st and 2nd person markers are of pronominal origin.

Note further that in intransitive sets, both present and past/irrealis, S is always located at the beginning. As for transitive sets, P is located at the beginning and A at the end in the present. By contrast, in the past/irrealis there is a split: if P is 1st or 2nd p., then this marker is located at the beginning (and A at the end if it

---

2. To simplify things, I include only singular persons in the table (the analysis of the plurals would follow similar lines). Moreover, space limitations oblige me to go directly to what is crucial for the extremely high chronology which the facts under consideration here belong to, and to skip grammatical categories which are deemed secondary from a diachronic point of view. In fact, in addition to the intransitive and transitive sets expounded in Table 1, by the insertion of an R-marker into finite forms of each of the previous set, Basque forms two more sets containing respectively S-R and P-R-A markers. Most of these are formed by means of derivations from the R-less variants, which are hence primary and diachronically older (see Ariztimuño, 2013, pp. 385–387, 389–394).

Table 1. Distribution of lexemes, TAM morphemes and singular core-argument markers in the synthetic verbs *e-torr-i* 'to come' (intransitive) and *e-karr-i* 'to bring' (transitive) in both present and past/irrealis conjugations

	Intransitive sets, present: S-VB.LEXEME			Transitive sets, present: P-VB.LEXEME-A			
	S	TAM	root	P	TAM	root	A
1SG	n	a	tor	n	a	kar	da
2SG	h	a		h	a		a/na
3SG		da			da		Ø

	Intransitive sets, past/irrealis: S-VB.LEXEME(-*a/e*)(-*n*)				Transitive sets, past/irrealis: P-VB.LEXEME-A(-*a/e*)(-*n*), if P = 1/ 2 person A-VB.LEXEME(-*a/e*)(-*n*), if P = 3 person & A = 1/ 2 person				
	S	TAM	root	Past	P/A	TAM	root	(A)	Past
1SG	n	en	tor	(a/e) + (n)	n	e(n)	kar	da	(a/e) + (n)
2SG	h	en			h	e(n)		a/na	
3SG		z/le				z/le		Ø	

is overt), whereas if P is 3rd person and A 1st or 2nd person, then A is located at the beginning. If both A and P are 3rd person, the first slot is occupied by a TAM morpheme, not by a person marker. For the sake of a better visualization of this layout, let us summarize the template of each possibility in Table 2.

Table 2. Templates of Basque finite verbal forms

1. In present forms
   1.1 Intransitive verbs: S – TAM.morph – vb.lex (- TAM.morph)
   1.2 Transitive verbs: P – TAM.morph – vb.lex (- TAM. morph) – A

2. In past/irrealis forms
   2.1 Intransitive verbs: S+TAM.morph – vb.lex (- TAM.morph) (- *a/en* if past)
   2.2 Transitive verbs: split!
        2.2.1 If P = 1 or 2 person: P+TAM.morph – vb.lex (- TAM. morph) – A (- *a/en* if past)
        2.2.2 If A = 1st or 2nd person and P = 3rd person: A+TAM.morph – vb.lex (- TAM. morph) – (- *a/en* if past)
        2.2.3 If both A and P = 3rd person: TAM.morph – vb.lex (- TAM.morph) (- *a/en* if past)

3.   History of the research on the Basque finite verb

The passive origin of the ergative defended in this paper, following Trask (1977), is anything but new. As early as the end of the 19th century, Hugo Schuchardt (1893) already saw that if, in the present, S-markers of intransitive verbs (*na-go* 'I am') were marked exactly as P-markers of transitive verbs (*na-kar-zu* 'you bring me'), then both had a common underlying pattern. His explanation was that the P of transitive verbs was really, i.e. underlyingly, the S of passive predicates.

> Das transitive Zeitwort des Baskischen ist passiv gedacht. *Na-kar-zu*, 'Sie tragen mich' heisst eigentlich 'ich werde getragen von Ihnen', und von dieser weitschweifigen Uebersetzungsweise lässt sich hier als in einer sprachgeschichtlichen Darstellung nicht abgehen.[3]   (Schuchardt, 1893, p. 2)

Thus, the identity of the first element of *na-go* and *na-kar-zu* was accounted for by interpreting that both were markers of the same grammatical role ("die des Subjekts"), and the problem boiled down to a synchronic interpretation of the data (Schuchardt, 1893, p. 3).

Until 1977, Bascologists were divided on this point (i.e. whether Basque transitive sentences were underlyingly passive or not). However, Trask (1977) objected that, setting out the problem that way, one was mixing up synchrony and diachrony. The following quote is, in my opinion, of dazzling clarity:

> Now, while I utterly reject the notion that the present day ergative should be regarded as passive in nature, I propose instead that it arose historically from a passive construction which was reanalyzed as active. This exactly accounts for the order of elements found in the Basque transitive verb.   (Trask, 1977, p. 208)

This explanation would account for the identity of the first slot in *n-a-tor* 'I come' and *n-a-kar* '(s)he brings me' in another manner: they once represented the same semantic role, that of S, and hence the latter was a passive structure. As for the formation of the finite verb from a previous clause, it is explained in the following elegant way:

> […] I propose that the SVO morphological order of Basque intransitive verbs should be taken as evidence for an older SVO syntactic order. That is, I propose that the originally independent pronouns came to be bound to the verb at a time when the structure of the Basque sentence was SVO, and that when this structure changed to SOV, the pronominal elements, being bound, could not move, but remained morphologically "frozen" in their earlier *syntactic* position. […].
>   (Trask, 1977, p. 208)

---

3.   'The transitive verb of Basque is passively conceived. *Na-kar-zu*, 'You carry me' actually means 'I will be carried by you', and this lengthy translation should not be omitted here as it might be in an historical linguistic representation.'

This is an application of Givón's principle according to which "today's morphology is yesterday's syntax". Now, if PB had passives, continued Trask, then it had actives too, and hence there should be a passivizing morpheme to make active verbs passive. What was it? His response is as follows:

> If the ergative is simply a reinterpreted passive, then we might expect to find the remains of an old passive marker in Basque transitive verb forms […]. / While there is no certain evidence for the reflex of a passive marker in Basque, it seems to me that there is an obvious candidate to be considered: the prefix *n-* which appears in most past tense forms. This morph *n* is always sandwiched between the personal prefix and the verbal root: e.g. *ninduen* 'he had me', *ginduzun* 'you had us'.
>
> (Trask, 1977, p. 210)

That is to say, the *-n-* between the person marker and the verbal root of past and irrealis finite forms (see Section 5) was proposed to be originally a passivizing morpheme, which lost its meaning when the verb was reanalyzed as an active. Trask was aware that this explanation ran into plenty of problems, such as the following: (1) why is there no *-n-* when the first slot is not a 1st or 2nd person marker (***ze-n-karr-en* but *ze-karr-en* '(s)he brought it')?, (2) why does this alleged passivizing *-n-* show up in S-markers (*ne-n-torr-en* 'I came')?, (3) why is there no trace of the opposition *-Ø-* vs *-n-* in 1st and 2nd p. plural markers (***ge-karr-en* vs ***ge-n-karr-en*, etc.)?

My proposal, defended in Martínez-Areta (2016), is that we don't need any passivizing morpheme, because verbal roots in PB were lexically unspecified for voice, as occurs in Chinese (Norman, 1988) or Lisu (Li & Thompson, 1976), in such a way that they expressed the simple verbal idea. Thus, the verbal root of a historical intransitive verb like *tor* (pfv. participle *e-torr-i*) would mean 'come', but that of a historical transitive verb like *kus* (pfv. participle *e-kus-i*) would not mean 'see', as in historical Basque, but 'see ~ be seen', depending on the context (both syntactic and pragmatic).

Restricting our analysis to the present, and applying Trask's passivizing view and Givón's principle, we could postulate two reconstructions like the ones in (3), where a historical transitive verb -a two-place predicate- could be complementized either by a direct object (3a), or by an agent complement (3b).

(3) a. **ni da-kus Peru*   > ~~Peru n-a-kus~~   (*kus* = active)
         I PROG-see/be.seen p.n.   ~~p.n. 1sGP-PRS-see~~
         'I am seeing Peru'   ~~'I see Peru'~~

     b. **ni da-kus Peru-ga*   > *Peru-k n-a-kus(a)*   (*kus* = passive)
         I PROG-see/be.seen p.n.-AG.C   p.n.-ERG 1sGP-PRS-see
         'I am being seen by Peru'   'Peru sees me'

Apparently, active predicates like (3a) disappeared from the present, and that is why present finite verbs always have P in the first slot, and never A, whenever this is occupied by a person marker.

What I am suggesting, then, is that transitive finite verbs with P in the first slot come from passive predicates, whereas transitive finite verbs with A in the first slot come from active predicates.

4.   The present

Let us begin this section by recalling the template of present finite verbs:

Intransitives: S.marker – TAM.morpheme – vb.lexeme
Transitives: P.marker – TAM.morpheme – vb.lexeme – A.marker

Now, while Trask's (1977) suggestion that ergative comes from passive has never caught on among Bascologists, the idea that clauses merged into finite verbs has, and several formalisms drawing on this idea have been developed since then (examples with *egon* 'be' and *etorri* 'come'):

(4)   a.   Trask (1977)
          1SG *na-da-go> nago*
          2SG *ha-da-go> hago*
          3SG *da-go> dago*
      b.   de Rijk (1992)
          1SG *n(i) dago → nago*
          2SG *h(i) dago → hago*
          3SG *dago → dago*
      c.   Gómez & Sainz (1995)
          1SG *ni-da-go> nago*
          2SG *hi-da-go> hago*
          3SG *Ø-da-go > dago*
      d.   Lakarra (2006)
          1SG *ni da-tor > *ni-da-tor   > *nyator   > nator*
          2SG *hi da-tor > *hi-da-tor   > *hyator   > hator*
          3SG *Ø da-tor   > Ø-da-tor   > dator   > dator*
          1PL *gu da-tor > *gu-da-tor(-(t)za) > *gwator-(t)za > gatoz*
          2PL *zu da-tor > *zu-da-tor(-(t)za) > *zwator-(t)za > zatoz*
          3PL *Ø da-tor   > *Ø-da-tor-(t)za   > *dator-(t)za   > datoz*

Lakarra (2008) has further proposed that *da-* was originally a progressive morpheme, with **dar* 'SIT, PUT' (as in *jarri* 'put' < **e-darr-i*) as its lexical source. This would form a serial verb construction with the following main verb. Although the view defended here does not rely on this, I consider Lakarra's suggestion most reasonable, and integrate this possibility into some of the formalisms below.

Now, if to the reconstruction in (5d) we add the assumption that the original predicates of the present were passive, then we would get a reconstruction along the lines of the one in Table 3 (examples with *ikusi* 'see').[4]

Table 3. Proposed development from passive predicates to transitive finite verbs in the present

1st phase (verbal roots expressing the simple verbal idea, unspecified for voice)	2nd phase (diathetical election to express the predicate)	3rd phase (promotion of the ag. compl. and active reanalysis of the clause)	4th phase (historical Basque)	Possibility which prevails
[A Peru, **P ni**, 'ikusi']				
Peru dar-kus ni	~~*Peru da-kus ni*~~			active
ni dar-kus Peru-ga	*ni da(r)-kus Peru-ga*	*Peru-ga n(i)(-)(d)a-kus*	*Peru-k nakus*	**passive**
[A Peru, **P hi**, 'ikusi']				
Peru dar-kus hi	~~*Peru da-kus hi*~~			active
hi dar-kus Peru-ga	*hi da(r)-kus Peru-ga*	*Peru-ga h(i)(-)(d)a-kus*	*Peru-k hakus*	**passive**
[A ni, **P Peru**, 'ikusi']				
ni dar-kus Peru	~~*ni da-kus Peru*~~			active
Peru dar-kus ni-ga	*Peru da(r)-kus ni-ga*	*ni-ga Peru da-kus*	*ni-k Peru da-kus-da*	**passive**

5. The past and the irrealis

Let us now recall the template of past / irrealis finite verbs:

Intransitives: S.marker+TAM.morpheme – vb. lexeme (- *a/en* if past)
Transitives: 3 possible templates

a.   If P = 1/ 2 person
     P.marker+TAM.morpheme – vb.lexeme – A.marker (- *a/en* if past)
b.   If P = 3 person and A = 1/ 2 person
     A.marker+TAM.morpheme – vb.lexeme (- *a/en* if past)
c.   If both P and A = 3 person
     TAM.morpheme – vb.lexeme (- *a/en* if past)

---

4.   The forms in brackets of the first phase represent three possible predicates with a semantic A and a semantic P. The two clauses below each possible predicate in brackets represent the two conceivable ways of expressing that predicate: actively (**Peru dar-kus ni* 'Peru sees me') or passively (**ni dar-kus Peru-ga* 'I am seen by Peru'), etc. In the second phase, a diathetical election is made which generalizes one of the variants for each predicate.

Now, if, as suggested above, P-initial finite verbs come from passive predicates but A-initial finite verbs come from active predicates, then for the past/irrealis we will need a reconstruction like the one in Table 4.[5]

Table 4. Proposed development from passive and active predicates to transitive finite verbs in the past/irrealis

1st phase (verbal roots expressing the simple verbal idea, unspecified as to voice)	2nd phase (diathetical election to express the predicate)	3rd phase (promotion of the ag. compl., active reanalysis of the clause)	4th phase (historical Basque)	Possibility which prevails
[A Peru, **P ni**, 'ikusi']				
Peru z/le-kus- ni	~~*Peru Ø/z/ le-kus- ni*~~			active
nen kus- Peru-ga	*nen kus- Peru-ga*	*Peru-ga nen(-)kus-*	*Peru-k nen-kus-*	**passive**
[A Peru, **P hi**, 'ikusi']				
Peru z/le-kus- hi	~~*Peru Ø/z/ le-kus- hi*~~			active
hen kus- Peru-ga	*hen kus- Peru-ga*	*Peru-ga hen(-)kus-*	*Peru-k hen-kus-*	**passive**
[A ni, **P Peru**, 'ikusi']				
nen kus- Peru	*nen kus- Peru*	*Peru ne(n)(-)kus-*	*Peru ne-kus-*	**active!**
Peru z/le-kus- ni-ga	~~*Peru Ø/z/ le-kus- ni-ga*~~			passive

As described in Section 3, in some past and irrealis finite forms, there is an *-n-* morpheme sandwiched between the person marker and the verbal root. This appears when the first slot is occupied by any 1st or 2nd person marker which is S or P: *nen-bil-en* 'I walked', *hen-bil-en* 'you walked', *gen-bil-tzan* 'we walked', *zen-bil-tzan* 'you (pl.) walked'; *nin-du-en* '(s)he had me', *hin-du-en* '(s)he had you', *gin-t-u-en* '(s)he had us', *zin-t-u-en* '(s)he had you (pl.)'.

However, when that 1st or 2nd person marker is A, then we have different results in the singular and in the plural. In the singular, there is no *-n-*: *ne-karr-en* 'I brought it', *he-karr-en* 'you brought it'.[6] By contrast, in the plural, some verbs do have *-n-*: *gen-du-en* 'we had it', *zen-du-en* 'you (pl.) had it'. Some others, in turn,

---

5. The hyphen after the verbal root indicates that this can be suffixed by *-a/en* (in the morphological sector antecedent of the past), or not (in the irrealis).

6. In the western strip of the Biscayan territory, however, 1st and 2nd person singular A-markers do have *-n-* in the verb **edun* 'have' (e.g. *nenduá* 'I had', attested in 1653), as already pointed out by Trask (1977, p. 211).

have no *-n-*, but actually *-n-e-*: *ge-n-e-karr-en* 'we brought it', *ze-n-e-karr-en* 'you (pl.) brought it'.

This *-n-* is probably the element on which the discussion about the finite verbal morphology gets most complex. Briefly, the main attempts to explain it are the following:

a. Oregi-Aranburu (1974) proposed *lehen* 'before' as its lexical source.
b. Trask (1977) suggested that it was originally a passivizing morpheme, which became devoid of any meaning after the active reanalysis.
c. Heath (1977) interpreted *-n-* as a morpheme triggering "inversion" (compare *ne-n-karr-en* '(s)he brought me' but *ne-karr-en* 'I brought it').
d. Aldai (2000) proposed *-n-* to be originally an imperfective marker, with the inessive case marker (as in *Bilbo-n* 'in Bilbao') as its source.
e. Lakarra (2008) identifies *-n-* as an originally perfective marker, with **den* 'FIN-ISH' (as in *ats-eden* 'rest') as its lexical form, making up a serial verb construction with the following main verb.

Heath's account is actually a synchronic one (and even as such it has serious problems; see Ortiz de Urbina, 1989, p. 62). As for the others, in my opinion a short-coming common to all is that they conceive the *-n-* as a part of the verbal complex (or of the periphrasis which gave rise to it), but this fails to explain why it doesn't appear when the first slot is not occupied by a 1st or 2nd person marker (e.g. **ze-n-karr-en, but *ze-karr-en* '(s)he brought it').

I suggest, instead, that, before the clause merged into a finite verb, *-n-* was originally part of the 1st and 2nd person pronouns, in such a way that there would be two sets of 1st and 2nd person pronouns, illustrated in (5):

(5)   Set A of pronouns Set B of pronouns
    1SG. *ni*   **nen*
    2SG. *hi*   **hen*
    1PL. *gu*   **gen*
    2PL. *zu*   **zen*

Set A would show up with finite forms antecedent to the present (**ni dar-kus*), set B in finite forms antecedent to the past/irrealis (**nen kus*), hence the representations in the left-hand column in Tables 3 and 4. In the singular, 1st and 2nd person A markers do not have this *-n-*: compare *ne-n-torr-en* 'I came', *ne-n-kus-en* '(s)he saw me', but *ne-kus-en* 'I saw it'. My proposal explains the last two forms as being originally the same form: **nen kus* 'I would see it ~ (s)he would see me' (the suffix-ation of *-(a/e)n* to form an aorist or a perfective would have been a later process, see Section 6). Once the active reanalysis had ensued and the clause had merged

into a finite verb, different clause-level contexts would have generated a split of predicates: *nen-kus Peru-ga* 'I would be seen by Peru' vs *nen-kus Peru* 'I would see Peru'. Now, since *n* may or may not drop out in a syllabic coda before a stop, this facultative dropping out in the A-marker *nen-* would have been employed to extend the S=P≠A alignment, which had already arisen at clause level as well as in the finite forms antecedent of the present (where S *n-* / P *n-* / A *-da*), to the finite forms antecedent to the past/irrealis.

The problem cannot be explained here in full detail, but there are two independent pieces of evidence for the existence of a set B of pronouns:

a. The independent pronouns of historical Basque correspond to set A above, to which case markers are suffixed (see Martínez-Areta, 2013). However, there is an exception: the old form for 'my' is not the expected *ni-re* (this exists, but is an analogical creation which first appears in the 18th century, in the west alone), but *ene* (the form common to all dialects in 16th-century Basque). I propose that the original form would be *neN-e* (set B pronoun + genitive marker), which would become *eNe* by dissimilation, and later *ene* 'my' in medieval Basque.

b. The second piece of evidence is a typological argument which will be put forward in Section 6.

6.   Split ergativity in Basque

At clause level, historical Basque NPs always show a P=S≠A morphological alignment, regardless of animacy, definiteness, etc., and hence noun morphology has no split whatsoever.[7] However, if we accept the reconstruction suggested for PB in the previous sections and compare Tables 3 and 4, we can see that PB did have a split at clause level. When clauses merged into finite verbs, this split became "frozen" in agreement markers of transitive finite verbs of historical Basque, giving rise to the alignment in Table 5 (a simplified version of Table 2, focused on the relative position of A and P).

---

7.   There are actually a few exceptions, but they have nothing to do with any parameter of Silverstein's hierarchy. The plural of (the three sets of) demonstratives, the personal pronoun *zu-ek* 'you (pl.)' (and in some early western attestations also *gu-ek* 'we'), and the quantifier *batzu(e)k* 'some' especially in peninsular dialects, have no distinction between the absolute and ergative forms (hence S=P=A). There is common agreement, however, that all these forms, as well as the plural in noun morphology in general, arose in a post-OCB period. Therefore, this syncretism is irrelevant for the chronology at stake in this paper.

Table 5. Split ergativity as "frozen" in agreement markers of transitive finite verbs of historical Basque

	P = 1st or 2nd person	A = 1st or 2nd person, and P = 3rd person
Present	P-…-A	P-…-A
Past / irrealis	P-…-A	A-… (P covert)

Now, the distribution within the past/irrealis is perfectly expected according to Silverstein's hierarchy. 1st and 2nd persons are by definition animate, definite, and known information. Therefore, they will be easily topicalized if they are P.

What is astonishing is that the more accusative pattern – with A- in some cases – appears precisely in the past/irrealis and not in the present. The anomaly of a verbal morphology with a past more A-oriented than the present was already detected by Ortiz de Urbina (1989, pp. 13–15). However, this is surprising if – and only if – we consider the past (perhaps coming from an aorist or a perfective) to be the original value of the morphological sector with that alignment. If we, instead, assume that the original value of that morphological sector is the irrealis, then the typological oddity vanishes, because then we will have, side by side, an irrealis and a present (most probably a progressive in origin). Of these, the former is more A-oriented according to Comrie's (1981) hierarchy, shown in the following figure:

A ⟵———————————————————————————————⟶ P
prospective   imperfective   perfect

That is to say, *ceteris paribus*, a given A will be more prominent the further to the left its correspondent predicate is (and vice versa).

Comrie put these three concepts (prospective, imperfective, perfect) in a coordinative relationship within a hierarchy. This can be somewhat confusing because, in the standard terminology, coined by Comrie himself in the 70s, the counterpart of imperfective is perfective. The perfect is a concept with both aspectual and temporal implications. Perhaps it would be more accurate to point out that all verbal categories have a logical implication for the prominence or salience of A and P, and that it would be preferable to split Comrie's hierarchy into three more specific ones: future > present > past in terms of tense, imperfective > perfective in terms of aspect, and irrealis > realis in terms of mood. In all three, A would be more prominent the further to the left its predicate is.[8]

---

8.   As Comrie (1981, p. 71) puts it, in a way summarizing these three hierarchies: "More generally, potentialities are more likely to be attributed to A than to P, so that we would expect prospective, future, potential, and irrealis to favor A-orientation, to correlate with active-accusative".

If we assume that the verbal sector corresponding to the historical past/irrealis was an irrealis in PB, and that an aorist or perfective was derived from it after clauses had merged into finite verbs (*nen-tor* 'I would come' → *nen-torr-en* 'I came'), then we would have had, immediately prior to the merger of clauses into finite verbs, an irrealis and a progressive.[9] This is much more compatible with Comrie's hierarchy (or hierarchies), provided that the irrealis is even more A-oriented than the progressive.

What I am suggesting is that, in the transition from the 1st phase to the 2nd in Tables 3 and 4 above, for the active variant to prevail as the expression of predicates with a semantic A and a semantic P, two conditions had to be met: (1) that the A be 1st or 2nd person and the P 3rd person, and (2) that the TAM value of the predicate be the more A-oriented of the existing ones in the language (i.e. the irrealis). If any of these conditions was not met, the prevailing variant was the passive one.

Moreover, I would go a step further and suggest that, even if we have to postulate an irrealis and a progressive for PB, the real opposition underlying the verbal system was irrealis / realis, the latter being realized aspectually as a progressive. This would allow us to link PB to the typology of languages set up by Bhat (1999), according to which there are: (1) tense-prominent languages, (2) aspect-prominent languages, but also (3) mood-prominent languages.

Some of the recurring features of the third type of languages are, interestingly, the following:

a. The opposition irrealis / realis is the one their verbal system is most sensitive to.
b. They usually have two sets of 1st and 2nd pronouns (see Bhat, 1999, p. 135), which is exactly what I proposed for PB in Section 5 on independent grounds.
c. They usually have a *pronominal* split ergativity (see Bhat, 1999, pp. 102, 146), always along the lines of the one in the Basque past/irrealis, i.e. a split based not on tense or aspect, but on which persons are the A and the P.

The split ergativity that we can observe frozen in the verbal morphology of historical Basque, and that I have postulated as occurring at clause level in PB, is actually more complex than the sheer pronominal split ergativity of languages like Tangut. In PB, the active morphosyntax would not occur in one half of the whole

---

9.   This does not mean that the aorist (a designation which may, admittedly, be an over-simplification) originated from the irrealis, but within the irrealis. Aoristic forms must have arisen from some irrealis forms which had the suffix *-(a/e)n* attached, in a process which I cannot address here. What I do defend is that the irrealis is diachronically prior to the aorist, an inference supported by the fact that it is the unsuffixed, hence formally unmarked member of the pair. Both Aldai (2000, pp. 61–62, 70–71) and Ariztimuño (2013, pp. 52–58) also appear to believe that, from a formal point of view and regardless of grammatical values, what corresponds to the historical past arose as a derivation of what corresponds to the historical irrealis.

spectrum of possible predicates, but rather in one half of a half (see Table 5, where its morphological stage is depicted). It would, then, represent a pronominal split ergativity within another split ergativity conditioned by the TAM value of the predicate, hence the double condition required for it to be triggered. I believe, however, that the marked nature of this split type and its association with mood-prominent languages is a strong indication that PB belonged to this type.

7. Several additional features that the ergative from passive hypothesis can explain

I believe that the passive-to-ergative hypothesis defended by Trask (1977) and enlarged upon here can also explain some other features of historical Basque which I have considered in Martínez-Areta (2016) but which cannot be discussed in detail here. I will just briefly mention the ones directly related to the steps assumed in Tables 3 and 4 and which are usually involved in attested passive-to-ergative processes.

If the ergative alignment comes from a reanalysis of passive predicates, then we would expect the ergative morpheme *-k* to be relatable to an agent complement. Now, Lakarra (2005, p. 442) has derived *-k* from **-ga*, labeling it as a "general locative" marker. This might be related, I add, to the locative suffix *-aga* of toponyms, as in *Ares-pacoch-aga* 'place of the solitary oak tree', to be segmented into *-a-ga*. Locative non-core NPs are a well-attested source of agent complements, and the origin proposed by Lakarra for the ergative marker *-k* is exactly what we would expect to have before the active reanalysis of passive predicates and the transformation of the agent complement marked as locative into a core A.

The passive-to-ergative hypothesis also offers a pragmatic explanation for the word-order change AVO > AOV, assumed by most scholars for PB. When passives were generalized to express any predicate with a semantic A and a semantic P, the semantic agent – i.e. the agent complement – always ended up in the position after the verb. Since the agent is typically animate, definite and known information, it was promoted to the beginning of the clause. Finally, the clause was reanalyzed as active, which brought about the rise of ergativity and of the AOV order. I would like to note that we can find an embrionic phase of this same account in Trask (1997, p. 247).

A final issue is what may have caused the merger of clauses into finite verbs. At least since Givón (1976), we have known that a common source of such processes, which of course also involve the rise of agreement, is the demarking of marked topicalized structures. We can unproblematically postulate such a process for PB. This issue cannot be taken up here, but the process is implied in the reconstructions proposed in Tables 3 and 4 (from the 2nd to the 4th phase).

8.   Conclusions

The passive theory is the best possible account to explain the rise of Basque ergativity, provided that:

a.   we interpret it in a diachronic way, and assign the existence of these passives to a previous phase, probably PB.
b.   we consider that the roots of historical transitive verbs were lexically unspecified for voice.

The split ergativity of Basque is best accounted for by assuming that, within the past/irrealis morphological sector, the original value was the irrealis. If we do this, it is much easier for its unusual distribution of A and P to fit in with the typological trends attested worldwide.

Acknowledgements

Abbreviations

ABL	ablative	PL	plural
ABS	absolutive	PROG	progressive
ART	article	PRS	present
AUX	auxiliary	S	single argument of canonical intransitive verb
ERG	ergative		
INES	inessive case	SAE	Standard Average European
OCB	Old Common Basque	SG	singular
PB	Proto-Basque	TAM	tense-aspect-mood
PFV	perfective		

References

Aldai, G. (2000). Split ergativity in Basque. The pre-Basque antipassive-imperfective hypothesis. *Folia Linguistica Historica*, 21, 31–98.

Aldai, G. (2009). Is Basque morphologically ergative? Western Basque vs. Eastern Basque. *Studies in Language*, 33, 783–831. https://doi.org/10.1075/sl.33.4.01ald

Ariztimuño, B. (2013). Finite verbal morphology. In: Martínez-Areta, M. (Ed.), *Basque and Proto-Basque. Language-internal and typological approaches to linguistic reconstruction* (pp. 359–427). Frankfurt am Main: Peter Lang.

Bhat, D. (1999). *The prominence of tense, aspect and mood*. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/slcs.49

Comrie, B. (1981). Aspect and voice: some reflections on perfect and passive. In: P. Tedeschi, & A. Zaenen (Eds.). *Syntax and semantics (Vol. 14). Tense and aspect* (pp. 65–78). New York/San Francisco / London: Academic Press.

Givón, T. (1976). Topic, pronoun and grammatical agreement. In: C. Li (Ed.), *Subject and topic* (pp. 149–188). New York/San Francisco/London: Academic Press.

Gómez, R. & Sainz, K. (1995). On the origin of the finite forms of the Basque verb. In: J. I. Hualde & J. A. Lakarra & R. L. Trask (Eds.), *Towards a history of the Basque language* (pp. 235–274). Amsterdam/Philadelphia: John Benjamins.

Heath, J. (1977). Remarks on Basque verbal morphology. In: W. A. Douglass & R. W. Etulain (Eds.), *Anglo-American Contributions to Basque Studies: Essays in Honor of Jon Bilbao* (pp. 193–201). Reno: University of Nevada.

Lakarra, J. (2005). Prolegómenos a la reconstrucción de segundo grado y al análisis del cambio tipológico en (proto)vasco. *Palaeohispanica*, 5, 407–470.

Lakarra, J. (2006). Notas sobre iniciales, cambio tipológico y prehistoria del verbo vasco. In: J. A. Lakarra & J. I. Hualde (Eds.). *Studies in Basque and historical linguistics in memory of R.L. Trask* (= Anuario del Seminario Julio Urquijo, 40) (pp. 561–621). San Sebastián / Bilbao.

Lakarra, J. (2008). Aitzineuskararen gramatikarantz (malkar eta osinetan zehar). In: X. Artiagoitia & J. A. Lakarra (Eds.). *Gramatika jaietan Patxi Goenagaren omenez* (= *Appendix of ASJU*, 51) (pp. 451–490). Bilbao: University of the Basque Country.

Li, C. & Thompson, S. (1976). Subject and topic: A new typology of language. In: Li, C. (Ed.), *Subject and topic* (pp. 457–489). New York / San Francisco / London: Academic Press.

Martínez-Areta, M. (2013). Demonstratives and Personal Pronouns. In: M. Martínez-Areta (Ed.), *Basque and Proto-Basque. Language-internal and typological approaches to linguistic reconstruction* (pp. 283–321). Frankfurt am Main: Peter Lang. https://doi.org/10.3726/978-3-653-02701-3/14

Martínez-Areta, M. (2016). Euskararen lerrokadura ergatiboaren sorreraz. *Anuario del Seminario Julio Urquijo*, 50.

Mitxelena, K. (1981/1987). Lengua común y dialectos vascos. In: *Palabras y textos* (pp. 35–55). Vitoria/Gasteiz: University of the Basque Country.

Mounole, C. (2014). *Le verbe basque ancien: étude philologique et diacronique* (= *Anuario del Seminario Julio Urquijo*, 48). University of the Basque Country.

Norman, J. (1988). *Chinese*. Cambridge: Cambridge University Press.

Oregi-Aranburu, J. (1974). Euskal-aditzaz zenbait gogoeta. *Fontes Linguae Vasconum*, 17, 265–283.

Ortiz de Urbina, J. (1989). *Parameters in the grammar of Basque*. Dordrecht: Foris.

Padilla-Moyano, M. (2013). Non-finite verbal morphology. In: M. Martínez-Areta, (Ed.), *Basque and Proto-Basque. Language-internal and typological approaches to linguistic reconstruction* (pp. 323–357). Frankfurt am Main: Peter Lang.

Rijk, R. P. G. de. (1992). 'Nunc' Vasconice. In: J. I. Hualde & J. A. Lakarra & R. L. Trask (Eds.), *Towards a history of the Basque language* (pp. 295–311). Amsterdam/Philadelphia: John Benjamins.

Schuchardt, H. (1893). Baskische Studien I: Über die Entstehung der Bezugsformen des Baskischen Zeitworts. In: *Denkschriften der Kaiserlichen Akademie der Wissenschaften in Wien. Philosofisch-Historische Classe. Band XLII* (pp. 1–82). Vienna.

Trask, R. (1977). Historical syntax and Basque verbal morphology: Two hypotheses. In: W. A. Douglass & R. W. Etulain (Eds.), *Anglo-American contributions to Basque studies: Essays in Honor of Jon Bilbao* (pp. 203–217). Reno: University of Nevada.

Trask, R. (1997). *The history of Basque*. London: Routledge.

Patterns, paradigms, & restructuring

Synchrony, diachrony, and indexicality

Henning Andersen
UCLA

This paper is about the current disconnect between synchronic and diachronic linguistics and the need for historical linguists to take the lead in addressing the big issues in morphology: how agglutination develops into fusion and symbolism, how these techniques are maintained through time, how they devolve into agglutinative or isolating patterns. The formation and reanalysis of allomorphy are central to these developments. The paper advocates a theory of the morphological sign by which allomorphs form subsidiary paradigms and the individual allomorph is both symbol and index: it stands for its exponendum and points to features of its environment. This articulate conception facilitates the detailed analysis and interpretation both of synchronic patterns and of the gradual typological transformations of inflectional systems.

Keywords: morphology, allomorphy, WP, paradigm, symbol, index, reanalysis, interfix, zero morph, cumulative exponence

1.   Synchrony, diachrony and history

It is sometimes claimed that there is no difference between synchrony and diachrony: in reality, all we can observe is gradual transitions and changes. But surely it should be possible to distinguish between our object of investigation, languages, and the different approaches we use in studying them.[1]

Languages are historical objects; as such they present the observer with patterns in flux or in transition and other patterns that appear stable, at least temporarily. In this regard languages are like other cultural systems.

To study these objects we adopt distinct perspectives, the synchronic and the diachronic. In the diachronic perspective we employ several methods to study observed and reconstructed elements and patterns, asking how they came into being,

---

1.   I am grateful to several anonymous readers for useful suggestions, not least for the suggestion that resulted in the present title of the paper.

and how they developed to the stage – or between the stages – at which they are observed. By contrast, in the synchronic perspective we analyse observed elements and patterns, asking how they function, that is, how they contribute to the transmission of information in speech acts.

The diachronic perspective implies the synchronic in two respects: First, diachrony implies a comparison between two or more synchronic 'states'. Secondly, it is only after the synchronic function of an element or pattern has been identified, that is, when we know what it is, that it makes sense to ask about its origin or development.

The historical dimension of language is obvious in the diachronic perspective. Historical linguists know that all change occurs through variation and have long since integrated the insights of sociolinguistics in their work. But the historical dimension is no less important in the synchronic perspective. Any synchronic 'state' may present structural alternations as well as stylistic and social variation. These are meaningful patterns that reflect part of the speakers' knowledge.

Thus a synchronic language 'state' calls not for the static description prescribed by Saussure (1973, pp. 164–199), but for a dynamic description that captures – as much as possible – the observed flux and gradual transitions which reflect ongoing historical development; cf. Jakobson's (1990, pp. 164, 632) "dynamic synchrony".

Yet, a hundred years after the *Cours*, leading theoreticians of synchronic morphology study corpora of inflectional wordforms without regard to synchronic variation, let alone such basic distinctions as innovation vs. archaism, or productive vs. unproductive patterns.

2. Word and Paradigm morphology

During the second half of the 1900s some students of synchronic morphology became convinced that the established practice of analysing morphological wordforms into recurrent elements of expression associated with recurrent elements of content (Bloomfield, 1935, p. 155) does not work. It seemed that there were too many deviations from the ideal, biunique relation between content and expression. The 1960s and 1970s saw the development of the new theoretical approach of Word and Paradigm (WP). Matthews (1972, 1974, 1991) in several monographs critiqued morphological analysis, mocking the structuralists' preoccupation with units, and advocating instead a direct relationship between grammatical content and whole wordforms. Matthew's views are echoed by many other theoreticians of morphology, from Anderson (1992) to Blevins (2016); see the survey in Stewart (2016).

This theoretic development seems to stem from a confusion of language and method (cf. Section 1): The principles of morphological analysis are not a claim

about languages that can be invalidated by observed data; it is a heuristic procedure that lays bare the relations between the forms and meanings of morphological elements, biunique content-expression pairings, portmanteaux, multiple exponence, zero signs, empty morphs; see the exposition in Bloomfield (1935, pp. 158–169, 207–246).

From the point of view of historical linguistics the aims of WP descriptions are totally out of touch with actual languages. Anyone who has some experience with historical morphology knows that the innovations that are made in the transmission of language provide plenty of evidence that speakers are preoccupied with units of form and meaning: Morphological innovations demonstrate the fine-grained analysis to which language patterns, including inflectional paradigms, are subjected as they pass from generation to generation.

Besides, descriptive frameworks for corpora of whole wordforms cover up the diversity of techniques in individual morphological systems, the details of the most remarkable long-term developments in language histories: the gradual typological transformations from agglutination to fusion to symbolism (Dahl, 2004) and the less spectacular, but no less instructive back-and-forth changes between fusion and agglutination, symbolism and isolation (Werner, 1987).

3.   Morphology for communication

Descriptive theorizing in WP morphology has accepted Saussure's sign theory, assuming that the relations between Saussurean *signifiés* and *signifiants* exhaust what has to be accounted for in morphology (Anderson, 1992; Stump, 2001; Blevins, 2016).

Interestingly, during recent decades there has been a growing understanding among other linguists that this 'bipolar' sign concept is inadequate. Mel'čuk (2006, p. 384) has long advanced a more articulate concept of the linguistic sign as a triad comprising content (exponendum), expression (exponent) and syntactics. Some American linguists have similarly recognized that there must be more to morphology than the tradition imagined. Carstairs-McCarthy (2000) has suggested that allomorphy is meaningful. Spencer (2016), like Mel'čuk (2006), views the individual morphological sign as a triad composed of content, expression, and syntax, adding as a fourth component a 'serial number' to ensure the identity of the individual sign through the vagaries of semantic, phonological, or syntactic variation.

This more articulate understanding acknowledges that signs occur in contexts and function in relation to other signs. In semiotic terms, it implies a recognition that in addition to the symbolic relations between exponents and exponenda there are indexical sign relations between morphological signs and their contexts.

In Charles S. Peirce's well-known three-way distinction of icons, indexes and symbols, the index is the sign that draws attention to its object, indicates or points to it. It does this thanks to its 'existential relation' to its object.[2]

In every-day terms, the Index is the smoke mentioned in 'Where there is smoke, there is fire'. In linguistic terms, if a sign rule introduces an expression for a content in a certain environment, the expression is established as a Symbol of the content, but at the same time as an Index of a feature or features of the environment. For instance, in a certain Russian noun declension, the acc.sg desinence is /-a/ if the stem refers to an animate entity, and otherwise -Ø. We can say that the /-a/ *stands for* acc.sg, but *points to* the animacy of the stem referent, whereas the -Ø *stands for* acc.sg and *points to* the inanimacy of the stem referent (Example (9.a) below). As we shall see below, there are Index values of several kinds.

The importance of Indexes in language was first clarified in Peircean terms by Jakobson (1971). Their relevance in historical morphology was illustrated (without the *Index* term) by Korhonen (1969, 1982), by Dressler's (1977, 1985), who spoke of morphonological 'cosignaling', but explicitly (as Peircean *indexes*) by Shapiro (1969, 1991), Anttila (1975), and Andersen (1980; also 2006, 2009). This work has remained unacknowledged in the post-Saussurean mainstream of morphological theorizing but has some definite value.

For example, if we recognize the semiotic aspects of morphology we can abandon the traditional, backwards description of morphological alternations. For over a century scholarly descriptions of alternations have been stuck in nineteenth-century cause-and-effect terms, recently revamped in twentieth-century, machine-age lingo, in which a given feature *F* is said to 'trigger' a certain alternation. This established jargon assigns a role ('triggering') to the feature *F*, but none to the effect, the output of the rule.

The semiotic perspective turns this picture around and allows us to see what the morphological alternation is there for: the output of such a rule is an Index that *points to* its conditioning feature and thereby to the morphological sign to which that feature belongs. In short, the expression of a morphological sign is both Symbol and Index. Its content comprises both an exponendum and an index value. Both are part of the message communicated to the addressee(s).

---

2. Since all language signs are conventional, they are Symbols, in Peirce's terminology. Indexes in language, then, are indexical symbols. For simplicity's sake they will be called Indexes in the following. To separate Peirce's terms from their every-day language homonyms I capitalize the terms for sign types and italicize the expressions that characterize their function. For definitions of the basic sign types referred to here, see the Appendix.

4.   Morphology by itself?

If the content of morphological signs comprises both exponenda and indexical values, morphology is – precisely by virtue of its system of Indexes – of a piece with the rest of grammar. Morphological indexes need to be seen in that wider grammatical context.

A.   Reference and presuppositions. All reference is indexical. Utterances *point to* situations by *depicting* parts of them. To the extent that they describe, they are Icons – Images, Diagrams, or Metaphors in Peirce's terms (See Appendix; Nöth, 1990, p. 123) – composed of Symbols. But they refer because they *direct attention*, that is, they act as Indexes.

   Besides, the content of utterances is selected by each speaker so as to *point to* the addressee's (supposed) presuppositions and thereby fulfill the addressee's Gricean expectations.

B.   Coherence and textuality. Text deixis is exophoric as well as endophoric. Exophoric elements – discourse markers, and other shifters (Jakobson, 1971) – anchor utterances in the speech-act context by *pointing to* participants, time, place, social dimensions, and shared elements of reference.

   At the same time, endophoric (anaphoric and cataphoric) deixis, lexical and pronominal, tense agreement (*consecutio temporum*), modal expressions, and clause-combining elements (e.g., adversative, copulative, switch-reference) create coherence within and between clauses. Recurrent Indexes throughout a monolog or dialog lend it its character of textuality.

C.   Information structure. The information structure of an utterance may be *represented* by distinct grammaticalized valency patterns, in which grammatical relations are *diagrammed* by morphological or syntactic means (e.g., active vs. passive constructions); or information structure may be *diagrammed* by constituent order and/or prosodic features.

D.   Concord and agreement. On the sentence level, agreement morphology is selected by speakers as Indexes that *point to* co-constituents of the phrase, the clause, the wider co-text, or the context.

   In endophoric agreement, determiners and specifiers may *point to* their semantic heads, and predicates may *point to* their arguments by means of a partial identity in features.

   Exophoric agreement includes the deviations from endophoric agreement traditionally termed 'semantic agreement', which it would be better to recognize as referential agreement. Politeness agreement, which tailors the utterance to fit the relation between speaker and addressee, and stylistic features that agree with circumstances of the speech act, are also exophoric.

E.   Case marking. In phrases, head and dependent are co-Indexes forming a *Diagram* of elements of a situation showing that their contents are to be computed. Morphological marking on a dependent *points to* its head whereas marking on a head *points to* a dependent. Adjacency *reflects* co-constituency and is often accompanied by fixed constituent order that *diagrams* a (default) head-dependent relation.

F.   Allomorphy in inflection. Morphological (morphophonemic) alternants – in addition to their grammatical or lexical content – are endophoric Indexes: They *point to* expression elements, morphological features, or grammatical or lexical content elements in their environments (se Section 5). They are a morphosyntactic counterpart of phrase and clause level agreement.
Social and stylistic morphological variables provide both endophoric and exophoric Indexes.

G.   Phonology. Endophoric Indexes in phonology were first recognized by Trubetzkoy (1958: 95), who characterized the function of allophonic features as 'auxiliary-sociative'. Social and stylistic phonological variables provide endophoric and exophoric Indexes.

The following pages present some examples of the diverse index values that are typically part of morphological signs (Section 5). A few of the perennial problems in morphological theory for which the triadic sign is relevant will be reviewed in Section 6. And we will look at the wider implications of recognizing this aspect of morphological signs in Section 7.

5.   Indexes in morphology

Andersen (2008, pp. 28–29) defines two sources of morphological indexes, reanalysis of phonological alternations (index grammation from below) and reanalysis of lexical or grammatical expressions (index grammation from above). This section exemplifies morphological index types, phonological, stem-class, gramcat, and lexcat indexes.

5.1   Phonological indexes

Affix allomorphy may *point to* phonolgical features in a stem.

The most elementary phonological indexes, perhaps, are vowel harmony, well known from Uralic and Turkic languages.

Reflexes of the Common Slavic vowel fronting after palatal consonants provide another example. In Old Church Slavonic, some declensions have two sets of

desinences; for each case, then, there is a paradigm of two allomorphs; one *points to* stem-final 'palatals' (…č, …š, …ž, …št, …žd, …ń, …ľ, …ŕ, …j) (1.a); the other (1.b) *indicates* stem-final nonpalatal consonants; for some cases there is a single desinence (here -*a*) whose initial vowel appears neutral (at least in the standard orthography); Huntley (1993, p. 140).

(1) Old Church Slavonic. Case endings in the Second 'two-fold' declension

	(a)	(b)
NOM.SG	*duš-a* 'soul'	*glav-a* 'head'
GEN.SG	*duš-ę*	*glav-y*
DAT.SG	*duš-i*	*glav-ě*
INS.SG	*duš-ejǫ*	*glav-ojǫ*
NOM.PL	*duš-ę*	*glav-y*
GEN.PL	*duš-ĭ*	*glav-ŭ*

Stem allomorphy may *point to* phonological features of affixes.

In some Italian verbs, IND.PRS.1SG and IND.PRS.3PL have a longer stem allomorph in …g- (2). The same allomorph occurs in some forms of the subjunctive. It appears that the allomorph in …g- occurs before back vowels: It is a phonological Index of back-vowel desinences. The …g- is the outcome of an old sound change; its phonetic motivation is long lost, but it was codified as a phonological Index, a value it presumably retains; see further Section 7.1.

(2) Italian stem alternants in Present Indicative and Subjunctive

	Indicative	Subjunctive
1SG	*veng-o* 'I come'	*veng-a*
2SG	*vien-i*	*veng-a*
3SG	*vien-e*	*veng-a*
1PL	*ven-iamo*	*ven-iamo*
2PL	*ven-ite*	*ven-iate*
3PL	*veng-ono*	*veng-ano*

Phonological indexes can develop from gramcat indexes.

In a productive class of verbs in Common Slavic the stem-final consonant underwent Grade 3 Mutation before the PRS.1SG desinence; cf. (3.A). In Old Polish, this mutation was extended to PRS.3PL. The two desinences had no content in common that would motivate such an extension, but both had an initial nasal vowel; see (3.A–B). The extension, then, resulted from a reanalysis of the stem mutation's index value; it was reanalysed from grammatical (/___+1SG) to phonological (/___-ã). A parallel example in Ukrainian is described in Andersen (1980, p. 33).

(3) Grade 3 Mutation extended in Old Polish and in Ukrainian:

	A	B	C
	LCS	Old Polish	Modern Polish
1SG	*noš-ǫ*	*noš-ã*	*noš-ę*
2SG	*nos-i-šĭ*	*noś-i-š*	*noś-i-š*
3SG	*nos-i-tĭ*	*noś-i*	*noś-i*
1PL	*nos-i-mŭ*	*noś-i-my*	*noś-i-my*
2PL	*nos-i-te*	*noś-i-će*	*noś-i-će*
3PL	*nos-ę-tĭ*	*noš-ã:*	*noš-ą*

5.2   Stem-class indexes

Stem-classes are typically indexed by interfixes (see Section 6.1), desinence allomorphs, or mutations. The Latin examples in (4) show the paradigm of desinence allomorphs for the gen.sg; each is selected to *point to* a given declension.

(4) Latin gen.sg allomorphs:
→ -ae /Decl. 1 +_; e.g., *port-ae* f. 'door', *naut-ae* m. 'sailor'
→ -ī /Decl. 2 +_; e.g., *domin-ī* m. 'lord', *fāg-ī* f. 'beech', *templ-ī* nt. 'temple'
→ -is /Decl. 3 +_; e.g., *pont-is* m. 'bridge', *virgin-is* f. 'virgin', *mar-is* nt. 'sea'
→ -ūs /Decl. 4 +_; e.g., *cant-ūs* m. 'song', *trib-ūs* f. 'tribe', *corn-ūs* nt. 'horn'
→ -ei /Decl. 5 +_; e.g., *di-ēi* m. 'day', *fid-ei* f. 'faith'.

5.3   Gramcat indexes

Indexes of grammatical categories are typically stem allomorphs or mutations that *point to* affixal categories, or affix allomorphs *pointing to* grammatical categories of stems or of other affixes.

In German verbs, vowel mutations *point to* categories of mood and tense, e.g.,

Gm. *hat-t-e*.IND.PST.3SG   vs.   *hät-t-e*.SBJ.PST.3SG,
stand-Ø.IND.PST.3SG   vs.   *stünd-e*.SBJ.PST.3SG.

Synchronically English has three morphological verb classes; from an historical perspective they form three chronological strata. First, there is the productive (weak) class with the /-d, -t, -əd/ paradigm of suffix allomorphs in the (finite and nonfinite) Past forms and an invariant stem (*walked*). Secondly, there is the unproductive (weak) class with the /-d, -t/ allomorphs in the Past forms and stem allomorphy (*bought*). Thirdly, there are the strong verbs, unsuffixed in the finite Past, with or without stem allomorphy (*eat-ate* vs. *beat-beat*), some with /-ən/ (*woven*) or a

third stem allomorph (*sung*) in the nonfinite Past. Typologically, the three classes exemplify agglutination, fusion, and symbolism.

The second class is the most interesting. These 'dental preterites' have stem mutations that *point to* the (finite and nonfinite) Past suffixes (5). They occur (a) with stem-vowel mutation and the regular suffix allomorphs /-d, -t/, some with stem-vowel mutation and a stem-final consonant mutation as well;[3] (b) with stem-vowel mutation and a /-t/ suffix; (c) with invariant stem vowel but deletion of stem-final …d before /-t/; and (d) with stem-vowel mutation and deletion of stem-final …(C)C before /-t/.

(5)   a.   *sell – sol-d, tell – tol-d, keep – kep-t, leap – leap-t; leave – lef-t, lose – los-t*
       b.   *dream – dream-t, lean – lean-t, kneel – knel-t*
       c.   *lend – lent, rend – rent, spend – spent, build – built*
       d.   *bring – brought, catch – caught, teach – taught, think – thought*

In the current synchronic state, these fusional Past forms can still be dissected into stem and suffix, but their small number and diversity makes one wonder whether their stem allomorphs still *point to* the Past suffix; perhaps they tend to be reanalysed and used as lexicalized unit symbols. We return to them in Section 7.

Polish has two sets of regular person allomorphs that *point to* the categories of tense; one set *points to* the Present stem; the other *points to* the gender/number suffixes of the Past stem (6), e.g., *mieszka-Ø-š*.PRS.2SG 'you live', *mieszka-ł-a-ś*. PST.F:SG.2SG 'you resided' (the latter also occurs, exceptionally, in the Present of 'to be' (Andersen, 1987, p. 34).

(6)   Polish person and number allomorphs
            1SG   2SG 3SG 1PL   2PL   3PL
       Present: -ę, -m -š   -Ø -my -će -ą
       Past: -m   -ś   -Ø -śmy -śće -Ø

In the regular inflection of verbs in Irish, initial Lenition *points to* Imperfect, Past, and Conditional desinences; cf. Present and Imperfect in (7) (Dochartaig, 1992, p. 70).

---

3.   Past forms of the types *kep-t*, *lef-t* and *los-t* are often mentioned as having the suffix /-t/, exemplified under (b), thus Embick (2015: 92). But synchronically they are phonologically parallel to *topped*, *coughed*, *tossed*, with the regular /-d, -t/ allomorphs. Synchronically, their …v →…f and …z → …s mutations are the reverse of the nominal alternations in *knife – knives*, *house – houses*.

(7) Irish Present and Imperfect, Imperative

	Present	Imperfect
1SG	*tógaim* 'I take'	*thógainn*
2SG	*tógair/tógann tú*	*thógthá*
3SG	*tógann sé/sí*	*thógadh sé/sí*
1PL	*tógaimid/togann muid*	*thógaimis/ thógadh muid*
2PL	*tógann sigh*	*thógadh sibh*
3PL	*tógaid/tógann siad*	*thógaidis/ thógadh siad*
Impersonal	*tógthar*	*thógthaí*

An example that has recently been discussed in the literature (Corbett, 2016, p. 67) is the Present paradigm of one class of Russian verbs. Following Aronoff (1994), the set of wordforms with -o- endings (written ë in (8.A)) is characterized as a 'morphome'. A closer look, however, reveals three notable features that escape notion in that superficial approach.

First, the 1sg ending is a person marker; its /-u/ occurs with this content in all but two Russian verbs (the exceptions being *da-m*.PRS.1SG 'give', *je-m*.PRS.1SG 'eat' and their prefixal derivatives). This means that the 1sg ending can be aligned with the other person markers as in (8.B). This alignment implies a segmentation that isolates the vowel that intervenes between stem and person (participant) marker.

(8)

	A	B	C
	Present	Present	Gerund
1SG	*pek-ú* 'I bake'	*pʲek-Ø-ú*	
2SG	*peč-ëš*	*pʲeč-ó-š*	
3SG	*peč-ët*	*pʲeč-ó-t*	
1PL	*peč-ëm*	*pʲeč-ó-m*	
2PL	*peč-ëte*	*pʲeč-ó-tʲe*	
3PL	*pek-út*	*pʲek-ú-t*	*znáj-a*
	Past	Past	
M.SG	*pëk*	*pʲók-Ø-Ø*	
N.SG	*pekl-ó*	*pʲek-l-ó*	
F.SG	*pekl-á*	*pʲek-l-á*	
PL	*pekl-í*	*pʲek-lʲ-í*	*zná-v*

Secondly, the resulting segmentation (8.B) highlights the fact that PRS.3SG and PRS.3PL are homonymous but differ by this intervening vowel. In other words, one of the four o-vowels in the supposedly meaningless morpheme is not totally devoid of meaning: it forms the background for the PRS.3PL /-ú-/ which *points to* the number of the PRS.3PL.

Thirdly, in the Past paradigm the verb stem should be separated from the Past suffix; contrast (8.A) and (8.B). Its /-l-/ has a regular alternant /-lʲ-/ in the Plural and

is dropped (represented by -Ø-) in the Masculine in some twenty obstruent-stem verbs, such as this one. The fact that the Past morpheme's /-Ø-, -l^j-, -l-/ alternants contrast with the /-Ø-, -ú-, -ó-/ that intervene between stem and participant marker in the Present suggests the interpretation that these alternants *stand for* 'Present': Past is expressed by consonants (and -Ø-), but Present by vowels (and -Ø-). This interpretation appears to be corroborated by forms of the Gerund (here from *znat'* 'know'), which distinguishes Past and Present precisely with a consonant ending in the Past and a vowel ending in the Present.

We conclude that the Present is expressed by -Ø- in prs.1sg, /-ú-/ in prs.3pl, and /-ó-/ elsewhere; the /-ú-/ *stands for* Present and *points to* 3pl, the /-ó-/ *stands for* Present. The Past is expressed by -Ø- in the masculine (of these verbs), /-l'-/ in the plural, and /-l-/ elsewhere; /-l'-/ *stands for* Past and *points to* Plural, /-l-/ *stands for* Past.

Russian dialects show that speakers in different areas have had different ideas of a rational distribution of these Present allomorphs (Avanesov & Orlova, 1965, p. 163; Andersen, 1980, p. 28). That diversity cannot be reviewed here, but it is significant as evidence of speakers' concern with morphological detail, evidence that should not be ignored by students of morphology, whether diachronic or synchronic.

5.4   Lexcat indexes

In Russian, there are two allomorphs in the accusative in several declensions (one singular noun declension, the plural declension of nouns, the masculine singular and the plural declension of adjectives and adjectival pronouns). In Decl. 1a, for instance, the acc.sg allomorphs are -a, -Ø. The -a allomorph *points to* animate nouns, the -Ø- *points to* inanimate nouns (9.a). Animacy is a lexical feature, typically part of a noun's semantic description. But the allomorphs can be selected by speakers to make explicit the (in)animacy of a referent. Typical examples are borderline animacy cases, such as bivalves (*ustrica* 'oyster' is inan as food) and microbes. Besides, the allomorphy can be exploited in metaphor; e.g., *Narjadí pnj-a*.ACC[ANIM].SG *i pen'-Ø*.NOM.SG *budet xoroš* 'Dress up a stump[anim], and the stump will look good' (a version of *Clothes make the man*).

Russian patrials are derivatives in /-an-/ that denote residents of localities (*Pariž-an-in-Ø* 'Parisian') or members of social groups (*ljuter-an-in-Ø* 'Lutheran', *politkatorž-an-in-Ø* 'political prisoner') (9.b). They are characterized by a predesinential singulative suffix /-in-/ in the Singular and the nom.pl desinence /-e/, which only occurs with patrials. Both the predesinential singulative suffix and the special nom.pl allomorph *point to* the lexical class of the stem. In other singulatives the

/-in-/ suffix is followed by another derivational suffix (*sneg* 'snow' – *snež-in-k-a* 'snow flake'); it is then not predesinential and does not *point to* a patrial stem.

In older Russian, many Decl. 1a mass nouns had /-u/ in gen.sg, but they could be converted to count nouns and then had /-a/ (9.c); thus, still, *čaška čaj-u*.GEN.SG 'cup of tea' vs. *čaška kitajskogo čaj-a*.GEN.SG 'cup of Chinese tea'; in the former the allomorph /-u/ *points to* the mass noun *čaj*; in the latter the adjective implies the existence of several kinds of tea and /-a/ *points to* the count noun *čaj*.

(9)   Russian lexcat indexes
    a.   Animacy
       ACC.SG: *vólk-a* ANIM. 'wolf'   *stól-Ø* INAN. 'table'
       ACC.PL: *volk-óv*   *stol-ý*
    b.   Patrial and social-group terms
       NOM.SG *armjan-ín-Ø* 'Armenian' *gražd-an-ín-Ø* 'citizen'
       NOM.PL *armján-e*   *grážd-an-e*
    c.   Mass vs. count noun, gen.sg
       Mass noun   *čáj-u* 'tea' *sáxar-u* 'sugar' *sýr-u* 'cheese'
       Count noun *čáj-a*   *sáxar-a*   *sýr-a*

6.   The importance of morphological indexes

Recognizing the index values of morphological signs opens up a long neglected dimension of meaning in morphology, gives us access to the actual richness of morphological signs, and offers us an understanding of their contribution to communication. It makes it clear that morphological alternations, far from being an unnecessary, meaningless encumbrance, are semiotic features that make morphology isomorphous with other levels of linguistic structure.

This section will look at three of the classic deviations from the simple Saussurean *signifiant-signifié* relation. Recognizing the triadic structure of the linguistic sign puts the tradition's 'empty morphs' (Section 6.1), 'zero signs' (Section 6.2), and overlapping exponence (Section 6.3) in a new light.

6.1   Signs with zero exponenda

The content of the triadic sign is not exhausted by its exponendum. Its syntactic properties are a kind of content, in as much as they define its index value.

Signs with zero exponenda ('empty morphs') are called interfixes (Mel'čuk, 2006, p. 298). By virtue of their syntactic properties they define the morphosyntactic relations of contiguous constituents and thereby link them.

In stem formation, an interfix may *point to* a class of stems and a class of desinences, defined in phonological terms. Thus the Italian …g- in (2), which *points to* (one of) a set of verbs with stem-final …n or …l and sets of back-vowel initial desinences, can be interpreted as an interfix /-g-/. A pre-desinential interfix may *point to* a class of inflectional allomorphs that defines the given inflectional class (10.a). Interfixes themselves may have allomorphs; an example is the /-ā-/ of Lat. *am-ā-re*, which alternates with /-a-/ (e.g., *am-a-t*.IND.PRS.3SG) and -Ø- (e.g., *am-Ø-ō*.IND.PRS.1SG, *am-Ø-e-m*.SBJ.PRS.1SG.

In compounds, interfixes indicate (language-specific) specifier-head relations. In German, interfix allomorphs are often homophonous with plural suffixes but are in principle indifferent to number (10.b).

(10) Interfixes in stem formation and compounding
   a. Latin verb class interfixes
       Lat. Inf. *am-ā-re, hab-ē-re, duc-e-re, fin-ī-re*;
   b. German interfixes in nominal compounds
       Gm. (*Liebe* →) *Liebe-**s**-paar*, 'pair of lovers', (*Held* →) *Held-**en**-leben* 'a hero's life', (*Kind* →) *Kind-**er**-hemd* 'child's (or child-size) shirt', (*Gans* →) *Gäns-**e**-feder* 'goose-feather', (*Buch* →) *Büch-er-regal* 'book-case', *Buch-Ø-messe* 'book fair' ….

Aronoff (1994, p. 37) has drawn attention to the Latin 'third stem', the stem of the traditional Past or Perfect Passive participle ending in -t- (~ -s-), e.g., *amā-t-us* 'loved', *habi-t-us* 'had', *duc-t-us* 'led', *finī-t-us* 'ended'. It occurs in a number of deverbal formations in which it cannot be ascribed Passive, Past tense, or Perfective aspect content (11).

(11) Lat. *amō* 'love'   *amā-t-ūr-us* fut.ptcp 'who will love'
                           amā-t-um supine 1 'in order to love'
                           amā-t-ū supine 2 'with regard to loving'
      canō 'sing'   *can-t-or* 'singer'
      cogitō 'think'   *cogitā-t-iō* 'thought'
      scrībō 'write'   *scrip-t-ūra* 'writing'
                           scrip-t-itō (frequentative) 'write often'
      emō 'buy'   *emp-t-uriō* (desiderative) 'want to buy'
      iaciō 'throw'   *iac-t-ō* (intensive) 'fling'

Aronoff concluded that its /-t-/ affix is an empty morph without any specific syntactic function (a 'morphome').

However, at second glance, one can see an ascriptive adjective that appears passive when formed from transitive verbs, e.g., *amātus* '(be)loved', but active when formed from deponent verbs, e.g., *ūsus* 'who has used', *frūctus* 'who has enjoyed'.

The /-t-/ affix also occurs in a few denominal adjectives, e.g., (*barba* 'beard' →) *barbātus* 'bearded', (*cēna* 'dinner' →) *cēnātus* 'who has had dinner'.

But in many formations in which the /-t-/ is followed by another derivational suffix, it is an interfix, a purely indexical sign that serves the morphosyntactic function of connecting a verb stem to any of a closed inventory of vowel-initial suffixes that form deverbal adjectives (*amātūrus*), adverbs (the Supines), nouns, and verbs, as the list in (11) shows.[4]

The well-known initial mutations in Celtic languages can have the function of interfixes. They typically have no exponendum but serve to link coconstituents by *pointing to* their respective features. (12) shows the citation form of Welsh /kiː/ 'dog', (b) shows Soft, (c) Nasal, and (d) Aspirate mutations that *point to* the respective pronouns (Awbery, 1983b).

(12) a.   /ejn kiː/ 'our dog'
      b.   /də giː/ 'your (sɢ) dog'
      c.   /və ŋhiː/ 'my dog'
      d.   ej xiː/ 'her dog'

6.2   Signs with zero expression

Signs with zero expression ('zero signs'), have long seemed more problematic than they really are (Mel'čuk, 2002). But they do need to be elucidated from several points of view.

First of all, all skepsis whether there is a valid distinction between a zero and nothing should be abandoned (Matthews, 1972, p. 431). The distinction is plain enough in the contrast between *£50* and *£5*. The 5 in *£50* is followed by a zero, that in *£5* is followed by nothing.

The final zero in *£50* (a) signifies 'no ones' and (b) by its presence *indicates* that the 5 signifies the number of tens. The nothing that follows *£5* signifies nothing, and so the 5 in *£5* signifies 'five ones'.

A language example would be the contrast between Russ. *osen'* 'autumn' and *očen'* 'very'. The lexeme osen' has a bound stem /osʲenʲ-/ capable of hosting a paradigm of desinences. When a wordform of osen' does not end in /-i/, /-ju/, /-ej/, /-ax/, /-am/, or /-amʲi/, it ends in -Ø. The wordform *očen'* of očen', by contrast, is an indeclinable /očenʲ/ that cannot host any desinence, let alone a zero desinence, but can be followed by nothing.

Secondly, an important point to mention in this connection is that morphosyntactic rules define position classes, and that these form a serial structure of indexes,

---

4.   Steriade (2016) shows the extent to which derivatives formed from the Latin -t-stem had passive meaning.

comparable to the serial structure of indexes in numbers that tells us which digit represents ones, which tens, which hundreds, etc. Reading Russ. *osen'ju* from right to left tells us that /-ju/ is a desinence and /os^{j}en^j-/ its stem; reading *osen'* from right to left tells us that the bound stem /os^{j}en^j-/ here precedes a zero desinence allomorph. It is the existence of position classes that explains how also -Ø- allomorphs can *point to* elements in their environment, as in (8).

Thirdly, there is an important distinction between zero allomorphs and zero as the sole exponent of a sign.

Most often zero expressions are allomorphs. In the case of /os'en'-Ø/ (Decl. 3), its desinence is part of two allomorph paradigms: (a) the NOM.SG allomorphs (Decl. 1.a -Ø, Decl. 1.b /-o/, Decl. 2 /-a/, and Decl. 3 -Ø); and (b) the acc.sg allomorphs (Decl. 1.a /-a/, -Ø (cf. (9.a)), Decl. 1.b /-o/, Decl. 2 /-u/, and Decl. 3 -Ø).

But consider ellipsis. Ellipsis is a type of zero expression, most often a stylistic variant. But a stylistic ellipsis can become reanalysed. An example is the verbal negation in southern Welsh, which is now regularly omitted, e.g., /ni ðaw/ ~ /ðaw/ '(s)he won't come' (/daw/ '(s)he will come') (Awbery, 1983a, p. 163); the verb-initial Soft mutation (d → ð) continues to *point to* the ellipted negation; or perhaps the mutation has already been reanalysed as the exponent for negation and *ni*, as a stylistic marker that *points to* the negated, Lenited verb form; compare the parallel in Fr. *Elle va pas venir ~ Elle ne va pas venir*.

Signs that have an invariant zero expression are worth a longer discussion than the topic can be accorded here; cf. Mel'čuk (2002).

6.3   Cumulative exponence

Cumulative exponence was discussed by Matthews (1972, p. 135) as an argument against a morpheme-based morphology. His example, cited here as Table 1, was intended to demonstrate the uselessness of morphemic analysis, but it merely illustrates the inadequacy of the Saussurean sign concept. Table 2 shows that if the requisite expression rules are written with due attention to the larger system of conjugation (see Andersen, 2010, p. 3), Matthew's seemingly senseless criss-cross of relations is resolved into exponendum relations (vertical) and index relations (slanting arrows).

Table 1. Latin *kukurristī* 'you.SG ran, have run'

Grammatical representation:	CURR-	+	Perfective	+	2nd	+	Singular
Phonological representation:	ku	+	kurr	+	is	+	ti:

Table 2. Latin *kukurristī* 'you.SG ran, have run'

Content:	'run'	+	Perfective	+	Present	+	2nd	+	Singular
Expression:	kukurr-	+	-is-	+	Ø	+	-tī-	+	-Ø

7. History and typology

The aim of this paper was to draw attention to the failure of synchronic theories of morphology to take a systematic interest in morphology below the level of the word, and the need for historical linguistics to take up the challenge of accounting for the major, typological drifts in morphological techniques, from agglutination to fusion to symbolism, and the devolution back to agglutination or isolation; cf. Hodge (1970); Werner (1987).

The triadic sign conception presented and exemplified here (Sections 5–6) facilitates the detailed analysis and interpretation of the kinds of change involved in the typological drifts and is an essential tool for this task. Indexes are essential in the creation (through reanalysis) of portmanteau, cumulative, and overlapping signs. Affix indexing of grammatical stem features and stem indexing of grammatical affixes are steps towards fusion. The reanalyses of stem indexes as exponenda and the corresponding loss of affix exponenda leads to symbolism through a stage in which affixes have only stylistic index value; cf. Korhonen (1969, 1982).

The English verb, despite its simplicity, is instructive. In Old English, the strong verbs could still be viewed as fusional: stem mutations *pointed to* strong-verb desinences and their desinences *pointed to* the strong-verb stems. With the leveling of the Present inflection and the loss of the rest of the system of person/number inflection, strong verb morphology became symbolic, with portmanteau exponents of both lexical and tense exponenda.

The diachronic study of morphological techniques has been troubled by the fact that synchronic systems are not typologically consistent; cf. Greenberg (1960). The language historian can turn this inconsistency to advantage when different techniques are reflected in distinct chronological strata (see Section 6.3). Such cases permit a clear differentiation between the productive parts of a system that reflect its synchronic type vs. typologically divergent patterns from its past that owe their continued existence to conservative norms of usage; cf. Coseriu (1970).

Abbreviations

ACC	accusative	NOM	nominative
DAT	dative	PL	plural
F	feminine	PRS	present
FUT	future	PST	past
GEN	genitive	PTCP	participle
Gm.	German	SG	singular
IND	indicative	SBJ	subjunctive
INS	instrumental	1	first person
LCS	Late Common Slavic	2	second person
M	masculine	3	third person
N	neuter		

References

Andersen, H. (1980). "Morphological change: towards a typology". In J. Fisiak (Ed.), *Recent developments in historical morphology* (pp. 1–50). The Hague: Mouton. https://doi.org/10.1515/9783110823127.1

Andersen, H. (1987). From auxiliary to desinence. In: M. Harris & P. Ramat. (Eds.). *Historical development of auxiliaries* (pp. 21–52). Berlin: Mouton de Gruyter.

Andersen, H. (2008). Grammaticalization in a speaker-oriented theory of change. In Thórhallur Eythórsson (Ed.), *Grammatical change and linguistic theory. The Rosendal Papers* (pp. 11–44). (Linguistik Aktuell/Linguistics Today, 113.) Amsterdam & Philadelphia: John Benjamins. https://doi.org/10.1075/la.113.02and

Andersen, H. (2010). From morphologization to demorphologization. In V. Bubenik & S. Luraghi (Eds.), *Companion to historical linguistics* (pp. 117–146). London: Continuum Companions in Linguistics.

Anderson, S. R. (1992). *A-morphous morphology*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511586262

Anttila, R. (1975). *The indexical element in morphology*. Innsbuck: Institut für Sprachwissenschaft.

Aronoff, M. (1994). *Morphology by itself. Stems and inflectional classes*. Cambridge, MA & London, England: MIT Press.

Avanesov, R. & Orlova, V. (1965). *Russkaja dialektologija*. Moskow: Nauka.

Awbery, G. M. (1983a). Moves towards a simpler, binary mutation system in Welsh. In H. Andersen (Ed.). *Sandhi phenomena in the languages of Europe*. (pp. 161–166). Berlin & New York & Amsterdam: Mouton de Gruyter.

Awbery, G. M. (1983b). Survey of sandhi types in Welsh. In H. Andersen (Ed.). *Sandhi phenomena in the languages of Europe*. (pp. 415–434). Berlin / New York / Amsterdam: Mouton de Gruyter.

Blevins, J. (2016). *Word and Paradigm morphology*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199593545.001.0001

Bloomfield, L. (1935). *Language*. London: George Allen & Unwin.

Carstairs-McCarthy, A. (2002). How stems and affixes interact: Stem alternants as morphological signata. In Bendjaballah et al. (Eds.) *Morphology 2000. Selected papers from the 9th Morphology Meeting, Vienna, 2000.* (pp. 49–58). Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/cilt.218.05car

Corbett, G. (2016). Morphomic splits. In A. Luís & R. Bermúdez-Otero (Eds.), *The morphome debate* (pp. 64–88). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198702108.003.0004

Coseriu, E. (1970). Synchronie, Diachronie und Typologie. In *his Sprache. Strukturen und Funktionen. 12 Aufsätze zur allgemeinen und romanischen Sprachwissenschaft.* Ed. by U. Petersen, H. Bertsch & G. Köhler (pp. 91–108). Tübingen: Gunter Narr.

Dahl, Ö. (2004). *The growth and maintenance of linguistic complexity.* Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/slcs.71

Dochartaigh, C. (1990). The Irish language. In D. Macaulay (Ed.), *The Celtic languages* (pp. 11–99). Cambridge: Cambridge University Press.

Dressler, W. (1977). *Grundfragen der Morphonologie.* Vienna: Österreichische Akademie der Wissenschaften.

Dressler, W. (1985). *Morphonology. The dynamics of derivation.* Ed. by Kenneth C. Hill. Ann Arbor: Karoma.

Embick, D. (2015). *Morpheme. A theoretical introduction.* Boston: DeGruyter Mouton.

Greenberg, J. (1960). A quantitative approach to the morphological typology of languages. *International Journal of American Linguistics* 26, 178–194. https://doi.org/10.1086/464575

Hodge, C. (1970). The linguistic cycle. *Language Sciences* 13, 1–7.

Huntley, D. (1993). Old Church Slavonic. In B. Comrie & G. C. Corbett (Eds.), *The Slavonic languages* (pp. 125–187). London & New York: Routledge.

Jakobson, R. (1971). Shifters, verbal categories and the Russian verb. In *his Selected Writings, 2. Word and Language* (pp. 130–147). The Hague & Paris: Mouton.

Jakobson, R. (1990). *On language.* Ed. by Linda R. Waugh & Monique Monville-Burston. Cambridge, MA & London: Harvard University Press.

Korhonen, M. (1969). Die Entwicklung der morphologischen Methode im Lappischen. *Finnisch-ugrische Forschungen* 37 (pp. 203–362). Reprinted as Die Entwicklung der morphologischen Technik im Lappischen. In: Korhonen (1996, pp. 17–142).

Korhonen, M. (1982). Reductive phonetic developments as the trigger to typological change: Two examples from the Finno-Ugrian languages. In A. Ahlqvist (Ed.), *Papers from the 5th International Conference on Historical Linguistics, Galway, April 6–10, 1981.* (pp. 207–212). Amsterdam / Philadelphia: John Benjamins. Reprinted in Korhonen (1996, pp. 207–212). https://doi.org/10.1075/cilt.21.23kor

Korhonen, M. (1996). *Typological and historical studies in language. A memorial volume published on the 60th anniversary of his birth.* Ed. by T. Salminen. Helsinki: Suomalais-ugrilaisen Seura.

Matthews, P. (1972). *Inflectional morphology. A theoretical study based on aspects of Latin verb conjugation.* Cambridge: Cambridge University Press.

Matthews, P. (1974). *Morphology: An introduction to the theory of word structure.* Cambridge: Cambridge University Press.

Matthews, P. (1991). *Morphology.* (Expanded and revised version of Matthews 1974.) Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139166485

Meľčuk, I. (2002). Towards a formal concept of 'zero linguistic sign'. Applications in typology. In S. Bendjaballah et al. (Eds.) *Morphology 2000. Selected papers from the 9th Morphology Meeting, Vienna, 2000*. (pp. 241–258). Amsterdam / Philadelphia: John Benjamins. https://doi.org/10.1075/cilt.218.20mel

Meľčuk, I. (2006). *Aspects of the theory of morphology*. Ed. by D. Beck. Berlin/New York: Mouton de Gruyter.

Nöth, W. (1990). *Handbook of semiotics*. Bloomington / Indianapolis: Indiana University Press.

Pierce, C. S. (1960–1966). *Collected papers*, 1–8. Ed. by C. Hartshorne & P. Weiss. (Second Printing.) Cambridge, MA: Harvard University Press.

Saussure, F. de. (1973). *Cours de linguistique générale*. Ed. by C. Bally & A. Sechehaye with A. Riedlinger. Critical edition by T. de Mauro. Paris: Payot.

Shapiro, M. (1969). *Aspects of Russian morphology: A semiotic investigation*. Cambridge, MA: Slavica.

Shapiro, M. (1991). *The sense of change: Language as history*. Bloomington: Indiana University Press.

Spencer, A. (2016). Stems, the morpheme, and meaning-bearing inflection. In A. Luís & R. Bermúdez-Otero (Eds.). *The morpheme debate* (pp. 207–227). Oxford: Oxford University Press.

Steriade, D. (2016). The morphome vs. similarity-based syncretism. In A. Luis & R. Bermúdez-Otero (Eds.), *The morphome debate* (pp. 112–172). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198702108.003.0006

Stewart, T. (2016). *Contemporary morphological theory: A user's guide*. Edinburgh: Edinburgh University Press.

Stump, G. (2001). *Inflectional morphology. A theory of paradigm structure*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486333

Trubetzkoy, N. (1958). *Grundzüge der Phonologie*, Göttingen: Vandenhoeck & Ruprecht.

Werner, O. (1987). The aim of morphological change is a good mixture, not a uniform language type. In A. Giacalone Ramat, O. Carruba & G. Bernini (Eds.), *Papers from the 7th International Conference on Historical Linguistics* (pp. 591–606). Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/cilt.48.43wer

Appendix. Some terminology and definitions

Icon   An Icon is a sign that *represents* its object by virtue of its similarity to the object. Icons *depict* their objects.[5]
"Anything whatever is an icon of anything in so far as it is like that thing and used as a sign of it." (2.247) [The icon acts as a sign] "by virtue of characters which belong to it in itself as a sensible object and which it would possess just the same were there no object in nature that it resembled, and though it never were interpreted." (4.447)

---

5. The following references are to volume and section in Peirce 1960–1966.

Index	An Index is a sign that represents its object by virtue of its contiguity or "existential relation" to the object. Indexes *draw attention to*, *point to*, *indicate* their objects. [Indexes] "refer to individuals … they direct attention to their objects by blind compulsion." (2.306) "In so far as the index is affected by the object, it necessarily has some quality in common with the object, and it is in respect to [this] that it refers to the object." (2.248) [The index] "is a real thing or fact which is a sign of its object by virtue of being connected with it as a matter of fact." (4.447)
Symbol	A Symbol is a sign that represents its object by virtue of a rule that warrants this interpretation. Symbols *stand for* their objects. "A symbol is constituted a sign merely or mainly by the fact that it is used and understood as such, whether the habit is natural or conventional, and without reference to the motives which originally governed its selection." (2.307) [Its] "fitness to represent just what it does represent lies in nothing but the very fact of there being a habit, disposition, or other effective general rule that it will be so interpreted." (4.447)
Hypoicons	Image, Diagram, Metaphor. The Image *resembles* its object, the Diagram *reflects* relations in its object, the Metaphor *suggests* its object. "Hypoicons may be roughly divided … [as follows]; those which partake of simple qualities … are images; those that represent the relations, mainly dyadic, or so regarded, of the parts of one thing by analogous relations in their own parts, are diagrams; those which represent the representative character of a representamen by representing a parallelism in something else, are metaphors." (2.277)

Ablaut pattern extension as partial regularization strategy in German and Luxembourgish

Jessica Nowak

University of Mainz

The present contribution considers the nature of ablaut pattern extension and the determinants contributing to it. Following Morphological Economy Theory (e.g., Werner, 1987), it argues that analogical pattern extensions are highly functional, i.e., they constitute frequency-driven partial regularization strategies within verbal systems. This assumption is supported by diachronic and empirical data from two case studies: In German, low-frequency strong verbs reduced stem allomorphy by adopting the Class II pattern [$o_{\text{PRET}}=o_{\text{PP}}$], a partial leveling that may precede the weakening of strong forms (e.g., *bellen – ball – gebollen* → *boll – gebollen* → *bellte – gebellt* 'to bark'). In Luxembourgish, a uniform preterite marker *ou* emerged as a consequence of drastic preterite loss.

Keywords: ablaut pattern extension, regularization, partial leveling, usage-based approach, German, Luxembourgish

1. Introduction

Ablaut pattern extensions are well-documented in the histories of West Germanic languages (see Durrell, 1999, 2001); compare, for instance, the analogical extension of the *cling*/*clung*-pattern to formerly weak verbs like *fling, sting, string* in English and its varieties in (1) (e.g., Bybee/Moder, 1983; Anderwald, 2009):

(1) analogical basis: e.g., *cling – clung – clung*
      string – stringed – stringed → *strung – strung*
      also: *fling, sling, sting*

Such ablaut pattern extensions are usually explained by the power of analogy, i.e. they are motivated by pure similarity with the analogical basis, in the cases under

(1) by the shared vowel+velar nasal in the infinitive (INF). These cases of analogy are rarely considered to be functionally motivated in terms of a regularization process, in the case of weak verbs becoming strong these changes are even classified as irregularizations (e.g., Marcus et al. 1992, p. 176). This assessment is rather surprising because the ablaut patterns that were analogically extended to other (mainly strong) verbs were usually structurally less complex, i.e., they exhibited only two different stem vowels, and in many cases, their analogical extension resulted in a reduction of stem allomorphy, and thus, in partial leveling; compare the extension of the Class II pattern $[o_{PRET}=o_{PP}]$ to several strong verbs from different ablaut classes in German, which reduced the inherited three-vowel alternations in verbs like *melken* 'to milk' and *heben* 'to heave' in (2):

(2) analogical model: Class II *fliegen*:   *flog – geflogen*
Class IIIb *melken*:   *malk – gemolken → molk – gemolken*
Class VI *heben*:   *hub – gehaben* → *hob – gehoben*

As will be shown in Section 2.2, the adoption of the $[o_{PRET}=o_{PP}]$ pattern constitutes a frequency-driven partial regularization strategy within strong verbs, and is thus functionally motivated. Another remarkable case of ablaut pattern extension is documented in Luxembourgish, see (3): Here, a uniform preterite with *ou* stemming from Classes II+VI emerged for all strong verbs, hence reducing the preterite formation to one single type.

(3) analogical model: Class II   *zéien – zouch – gezunn*   'to pull'
Class VI   *stoen – stoung – gestan(en)*   'to stand'
Cl. I   *bliwwen – blouf – bliwwen*   'to stay'
Cl. IIIa   *kléngen – kloung – geklongen* 'to sound'
Cl. VII   *falen – foul – gefall*   'to fall'

The fact that Luxembourgish significantly diverges from the ablaut pattern extensions usually found across West Germanic languages,[1] where a vowel alternation pattern for the preterite (PRET) and the past participle (PP) was extended as in the German case above can be best accounted for from a functional and usage-based perspective: Paradigm leveling yielding stem allomorphy reduction or uniformity is a matter of token and/or category frequencies (see Sections 2.1–2.3), thus undermining paradigm uniformity as independent force.

Another key aim of this paper is to make a major contribution to the nature of analogical processes, i.e., why specific patterns are preferred over others in

---

1. For Dutch and Swedish see Schmuck (2010), for English and its varieties, where the $[o_{PRET}=o_{PP}]$ pattern has also been extended to some degree, see Steiner (2007) and Nowak (2010).

analogical extension, and how the extension itself can be best accounted for. In Luxembourgish, the generalization of the preterital *ou* is product-oriented in the sense of Zager (1980) and Bybee (e.g., 1988), with new forms being coined on the basis of the Class II+VI preterites, irrespective of the phonological properties of the source (i.e. the infinitive; see Section 2.3). In German, product- and source-oriented generalizations are mutually reinforcing and thus highly entangled (see Section 2.2.1). The productivity of the ablaut patterns under discussion results from a complex network of interacting factors such as high type-frequency, the openness of the phonological defining properties of the pattern, shared interparadigmatic features between the analogical model and the target (e.g., the PP-*o* of Class II verbs in German) as well as the degree of cue validity of the extended ablaut vowel with respect to the [PAST] function (Sections 2.2–2.3).

2.   Two case studies

Before going into more detail concerning the cases of ablaut pattern extension in German (Section 2.2) and Luxembourgish (Section 2.3), we must distinguish between the notions of paradigm leveling and pattern extension first, and then set out the usage-based theoretical foundation on their causes (Section 2.1).

2.1   Paradigm leveling, pattern extension, and regularizations

Paradigm leveling, also intraparadigmatic or analogical leveling, is defined as the removal of allomorphic alternations within an inflectional paradigm. In verb morphology, strong-to-weak class shifts can be considered a prime example of this type of analogical change, compare English *to bake* whose strong preterite and past participle forms became weak: *bōk – baken → baked – baked*. However, strong verbs need not necessarily abolish stem allomorphy completely, thus yielding partial leveling only, e.g., *to swell*: *swelled*$_{\text{PRET}}$, but still *swollen*$_{\text{PP}}$.

    Partial leveling can either primarily relate to the formal similarity between stem allomorphs as in German *schlagen*[2] 'to hit' under (4) or to the number of stem allomorphs within a paradigm as in the case of English *to swell* above or the German examples in (5) (see Corbett, 2007, criteria 2&4; Dammel, 2008, pp. 6–9, 17).

---

2.   The Middle High German (MHG) diphthong /uo/ became monophthongized to /u:/ in New High German (NHG), the initial MHG *s-* became [ʃ] <sch> during the MHG period.

(4) *slahen – sluog – geslagen* → *schlagen – schlug – geschlagen*

(5) a. *helfen*: *half, hulfen, geholfen* → *half, geholfen*
    b. *melken*: *malk, mulken, gemolken* → *malk, gemolken* → *molk, gemolken*

In (4), formally more related stem allomorphs result from the leveling of the *h – g* alternation. In cases where the alternating stem vowels are unified, the number of stem allomorphs is automatically reduced, in the case of *helfen* 'to help' from four to three different stems, see (5a). However, some strong verbs simplified their ablaut pattern even to a greater extent, compare *melken* 'to milk' in (5b), which developed a two-vowel alternation pattern.

   As correctly pointed out by Campbell (2013, pp. 106–107), the uniformation of the PRET and PP stem vowels is another instance of partial leveling, proving that paradigm leveling is, as Werner (1987, p. 600) suggests, no "all-or-nothing-phenomenon". In this context, the regularities detected by Bittner (1996) in the strong-to-weak shift of German verbs are highly instructive, see (6):

(6)

IMP	PRES	PRET	PP
*bill>***bell!**	*bill(s)t>***bell(s)t**	*boll>***bellte**	*gebollen >***gebellt**
back!	*back(s)t>***back(s)t**	*buk>***backte**	*gebacken*
*milk>***melk!**	*milk(s)t>***melk(s)t**	**melkte**/*molk*	*gemolken*
*schier!>***scher!**	*schier(s)t>***scher(s)t**	*schor*	*geschoren*

Depending on the number of different stems, this change involves a series of stem allomorphy reducing steps; compare, for instance, the leveling of the *e/i*-raising in the 2nd PS.SG.IMP. and 2nd/3rd PS.SG.PRES.IND. respectively in *melken*. As these single changes are directly sequential, they always result in partial leveling because stem allomorphy is retained in several other paradigmatic positions; compare partially leveled strong verbs which have either retained strong preterites and past participles (*scheren* 'to shave'), strong past participles only (*backen* 'to bake'), or which oscillate somewhere between verbs with and without strong preterites (*melken* 'to milk'), see (6). Only if the last leveling step is completed (i.e. the loss of the strong PP) is stem allomorphy totally relinquished, see *bellen* 'to bark'. As will be shown in 2.2.2, the adoption of $[o_{PRET}=o_{PP}]$ can be implemented as a further regularization step in Bittner's scale.

   It has often been claimed that the phenomenon of paradigm leveling is nothing other than a type of **analogical extension** of existing patterns (see Hill, 2007; Garrett, 2008). This is indeed true for all the cases discussed above, as analogical models can easily be detected, be it the inflectional pattern of weak verbs (i.e., adding the dental suffix to the unchanged verb stem) or a concrete ablaut pattern. However, the notion of "leveling" captures the result of the analogical process more

adequately than the term "extension", since the latter may not necessarily yield a reduction of allomorphy but may also introduce stem allomorphy anew, as it has been the case for formerly weak verbs that became partially or totally strong (e.g., *dive – dove/dived – dived* or *winken – winkte – gewunken/gewinkt* 'to wave') or for strong verbs that developed a more distinctive ablaut pattern via analogical borrowing of the *ie*-preterites of Class VII in Dutch (e.g., *helpen – holp – geholpen → hielp – geholpen*; see Nowak, 2010). Whereas the notion of leveling is strongly associated with processes operating exclusively on the intraparadigmatic level, Nowak/ Schmuck (2013) have argued that leveling can also operate on the interparadigmatic level, i.e., allomorphy is reduced within the strong verb system as a whole through the emergence of patterns which are few in number but which are well-stocked. Emblematic cases of this leveling type are the emergence of Class VII in West (and North) Germanic with a uniform preterite vowel for the formerly reduplicating verbs, and – as will be shown in 2.3 – the Luxembourgish *ou*-preterite.

Finally, it is important to touch upon the **causes** of paradigm leveling and pattern extension within strong verbs that have been put forward so far in research. Morphological theories often explain the shift from ablauting, i.e. strong verbs to the weak class in the history of Germanic languages by the universal preference for uniform stem exponence and additive marking over stem changing processes. This is because the notion **regularity** is based on the "one form – one function" principle (e.g., Mayerthaler, 1981; Bittner, 1996; Corbett, 2007). As weak verbs never undergo stem changes to express morphosyntactic features (e.g., *lachen – lachte – gelacht* 'to laugh'), their inflectional behavior fulfills this Uniformity Principle (e.g., Bittner, 1996). Hence, strong verbs approach this regular pole by abolishing stem changes, as exemplified in (6) above. Additionally, this class shift benefits from language-specific systemic pressures, i.e. the Class-Preference Principle best-known from Wurzel's (1984) system-dependent approach to Natural Morphology: the most type-frequent weak verb class (c. 4,000 members in NHG) wins out at the expense of the smaller class of strong verbs (c. 150 members in NHG) in the long-run (see also Bittner, 1996; Duden, 2016, § 704). However, these accounts not only fail to explain the fact that many strong verbs have been preserved in all[3] Germanic daughter languages but also the fact that several strong verbs have only partially regularized, either by developing weak preterite forms only (e.g., Mod.Grm. *backte*, but *gebacken*) or by simplifying their ablaut pattern through [$o_{PRET}=o_{PP}$].

---

3.   However, intensive language contact may also contribute to morphological simplification, as has been the case in Afrikaans, the only Germanic language which has not preserved the strong conjugation (see Donaldson, 1993, pp. 217–299).

Dual-system approaches (e.g., Marcus et al., 1992; Pinker, 1999), which assume that weak/regular and strong/irregular verbs are processed by two distinct modules – a rule-based mechanism for regular verbs and an associative memory system for irregular verbs – provide a better approximation to these diachronic developments. Since irregular forms (including strong verbs) are retrieved from associative memory, they are expected to be sensitive to token and type frequency as well as to phonological similarity effects. This prediction is largely in line with **single-system approaches**, which claim a memory-based associative system for both regular and irregular forms (e.g., usage-based approaches, compare Bybee, 1995, 2010).

Token frequency impacts the cognitive representation and processing of linguistic structures by strengthening memory traces under repeated use of and exposure to forms (i.e. they become more entrenched in the sense of Langacker, 1987), making high-frequency items not only easier and faster to access but also more resistant to analogical, regularizing changes (conserving frequency effect). The exact opposite applies to low-frequency items and those affected by drastic frequency losses. The history of strong verbs underpins this correlation between token frequency and regularization. As Augst (1975, p. 255) has shown for German, more than 50% of its strong verbs were lost on the threshold of NHG, either by becoming weak or by dying out (MHG 339 > NHG 169 verbs). However, Augst's (1975, p. 258) analysis of strong verb losses reveals that only peripheral and thus low-frequency items of the lexicon were affected (MHG 308 > NHG 132 verbs), whereas core items hardly suffered from any losses (MHG 41 > NHG 37 verbs). Similar findings have been put forward by computational models simulating verbal class shifts for English (e.g., Hare/Elman, 1995; Lieberman et al., 2007) and Dutch (e.g., Pijpops/Beuls/Van de Velde, 2015), and by studies on first language acquisition[4] of regular and irregular (verb) morphology (e.g., Marcus et al., 1992). However, a major challenge for dual-system approaches concerns the prediction that the application of a rule is frequency-insensitive, i.e. it is not based on generalizations made across stored inflected forms (e.g., Pinker, 1999). As the Luxembourgish case in Section 2.3 will show, the regular/weak preterite formation was completely lost due to massive preterite loss, suggesting that regular forms are also stored in memory as proposed in usage-based unitary frameworks (e.g., Bybee, 1995, 2010).

Moreover, **Morphological Economy Theory** claims that the ratio behind frequency-driven regularization processes on the one hand and the conservation or

---

4.   It should be mentioned that some frameworks assume that language change relies on first language acquisition rather than on usage itself (e.g., Pinker, 1999, p. 68). However, Bybee (2010, pp. 114–119) provides convincing counterevidence to this assumption, e.g., by pointing out that children's errors/innovations not only deviate from those attested in language change but usually do not persist into adulthood.

even emergence of irregularity on the other is highly functional (see Werner, 1987; Ronneberger-Sibold, 1988; Nübling, 2000; Dammel, 2011; Nowak, 2015): In the high-frequency domain, performance-related needs are favored, i.e. shortness of expression going along with low articulatory effort, e.g., *I have done > I've done, wir haben > ham* (colloquial/spoken German 'we have'). In cases of drastically increasing token frequencies as in German *haben*, originally a fully regular verb, irregularities have been introduced. The lack of morphological transparency which goes along with irregular forms stored as whole units becomes a problem in terms of cognitive requirements if the "cost-benefit analysis" (Nübling, 2001, p. 71) is put out of balance. This is the case for irregular items of low(er) and decreasing frequencies. Here, competence-related needs in terms of low cognitive costs are favored, i.e. coding techniques conforming to the Uniformity Principle that apply in a rule-like behavior to many items. Consequently, even systemic pressures like the Class-preference Principle only apply in the low-frequency domain.

Type frequency, in turn, may on the one hand also contribute to the conservation of individual patterns and to their productivity on the other. Thus, infrequent strong verbs have remained strong if they belonged to high-type frequency classes – especially the ones exhibiting phonological consistency –, whereas low-frequency strong verbs stemming from small classes with members of phonological heterogeneity were more prone to become weak (e.g., Hare/Elman, 1995; Pijpops/Beuls/Van de Velde, 2015). Moreover, patterns of high type-frequency are cognitively highly entrenched and thus more readily available to apply to new items, especially if the phonological shape of the pattern is less or hardly specified as in the case of weak verbs (e.g., Bybee, 1995, p. 432; Köpcke, 1999, p. 57; Pijpops/Beuls/Van de Velde, 2015, p. 93). These productivity determinants are supported by the German and Luxembourgish data in Sections 2.2–2.3. However, they will be complemented by two more factors: cue validity of the ablaut vowel with respect to the grammatical category it expresses, and interparadigmatic similarity between the target and the analogical basis.

2.2   German

2.2.1   *The analogical basis of $[o_{PRET} = o_{PP}]$*
As Table 1 illustrates, the $[o_{PRET}=o_{PP}]$ pattern results from the ablaut leveling within the preterite forms of Class II, see *bieten* 'to offer': Here, the preterite singular stem *bōt(-)* (boldface) was analogically extended to the plural (with the older form listed in parentheses), probably under the influence of the *o* of the PP (see Nübling, 1999; type-frequency counts based on Solms, 1984, § 95).

Table 1. Preterite number ablaut leveling in Early New High German

Ablaut series	1. INF.	2. SG.PRET.	3. PL.PRET,	4. PP	Type frequency
I	*rīt-en* 'to ride'	*rit (reit)*	**rit-en**	*ge-rit-en*	90
II	*biet-en* 'to offer'	**bōt**	*bōt-en (but-en)*	*ge-bot-en*	54
III a	*bind-en* 'to bind'	**bant**	*band-en (bund-en)*	*ge-bund-en*	51
b	*werf-en* 'to throw'	**warf**	*warf-en (wurf-en)*	*ge-worf-en*	47
IV	*stel-en* 'to steal'	*stāl. (stal)*	**stāl-en**	*ge-stol-en*	30
V	*geb-en* 'to give'	*gāp (gap)*	**gāb-en**	*ge-geb-en*	32
VI	*fur-en* 'to drive'	*fuor, -en*		*ge-far-en*	27
VII	*siāf-en* 'to sleep'	*slief, -en*		*ge-slāf-en*	44

In Early New High German (ENHG, c. 1350–1650), the pattern of Class II was analogically extended to 24 strong verbs of the Classes III–VI, see Table 2, which arranges the new members of the $[o_{\text{PRET}}=o_{\text{PP}}]$ pattern according to their historical class (adapted from Nowak, 2015, p. 12).

Table 2. The new members of $[o_{\text{PRET}} = o_{\text{PP}}]$ pattern in Early New High German

Class	Verbs	Total
IIIa	*glimmen* 'to glow', *klimmen* 'to climb'	2
IIIb	*melken* 'to milk', *quellen* 'to well', *(er)schallen* 'to resound' (< *schellen*), *schmelzen* 'to melt', *schwellen* 'to swell'; now weak: *bellen* 'to bark', *(ver)wirren* 'to disarray' (< *werren*)	7
IV	*dreschen* 'to thresh', *fechten* 'to fence', *flechten* 'to braid', *löschen* 'to extinguish', *scheren* 'to shear sth.', *schwüren* 'to fester', now weak: *rächen* 'to revenge', *verhehlen* 'to disselmble'	8
V	*gären* 'to ferment', *pflegen* 'to cultivate sth.', *weben* 'to weave', *bewegen* 'to talk so. into sth.', *wägen* 'to weigh sth.'	5
VI	*heben* 'to heave', *schwören* 'to swear' (< *swern*)	2

The fact that Class II served as an analogical model is, at least at first sight, hardly surprising: With a total of c. 50 members, it is, along with series I (c. 100 verbs) and IIIa (c. 50 verbs), one of the three biggest strong verbs classes in MHG/ENHG, thus fulfilling one of the major criteria for pattern productivity already introduced in Section 2.1: high type-frequency (see also Table 1). The type-frequency/productivity correlation for ablaut patterns is supported by diachronic data from several Germanic languages; compare, for instance, the extension of the *cling*/*clung*-pattern in English (e.g., Bybee/Moder, 1983), of the Class I–III patterns in Swedish (see Schmuck, 2010), and also of the Class I and IIIa patterns in German (e.g., Köpcke,

1999, p. 56; for Luxembourgish see 2.3). Whereas in the cases of Classes I and IIIa in German, the new members matched the phonological structure of their analogical bases (compare formerly weak verbs like *pfeifen* 'to whistle' and *winken* 'to wave' with strong verbs like *greifen* 'to grip', Cl. I, and *sinken* 'to sink', Cl. IIIa), this was definitely not the case with the verbs that joined the $[o_{\mathrm{PRET}}=o_{\mathrm{PP}}]$ pattern; compare *melken* 'to milk' and *heben* 'to heave' with Class II verbs like *fliegen* 'to fly' or *bieten* 'to offer' (see also Tables 1 and 2). They neither shared the stem vowel *ie* of Class II (MHG /ie/ > NHG /iː/) nor – at least in many cases – the phonotactic property of a single stem final consonant.

However, on closer inspection of Table 1, most strong verbs that joined the $[o_{\mathrm{PRET}}=o_{\mathrm{PP}}]$ pattern shared the *o* of the PP with Class II. It is mainly this interparadigmatic similarity that serves as gateway for the analogical extension (for Dutch see Nowak, 2015, pp. 260–264). This explains the quantitative distribution of the $[o_{\mathrm{PRET}}=o_{\mathrm{PP}}]$ pattern across Classes III–VI: More than half of the new members stem from Classes III and IV where the PP-*o* was inherited or – in very few cases – introduced by sound change (i.e., the lowering of *u* > *o* before historical double nasals, e.g., MHG *geglummen*, *geklummen* > (E)NHG *geglummen*, *geklommen* 'to glow' and 'to climb'). Only a handful of verbs with a PP-vowel other than *o* joined the Class II pattern: *weben* – *geweben* 'to weave' (Cl. V) and *heben* – *gehaben* 'to heave' (Cl. VI) (also *bewegen*, *wägen*, *gären*, *pflegen*, *schwören* in Table 2). Interestingly, these verbs accommodated to the $[o_{\mathrm{PRET}}=o_{\mathrm{PP}}]$ pattern by acquiring the *o* in the PP first, and, once the PP-*o* had been consolidated, the *o* of the PP in the preterite forms (see Nowak, 2015, pp. 253–255), see (7).

(7) PP *before* PRET
ENHG *weben*   PP *geweben*   PRET *wab*
     (1) → *gewoben*   (2) → *wob*
ENHG *heben*   PP *gehaben*   PRET *hub*
     (1) → *gehoben*   (2) → *hob*

Again, the direction of analogical extension in (7) can be easily accounted for by type frequency: PPs with *o* as a stem vowel are by far the most type-frequent pattern within strong verbs, comprising c. 40% of all cases in ENHG (Cl. III–IV in Table 1; see Nowak, 2015, p. 253). Hence, Class V and VI verbs rebuilt their PP on the basis of the most type-frequent PP-pattern.

This interparadigmatic analogy is additionally supported by a source-oriented generalization: The verbs of Classes III–IV that adopted $[o_{\mathrm{PRET}}=o_{\mathrm{PP}}]$ analogically almost exclusively exhibited infinitival *e*, thus establishing a strong basic-derived relationship between the *e*-infinitive and the *o* of the PRET/PP (see also Table 2). Consequently, this source-oriented relationship could be easily applied to the verbs of Classes V and VI, as they also had infinitival *e* (e.g., *weben*, *heben*).

It seems plausible that the extension of the $[o_{\text{PRET}}=o_{\text{PP}}]$ pattern can be described in terms of a source-oriented generalization within the functional domain [PAST], as already suggested by Hempen (1988, pp. 273–274), i.e., the Class II pattern shows that a PP-*o* corresponds to an *o* in the PRET (see (8a)), a relationship that is applied to verbs with an inherited or analogical *o* in the PP (see (8b)).

(8) a. Cl. II: *geboten* : *bot*
    b. Cl. IIIb: *gemolken* : X   => X=*molk*
       Cl. IV: *geschoren* : X   => X=*schor*

The generalization exemplified in (8) was made possible by the fact that, in ENHG, the $[o_{\text{PRET}}=o_{\text{PP}}]$ pattern of Class II was not so strongly associated with the infinitival stem vowel as in other ablaut classes: Whereas the latter was well-defined in Classes I and III–VI, allowing for an analogical extension of their patterns to verbs of identical phonological shape only (i.e. the combination of stem vowel+stem final consonant(s)), Class II verbs exhibited /u:/ (e.g., *sûgen* 'to suckle'), /y:/ (e.g., *liugen* 'to lie'), and /ou/ (e.g., *soufen* 'to swig') alongside prototypical /iə/ (e.g., *fliegen* 'to fly'; see Nowak, 2015, pp. 233–234). Thus, the openness of the infinitival stem vowel allowed for a product-oriented generalization in the sense of Zager (1980) and Bybee (1995, p. 443). All in all, the product-oriented and source-oriented generalizations are strongly inter-linked and mutually reinforcing operating principles.

Finally, the high cue validity of the stem vowel *o* in terms of monofunctionality, more precisely with respect to the [PAST] function should be noted: Apart from *stoßen* 'to push' and *kommen* 'to come', the *o* appears exclusively in the derived past tense values of strong verbs. This factor combined with high type-frequency, interparadigmatic similarity, the openness and lower complexity of the Class II pattern sets the course for its analogical extension.

2.2.2 *The pattern $[o_{\text{PRET}} = o_{\text{PP}}]$ as a partial regularization strategy*

The key to understanding the ratio behind the analogical extension of the $[o_{\text{PRET}}=o_{\text{PP}}]$ pattern in German lies in the functional and usage-based perspective to verbal change as outlined in Section 2.1.

As already indicated in the introduction, the adoption of $[o_{\text{PRET}}=o_{\text{PP}}]$ is a prime example of partial leveling as the vast majority of the new members would have regularly developed the more complex three-vowel alternation structure ABC (20 out of 24, see also Table 2). The fact that the analogy yielded a uniform exponence, the *o*, for the past tense values and at the same time an iconic formal distinction between past as opposed to the present is a welcome side-effect, especially when considering the fact that the preterite and perfect largely functionally converged in German. Only in the case of the few ABA-verbs lacking a salient present/past

opposition was stem allomorphy not reduced (e.g., *bewegen*, *gären*, *wägen*, and *weben*, see also Table 2), e.g., *weben – wab – geweben → weben || wob – gewoben*.

More importantly, the analogical adoption of $[o_{PRET}=o_{PP}]$ can be easily implemented as a further regularization step in Bittner's (1996) scale of strong-to-weak class changes introduced in Section 2.1, see Figure 1:
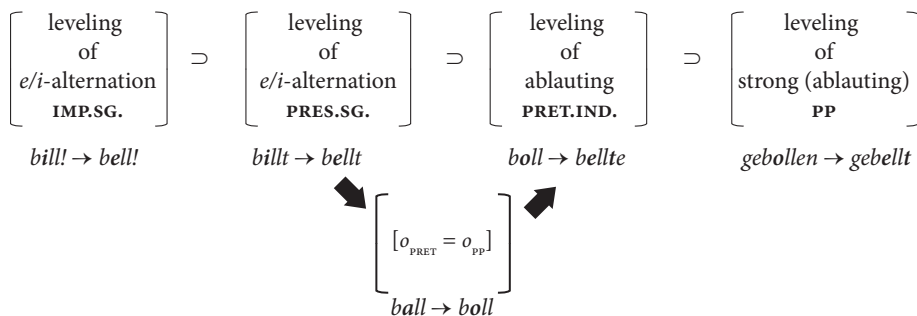
Figure 1. Implementation of Class VIII in Bittner's implicational scale

Verbs with the two left-most strong class features like *bellen* 'to bark' got rid of the *e/i*-alternation in the imperative and the present tense forms before simplifying their pattern to $[o_{PRET}=o_{PP}]$. This change is then followed by the loss of the (new) strong preterites and finally of the strong PP. Whereas the adoption of $[o_{PRET}=o_{PP}]$ has been an intermediate step for some strong verbs on their way to the weak class (e.g., *rächen*, *verhehlen*, *pflegen*), the great majority have remained strong to this day (see Nowak, 2015, p. 199; see also Table 2). However, it should be kept in mind that the accommodation to $[o_{PRET}=o_{PP}]$ is a facultative step in the regularization process, whereas Bittner's steps are obligatory.

The pressing question is now how to account for the fact that other strong verbs from Classes III–VI did not simplify their inherited pattern, compare, for example, *helfen* 'to help' (Cl. IIIb) and *geben* 'to give' (Cl. V), whose inflectional behavior was identical to that of verbs like *melken* 'to milk' (Cl. IIIb) or *weben* 'to weave' (Cl. V) in ENHG, compare (9)–(10).

(9)   a. *helfen*   *half – geholfen*   → no change
      b. *melken*   *malk – gemolken*   → simplification: *molk – gemolken*

(10)   a. *geben*   *gab – gegeben*   → no change
       b. *weben*   *wab – geweben*   → simplification: *wob – gewoben*

The selectivity of the $[o_{PRET}=o_{PP}]$ extension can be easily explained in terms of frequency; compare the frequency counts for MHG (Figure 2a) and for NHG (Figure 2b), respectively, which contrast the average token frequencies per million words (pmw)

of the 24 verbs that adopted the $[o_{\text{PRET}}=o_{\text{PP}}]$ pattern with their analogy-resistant counterparts of Classes III–VI for both periods of German (for details see Nowak, 2015, pp. 178–179). As Figure 2a shows, the verbs that simplified their pattern to $[o_{\text{PRET}}=o_{\text{PP}}]$ were of much lower frequency than the ones that preserved their inherited pattern: 74 tokens pmw vs. 503 tokens pmw in MHG. More importantly, the analogy-resistant items did not suffer from such drastic frequency losses on the threshold to NHG as their low-frequency counterparts, see Figure 2b: The average frequency of the latter fell down to only 5 tokens pmw, whereas the analogy-resistant cases are ten times as frequent (148 tokens pmw).
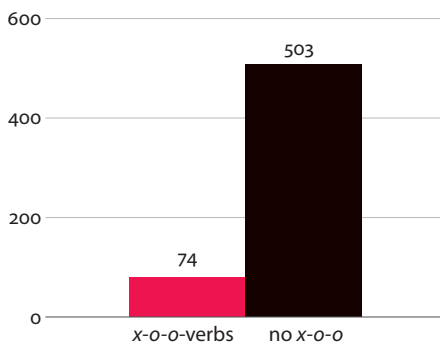
a. Middle High German
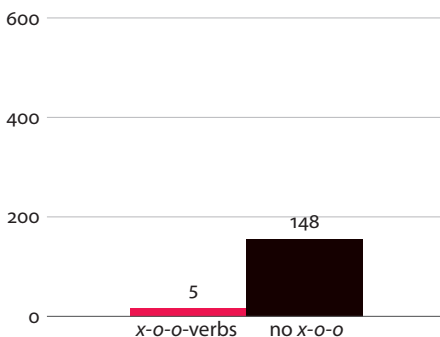
b. New High German

Figure 2. Average token frequencies of $[o_{\text{PRET}}=o_{\text{PP}}]$ verbs and their analogy-resistant counterparts in MHG and NHG (< Classes III–VI)

Interestingly, the analogical extension of $[o_{\text{PRET}}=o_{\text{PP}}]$ is still ongoing in contemporary German, catching up with less frequent ABC-verbs with a PP-*o* like *schwimmen* 'to swim', *spinnen* 'to spin', *sinnen* 'to reflect', and *rinnen* 'to flow' (see Nowak, 2015,

pp. 72–81), see Table 3b: They exhibit an average frequency of 8 tokens pmw each, whereas the two very frequently used verbs *gewinnen* and *beginnen* (c. 300 tokens pmw each, see Table 3c) have no leveling tendencies at all. In ᴇɴʜɢ, however, verbs like *schwimmen* were obviously still frequent enough to resist the analogy compared to their extremely infrequent counterparts *glimmen* 'to glow' and *klimmen* 'to climb', which exhibited an average frequency of 2 tokens pmw each (see Table 3a).

Table 3. Average token frequencies pmw of *beginnen*, *gewinnen* compared to *rinnen*, *schwimmen*, *sinnen*, *spinnen*, and *glimmen*, *klimmen*

	$[o_{PRET}=o_{PP}]$ tendency	MHG	NHG	VERBS
(a)	adoption in ᴇɴʜɢ	2	2	*glimmen* 'to glow', *klimmen* 'to climb'
(b)	in ɴʜɢ	68	8	*rinnen* 'to flow', *schwimmen* 'to swim', *sinnen* 'to reflect', *spinnen* 'to spin'
(c)	no tendency	956	308	*beginnen* 'to begin', *gewinnen* 'to win'

This correlation between token frequency and ablaut pattern complexity also holds true for the strong verb class as a whole (see Nowak, 2015, p. 167 for German, p. 172 for Dutch): Table 4 shows the average token frequencies per million words for the three structural patterns of ɴʜɢ strong verbs, which were extracted from the corpus "Projekt Wortschatz deutsch" for written Modern German (size: c. 217 million words, for details see Nowak, 2015, p. 19): ABB (e.g., *lügen – log – gelogen* 'to lie'), ABC (e.g., *sitzen – saß – gesessen* 'to sit'), and ABA (e.g., *lassen – ließ – gelassen* 'to let'). The average frequencies of verbs with the more distinctive three-vowel alternation ABC are more than twice as high as those of the less complex ABB-type, c. 100 vs. 40 tokens pmw, suggesting two conclusions: First, less frequent verbs have remained strong precisely because they exhibit a less complex ablaut pattern ABB. Second, only high-frequency strong verbs can "afford" the complex three-vowel alternation ABC-pattern.

Table 4. Average token frequencies of strong verb patterns [only simplexes]

	ablaut alternation types		
	ABB	**ABC**	**ABA**
tokens	41	97	143
types	82	49	32
e.g.,	*lügen* 'to lie', *reiten* 'to ride'	*sitzen* 'to sit', *helfen* 'to help'	*lassen* 'to let', *geben* 'to give'

The fact that ABA-verbs average the highest frequencies (c. 140 pmw) can be accounted for by their higher vulnerability to become weak, once their strong preterites have been regularized, compare *backen – (buk) – gebacken > gebackt*

'to bake' (e.g., Dammel, 2011, pp. 196–199). Furthermore, the vowel identity between the stems of the PRES and PP yields no lower cognitive costs because these category values are functionally distinct. Hence, both stems need separate lexical representations, whereas this is not the case for ABB-verbs, as the preterite and perfect (with the PP as part of the construction) have largely functionally converged (see Dammel/Nowak/Schmuck, 2010). It is therefore anything but surprising that at least some low-frequency ABA-verbs adopted the $[o_{\text{PRET}}=o_{\text{PP}}]$ pattern.

2.3   Luxembourgish

The history of Luxembourgish is closely intertwined with that of High German and particularly with Moselle-Franconian dialects of German. It was only emancipated from the latter in the course of the 19th century (see Nowak, 2015, p. 6). As a consequence of its "close linguistic-genealogical relationship with German" (Gilles et al. 2011, p. 65), MHG[5] will serve as a reference point for the description of the Luxembourgish' developments.

As already pointed out in the introduction, Luxembourgish developed a uniform ablaut vowel *-ou-* within the strong preterite forms. However, as Luxembourgish has been subject to massive preterite loss, only a handful of preterites still bear witness to this analogy, see (11), which arranges the verbs according to their historical class (adapted from Nowak, 2015, pp. 121–122):

(11)	Class	INF	pret	PP	
	I	bleiwen	blouf	bliwwen	'to stay'
	II	zéien	zouch	gezunn	'to pull'
	IV	kommen	koum	komm	'to come'
	V	gesinn	gesouch	gesinn	'to see'
		ginn	gouf	ginn	'to give'
		leien	louch	geleeën	'to lie'
		sëtzen	souz	gesiess	'to sit'
	VI	stoen	sto(u)ng	gestan(en)	'to stand'
	VII	fänken	fo(u)ng	gefaang(en)	'to catch'
		goen	go(u)ng	gaang	'to go'
		hänken	ho(u)ng	hé(i)ng	'to hang'

Strikingly, the analogical extension of preterital *-ou-* was not restricted to specific ablaut classes as in German, suggesting that the analogical extension once

5.   It goes without saying that the MHG system presented here is an abstraction over the wide range of MHG varieties (e.g., dialects, sociolects), i.e. it is oriented to normalized MHG as found in reference grammars.

consequently applied to all strong verbs (at least to those that had not yet lost their preterites due to preterite loss). This assumption is confirmed by two facts: first, by the compilation of archaic preterite forms based on the Luxembourgish Dictionary (LWB, 1954–77, vol. I–V) which documents at least twenty more *ou*-preterite cases, including athematic *doen/dinn* 'to do' and even historically weak verbs like *froen* 'to ask' or *maachen* 'to make'; compare the examples in (12). Second, by dialectal data: the Ösling dialect located at the northern point of Luxembourg and, hence, north of the preterite/perfect-isogloss has preserved approximately 100 cases of uniform preterites (mainly from historical strong verbs; for a full list see Nowak, 2015, pp. 126–138).

(12)

Class	INF	PRET	PP	
II	*krauchen*	*krouch*	*gekroch(t)*	'to crawl'
III	*klammen*	*kloum,*	*geklommen*	'to climb'
	kléngen	*kloung*	*geklongen*	'to cling'
IV	*stiechen*	*stouch*	*gestach*	'to sting'
V	*geschéien*	*geschouch*	*geschitt*	'to happen'
	friessen	*frouss*	*gefriess*	'to devour'
	iessen	*ouss*	*giess*	'to eat'
VI	*droen*	*drouch*	*gedroen*	'to carry'
VII	*biosen*	*blous*	*geblos(en)*	'to blow'
	jalen	*foul*	*gefall*	'to fall'
	halen	*houl*	*gehalen*	'to hold'
	heeschen	*housch*	*heeschen, geheescht*	'to be named'
	lafen	*louf*	*gelaf*	'to walk'
	loossen	*louss*	*gelooss*	'to let'
athematic	*doen/dinn*	*doung*	*gedinn*	'to do'
weak	*froen*	*frouch*	*gefrot*	'to ask'
	kafen	*kouf*	*kaf kaafl, kat*	'to buy'
	kréien	*kr(o)uch*	*kritt*	'to get'
	maachen	*mouch*	*gemaach(t), gemeet*	'to make'

The emergence of a uniform preterite marker has to be seen in the context of the extreme preterite loss Luxembourgish was subjected to. Besides lexical token frequency, category frequency is likewise decisive in determining the extent to which allomorphy is manifested in the paradigm: the more frequent a category bundle (e.g., 3PS.PRES.IND.) or a category value (e.g., PRES.), the greater the degree of fusion of forms, the amount of allomorphy, and the resistance to regularizations, and vice versa (see Nübling, 2000, pp. 217–273; Dammel/Nübling, 2006). The immense decrease in category frequency, in our case of the preterite, triggers the leveling of the different preterite vowels in favor of a uniform marker *-ou-*, thus reducing

memorizing costs. Again, uniformity is only a means to an end, but not the driving force behind this process. Following Dammel/Nübling (2006, pp. 109–110), I claim that the emergence of uniformity must be interpreted as "a symptom of category weakening", in our case, of the preterite. As argued in Dammel (2011, p. 192) and Nowak (2015, p. 211), this type of simplification can also be interpreted as a partial regularization step, in that it is an intermediate step towards preterite loss.

Category frequency also interacts with lexical frequency, i.e. preterite forms of low-frequency verbs are more vulnerable to preterite loss than those of high-frequency verbs. This tendency explains why the remnants of the analogical *ou*-preterite belong to the high-frequency domain, whereas in German, only less frequent strong verbs relied on the [o_{PRET}=o_{PP}] analogy. The interaction between category and lexical frequencies can be best illustrated if verbs of different lexical frequency domains are compared with one another; see Table 5, which shows the average absolute frequencies crawled from Google.lu in 2012 for randomly selected verbs with no preterite forms (e.g., *bäissen* 'to bite'), the verbs with *ou*-preterites, and the two suppletive verbs *sinn* 'to be' and *hunn* 'to have' (see Nowak, 2015, pp. 208–209): Verbs that preserved their *ou*-preterite are twice as frequent as their preterite-less counterparts, c. 300,000 vs. 150,000 tokens, whereas the extremely frequent cases *sinn* and *hunn* (c. 1.2 million tokens) did not merely preserve the synthetic preterites but also resisted the analogy, thus exhibiting suppletive forms (*war* and *hat*).

Table 5. Average absolute tokens frequencies of verbs with no preterite forms compared to verbs with *ou*-preterites and suppletive forms

	No pret.	*ou*-pret.	Suppl. pret.
absolute token freq.	c. 150,000	c. 300,000	c. 1,200,000
types	*bäissen, bannen, bedréien, briechen, drénken, fléissen, friessen, fueren, hellefen, hiewen*	*bleiwen, fänken, gesinn, ginn, goen, hänken leien, setzen, stoen, zéien*	*sinn – war* *hunn – hat*t

The Luxembourgish case also suggests that the regular/weak preterite formation was based on generalizations made across stored inflected forms with a high sensitivity to frequency of use rather than on a frequency-insensitive rule as proposed in dual-system approaches (see also Section 2.1). If the latter were the case, the weak pattern should have prevailed instead of being completely lost due to decreasing category frequencies, i.e. preterite loss (for an intensive discussion see also Nowak, 2015, pp. 296–345).

What has not yet been addressed is how the analogical extension of the preterital *ou* was constrained. Here, Classes II and VI, which merged in their preterite forms,

served as an analogical model (for details see Nowak, 2015, pp. 128–136). Thus, a large, strong class of c. 92 verbs emerged – an important prerequisite for productivity. However, high type-frequency alone is not sufficient to explain how and why the *ou* must have been reinterpreted as preterite marker in order to be analogically extended to any strong verb. To understand this reinterpretation process, it is first necessary to consider the development of ablaut relations between the INF/PRES and PP in Classes II and VI. These developments are illustrated in Table 6 for MHG and LUX (see Nowak, 2015, pp. 284–285). As for MHG, we have clear patterns in both classes: an /ie/ in the infinitive corresponds to an /o/ in the PP (Cl. II), and an /a/ in the infinitive to an /a/ in the PP (Cl. VI). However, Luxembourgish underwent massive phonological changes, which destroyed the clear ablaut relations of MHG. As Table 6 shows, the MHG /ie/-/o/ pattern of Class II disintegrated into seven different patterns, that of Class VI (i.e., /a/-/a/) into six. Consequently, the number of patterns increased sixfold, from MHG two to twelve alternations in LUX (the /ue-ue/ pattern emerged in two ways, from /ie-o/ and /a-a/, see *sueden* 'to seethe' and *fueren* 'to drive').

Table 6. Disintegration of ablaut relations in Luxembourgish (Cl. II and VI)

MHG class	MHG inf. vowel	LUX inf. vowel	LUX PP vowel	MHG PP vowel	Lux. example	Translation
II		ei	u		*fiélen – geflunn*	'to fly'
		ei	o		*fléissen – gefloss*	'to flow'
		ei	uə		*fréieren – gefruer*	'to freeze'
	ie	i	o	o	*riehen – geroch*	'to smell'
		i	uə		*bidden – gebueden*	'to offer'
		uə	uə		*sueden – gesueden*	'to boil'
		au	o		*krauchen – gekroch*	'to crwal'
VI		iə	uə		*hiewen – gehuewen*	'to heave'
		a:	a:		*baken – gebak(en)*	'to bake'
	a	ε	ε	a	*wäschen – gewäsch*	'to wash'
		o:	o:		*droen – gedroen*	'to carry'
		uə	uə		*fueren – gefuer*	'to drive'
		o	a:		*stoën – gestaan*	'to stand'

As a consequence of the fragmentation of the ablaut relations, the *ou*-preterite of Classes II+VI combined with (almost) any vowel that is attested for strong verbs in the infinitive (eight out of eleven different vowels) and the PP (five out of nine different vowels), see Table 7 (see Nowak, 2015, pp. 285–286).

    This wide range of possible combinations of an INF/PP vowel with a preterital *ou* contributed to the decoupling of the *ou* from its historical source, and as a consequence of that, to the reanalysis of *ou* as preterital marker on its own (see Durrell, 2001, pp. 10–11). This process additionally benefitted from the high cue validity

Table 7. Dissociation of the *ou*-preterite from the infinitive and past participle

nr.	infinitive vowel	preterite vowel	past participle vowel	nr.
1	a:			
2	ɛ		a:	1
3	i		ɛ	2
4	o, o:		o, o:	3
5	au	*ou*	u	4
6	ei		uə	5
7	iə			
8	uə			

of *ou* with respect to the category value [PAST], as it almost never appears either in the INF/PRES nor in the PP of strong verbs (only in 1% and 3%, respectively, of the cases; see Nowak, 2015, p. 243). Hence, the analogical extension of preterital *ou* represents an emblematic case of product-oriented generalization, rendering phonological similarities between the analogical basis and the target totally obsolete.

3.   Crux of the matter

The usage-based perspective gives new insights into paradigm leveling and the ratio behind ablaut pattern extensions. The present study on ablaut pattern extension in German and Luxembourgish offers three major contributions to the explanation of such changes. First, the key determinant is frequency of use, either lexical as in German and/or categorical as in Luxembourgish. Second, pattern extensions are by no means solely a matter of phonological similarity but are, rather, functionally motivated, i.e. frequency-driven partial regularization strategies within strong verbs. In German, low-frequency verbs simplified their three-vowel alternations through the adoption of the two-vowel alternation $[o_{\text{PRET}}=o_{\text{PP}}]$ of Class II, a regularization process which is an intermediate step towards the weak conjugation. In Luxembourgish, in turn, a uniform preterite *ou* emerged for strong verbs as a result of massive preterite loss, thus drastically reducing preterite formation to one single type. Hence, these different directions of leveling (intraparadigmatic vs. interparadigmatic simplification) as well as the varying degrees with which the analogical extension applied (24 strong verbs in German vs. all strong verbs in Luxembourgish) can be easily accounted for. Third, the productivity of ablaut

patterns is not only related to high type-frequency, but also to their openness and to interparadigmatic similarities between the target and the source (i.e. the analogical basis), thus allowing for output- as well as source-oriented generalizations, and finally, to the cue validity of the ablaut vowel that is extended.

Abbreviations

ENHG	Early High New German	NHG	New High German
IMP	imperative	PL	plural
IND	indicative	PP	past participle
INF	infinitive	PRS	present
LUX	Luxembourgish	PRET	preterite
MHG	Middle High German	SG	singular
Mod.Grm	Modern German		

References

Anderwald, L. (2009). *The morphology of English dialects: Verb-formation in non-standard English*. Cambridge/New York. https://doi.org/10.1017/CBO9780511576539

Augst, G. (1975). Wie stark sind die starken Verben? Überlegungen zur Subklassifikation der nhd. Verben. In G. Augst (Ed.), *Untersuchungen zum Morpheminventar der deutschen Gegenwartssprache* (pp. 231–281). Tübingen.

Bittner, A. (1996). *Starke ,schwache' Verben und schwache ,starke' Verben. Deutsche Verbflexion und Natürlichkeit*. Tübingen.

Bybee, J. L. (1995). Regular morphology and the lexicon. *Language and cognitive processes*, 10(5), 425–455. https://doi.org/10.1080/01690969508407111

Bybee, J. L. (2010). *Language, Usage and Cognition*. Cambridge. https://doi.org/10.1017/CBO9780511750526

Bybee, J. L. & Moder, C. L. (1983). Morphological classes as natural categories. *Language*, 59(1), 251–270. https://doi.org/10.2307/413574

Campbell, L. (2013). *Historical Linguistics. An Introduction*. Edinburgh.

Corbett, G. (2007). Canonical typology, suppletion and possible words. *Language*, 83, 8–42. https://doi.org/10.1353/lan.2007.0006

Dammel, A. (2008). Flexionsmorphologische Irregularität – ein graduelles Phänomen. Doch wie lassen sich Grade bestimmten? In C. Stroh & A. Urdze (Eds.), *Morphologische Irregularität. Neue Ansätze, Sichtweisen und Daten* (pp. 1–28). Bochum.

Dammel, A. (2011). *Konjugationsklassenwandel. Prinzipien des Ab-, Um-, Aus- und Aufbaus verbalflexivischer Allomorphie in germanischen Sprachen*. Berlin/New York.

Dammel, A. & Nowak, J. & Schmuck, M. (2010). Strong verb paradigm levelling in four Germanic languages. A category frequency approach. *JGL*, 22(4), 337–359. https://doi.org/10.1017/S1470542710000097

Dammel, A. & Nübling, D. (2006). The Superstable Marker as an Indicator of Categorial Weakness? *Folia Linguistica*, 60(1–2), 97–113. https://doi.org/10.1515/flin.40.1part2.97

Donaldson, B. C. (1993). *A Grammar of Afrikaans*. Berlin/New York. https://doi.org/10.1515/9783110863154

Duden. (92016). *Die Grammatik. Duden Band 4*. Mannheim et al.

Durrell, M. (1999). Zum Ausgleich der Ablautalternanzen im Niederdeutschen. *ZDL*, 105, 39–49.

Durrell, M. (2001). Strong verb Ablaut in the West Germanic languages. In S. Watts et al. (Eds.), *Zur Verbmorphologie germanischer Sprachen* (pp. 5–18). Tübingen. https://doi.org/10.1515/9783110918656.5

Garrett, A. (2008). Paradigmatic uniformity and markedness: Historical convergence and universal grammar. In J. Good (Ed.), *Explaining linguistic universals* (pp. 125–143). Oxford. https://doi.org/10.1093/acprof:oso/9780199298495.003.0006

Gilles, P., Seela, S., Sieburg, H. & Wagner, M. (2011): Languages and Identities. In IPSE – Identités, Politiques, Sociétés, Espaces (Ed.), *Doing Identity in Luxembourg. Subjective Appropriations – Institutional Attributions – Socio-Cultural Milieus* (pp. 65–106). Bielefeld.

Hare, M. & Elman, J. (1995). Learning and morphological change. *Cognition*, 56(1), 61–98. https://doi.org/10.1016/0010-0277(94)00655-5

Hempen, U. (1998). *Die starken Verben im Deutschen und Niederländischen. Diachrone Morphologie*. Tübingen.

Hill, E. (2007). Proportionale Analogie, paradigmatischer Ausgleich und Formerweiterung: ein Beitrag zur Typologie des morphologischen Wandels. *Diachronie*, 24, 81–118. https://doi.org/10.1075/dia.24.1.05hil

Köpcke, K.-M. (1999). Prototypisch starke und schwache Verben der deutschen Gegenwartssprache. In M. Butt & N. Fuhrhop (Eds.), *Variation und Stabilität in der Wortstruktur* (pp. 45–60). Hildesheim.

Langacker, R. (1987). *Foundations of cognitive grammar: theoretical preriquisites*. Stanford.

Lieberman, E. et al. (2007). Quantifying the evolutionary dynamics of language. *Nature*, 449, 713–716. https://doi.org/10.1038/nature06137

LWB (1950–1977) = *Luxemburger Wörterbuch*, Volumes I-V. Luxemburg.

Marcus, G. F. et al. (1992). *Overregularization in language acquisition. Monographs of the Society for Research in Child Development*, 57(4), 1–178.

Mayerthaler, W. (1981). *Morphologische Natürlichkeit*. Wiesbaden.

Nowak, J. (2010). On the emergence of an "8th ablaut class" in German and Dutch. *JGL*, 22(4), 361–380. https://doi.org/10.1017/S1470542710000103

Nowak, J. (2015). *Zur Legitimation einer 8. Ablautreihe. Eine kontrastive Analyse zu ihrer Entstehung im Deutschen, Niederländischen und Luxemburgischen*. Hildesheim.

Nowak, J. & Schmuck, M. (2013). Ablautausgleich in den germanischen Sprachen. In F. Gruzca & J. Zhu (Eds.), *Akten des XII. Internationalen Germanistenkongresses Warschau 2010, Vol.* 15 (pp. 21–26). Frankfurt a.M.

Nübling, D. (1999). *Wie die Alten sungen*… Zur Rolle von Frequenz und Allomorphie beim präteritalen Numerusausgleich im Frühneuhochdeutschen. *ZS*, 17(2), 185–203.

Nübling, D. (2000). *Prinzipien der Irregularisierung*. Tübingen. https://doi.org/10.1515/9783110915082

Nübling, D. (2001). The Development of "junk". Irregularization Strategies of HAVE and SAY in the Germanic Languages. *Yearbook of Morphology 1999*, 53–74. https://doi.org/10.1007/978-94-017-3722-7_3

Pijpops, D., Beuls, K. & Van de Velde, F. (2015). The rise of the verbal weak inflection in Germanic. An agent-based model. *Computational Linguistics in the Netherlands Journal*, 5, 81–102.

Pinker, S. (1999). *Words and rules. The ingredients of language*. New York.

Ronneberger-Sibold, E. (1988). Entstehung von Suppletion und Natürliche Morphologie. *ZPSK*, 41(4), 453–462.

Schmuck, M. (2010). Relevanzgesteuerter verbalmorphologischer Wandel im Deutschen und Schwedischen. In Dammel, A./Kürschner, S./Nübling, D. (Eds.), *Kontrastive Germanistische Linguistik* (pp. 523–552). Hildesheim et al.

Solms, H.-J. (1984). *Die morphologischen Veränderungen der Stammvokale der starken Verben im Frühneuhochdeutschen. Untersucht an Texten des 14.-18. Jahrhunderts*. Bonn.

Steiner, J. (2007): Regel – Schema – Lexikon. Untersuchungen zum Stark-Schwach-Kontinuum im Deutschen und Englischen. Master's thesis, Mainz University.

Werner, O. (1987). The aim of morphological change is a good mixture – not a uniform language type. In G. A. Ramat et al. (Eds.), *Papers from the 7th international Conference of Historical Linguistics* (pp. 591–616). Amsterdam. https://doi.org/10.1075/cilt.48.43wer

Wurzel, W. U. (1984). *Flexionsmorphologie und Natürlichkeit*. Berlin.

Zager, D. (1980). *A real time process model of morphological change*. PhD, State University of New York.

Remotivating inflectional classes

An unexpected effect of grammaticalization

Livio Gaeta

Università di Torino

In Titsch, a Walser German dialect spoken in the Aosta Valley, a remodeling of
the old strong/weak verb classes found in the rest of the Germanic languages
has taken place, whereby verbs belonging to the strong class turn out to display
both strong and weak past participles. This outcome results from the reuse
or exaptation of the original morphological differentiation based on a purely
lexically-governed distribution which has been remotivated by associating the
morphological features with specific constructional schemas resulting from sev-
eral processes of grammaticalization.

Keywords: morphology, verb, inflection, grammaticalization, exaptation

1. Introduction

The Walser German varieties, found at the South-Western edge of the German
dialectal continuum, belong to the so-called Highest Alemannic branch and are
fairly well known for their extreme conservatism (cf. Russ, 1990 for an overview).
On the other hand, due to intense contact with the Romance languages, they also
present highly interesting innovative features. This is especially true of those vari-
eties which still survive as linguistic islands in Northern Italy, even though they are
subject to a process of language decay which has become dramatically fast-paced in
the last decades (cf. Del Negro, 2004 for details and discussion). In this paper, data
collected in the village of Gressoney in the Aosta Valley will be discussed, as part
of the ongoing research project ArchiWals, an initiative which aims to safeguard
the cultural and linguistic heritage of these endangered minorities (cf. Angster
et al., 2017 and the website www.archiwals.org). This variety has been subject to
intense contact with Piedmontese, Francoprovençal, French, and Italian. Especially
after World War I during the Fascist Regime, the local Walser German variety
called Titsch was expressly forbidden; moreover, after World War II the impact of

Standard Italian has increased dramatically also because of strong processes of in- and out-migration with the consequent mixed marriages. At any rate, in Gressoney the Walser community is still quite alive, also producing a significant number of written records which have been collected and stored in the digital archive so that they are freely accessible online. This will provide the main textual basis for the analysis presented below.

This paper will focus on the development of the so-called strong/weak verb classes in Titsch which have been reorganized in the context of a particular bundle of morphosyntactic features. The paper is structured as follows: in § 2 grammaticalization as an adaptive change is introduced, as well as its effect on the so-called complexity mismatch and on decategorialization, which is at the heart of the loss of compositionality and motivation of the original construction; in § 3 the complexity mismatch is discussed with regard to Titsch, providing a brief sketch of the properties of its verbal complex and depicting the development of the complexity mismatch in this variety. Finally, § 4 presents the conclusion.

2.   Grammaticalization as an adaptive change

Relying basically on analogous concepts exploited in evolutionary biology (see Lass, 1990, and the papers in Norde & Van de Velde, 2016), in Gaeta (2016) the conceptual pair adaptive vs. exaptive change was introduced. By adaptive changes a typology of changes is understood which are characterized as oriented and responding to a general design of economy and plasticity which manifests itself in the occurrence of developmental patterns which are repeatedly found language after language. In contrast, exaptive changes are normally non-oriented and result from the refunctionalization of extant linguistic material, partially or entirely devoid of any apparent function (see Smith, 2011 for a discussion of this point). Grammaticalization, basically intended in its most widespread meaning as the development of a former lexical item into a grammatical one, is deemed to represent an adaptive change insofar as it generally instantiates recurrent channels or paths of meaning extension leading from concrete to abstract; it is usually oriented, i.e. unidirectional, and presupposes a well-defined set of ecological conditions in which it is likely to take place, including communicative or social benefits attained by means of the extravagance of the perceptual benefits (the salience, the communicative efficacy, etc.) of the expressions undergoing grammaticalization. Furthermore, in light of the increased conventionality of the resulting markers one can also assume that grammaticalization has the effect of increasing the overall complexity of the grammar insofar as the latter has become more conventional or arbitrary as a consequence of the change.

To provide one concrete example, we can discuss the well-known case of the rise of two distinct verbal macro-classes observed in the Germanic languages, i.e. the older strong apophonic class, e.g. Gothic *hafjan* 'to raise' / *hōfum* 'we raised', and the newer weak class, e.g. Goth. *nasjan* 'to save' / *nasidēdum* 'we saved'. While the former goes back to an old Proto-Indo-European pattern, the latter results from a grammaticalization process which led the old causative DO to develop into an inflectional marker for the past tense (cf. Lehmann, 1989; Ringe, 2006, p. 167, Hill, 2010):[1]

(1) *nasjan dēdum* 'we did save' > *nasidēdum* 'we saved'

Many verbs belonging to this so-called "weak" class resulted from derivational processes based on other primary lexemes, namely strong class verbs, nouns, and adjectives (cf. Ringe, 2006, p. 252 and passim). By recruiting the causative verb DO, it was possible to build a past tense for these – mainly causative – verbs which could not exploit the older apophonic pattern. The grammaticalization of the DO periphrasis can be understood as an instance of an adaptive change insofar as it is strictly connected with favorable ecological conditions like the increase of the salience via emphasis and consequently of the communicative efficacy of the expression containing the originally causative verb. By developing such a pro-verb function, the latter also underwent a process of phonological reduction and of semantic bleaching, ultimately leading to its grammatical status as an inflectional marker (cf. Szczepaniak, 2009, pp. 112–117 for a detailed reconstruction). The process can be observed in several related changes described throughout the world's languages (cf. Heine & Kuteva, 2004, p. 119 for an overview).

2.1 Grammaticalization and the complexity mismatch

Once the new class is fully grammaticalized, it gives rise to the effect of paradigm layering, i.e. to inflectional classes, which is an instance of complexity mismatch in the sense of Francis & Michaelis' (2003) because a discrepancy is observed in the number of elements involved (and, consequently, in the complexity of the structure) at different levels of representation (see also Gaeta, 2008):

---

1. Actually, Ringe assumes that the original periphrasis contained a past participle: *nasida(n) dēdum* 'to make so. saved', subsequently reduced because of the loss of inflection and haplology. It must be added that other analyses have been suggested as well (see Hill, 2010 and references therein).

(2)

$$|\text{PAST}|$$

preterite	*hōfum*	*nasidēdum*
participle	*hafj-an-s*	*nasi-þ-s*
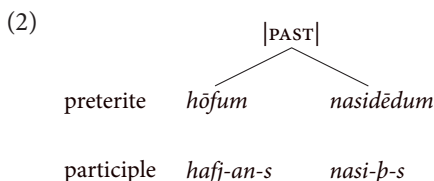
The same content |PAST| is associated with two different coding forms. Notice that the different morphological patterns found in the preterite are mirrored in the different suffixes selected in the past participle, a nasal and a dental, respectively. This paradigm layering is still found in every modern Germanic language, even though the strong class is observed to be generally recessive insofar as it loses members in favor of the newer dental class, while the opposite trend is seldom attested:

(3) a. MHG *melken / molk / gemolken* 'to milk / milked / milked'
              weben / wob / gewoben 'to weave / wove / woven'
                                ⇓
      MSG *melken / molk ~ melkte / gemolken* (*gemelkt*)
              weben / webte (rare *wob*) / *gewebt*

    b. Afr *breek* 'to break' / *ek het gebreek* 'I broke'
              werk 'to work' / *ek het gewerk* 'I worked'

In (3a) the two Middle High German (= MHG) verbs *melken* and *weben* used to belong to the strong class while their Modern Standard German (= MSG) followers shifted to the weak class. This happened almost entirely in the case of *weben* in which the older form *wob* is only found in stylistically marked contexts or in the more conservative varieties found in the South. On the other hand, in the case of *melken* the strong and weak forms co-exist in a classical context of ongoing variation in which the strong form of the past participle is more frequent while both forms are commonly used in the preterite. In the case of Afrikaans (= Afr) (3b) the preterite has completely disappeared while at the same time no apophonic alternation has survived in the past participle, with the exception of a couple of cases in which the preterite is still attested such as *dink* 'to think' / *dag ~ dog* 'thought (pret.)' / *gedink ~ gedag ~ gedog* 'thought (past part.)' and *weet* 'to know' / *wis* 'knew' / *geweet* 'known' (cf. Donaldson, 1993: 222). Besides, in a consistent number of participles used as adjectives the strong / weak forms are still preserved with the strong form typically reflecting a translated meaning: *'n gebroke hart* 'a broken heart' vs. *'n gebreekte koppie* 'a broken cup'. In other words, in all Germanic languages the strong class proves to be recessive and to cede to the weak class. In Afrikaans, this process has almost reached its end, radically eliminating the complexity mismatch.

2.2   Grammaticalization and decategorialization

One of the effects of grammaticalization is decategorialization (cf. Hopper & Traugott, 2003: 106) whereby morphosyntactic features of the original elements undergoing the change are lost. Moreover, such loss is normally held to be a clear-cut signal of the completion of the grammaticalization process. For instance, while in Old High German (= OHG) the past participle found in combination with the verb HAVE still displayed agreement with the object of HAVE and can accordingly be interpreted as an attributive participle (4a), this is not possible anymore after the completion of the change as shown by the MSG example in (4b):

(4)   a.   *phigboum   habeta   sum   giflanzot-an   in sinemo wingarten*
fig:tree[M]   had:3SG   some   planted-M.SG.ACC in his:DAT vineyard
'Somebody had a fig tree planted in his vineyard'   (Tat. 102, 2)

   b.   *Jemand   hat in seinem Weingarten ein-en   Feigenbaum gepflanzt*
somebody has in his:DAT vineyard   a-M.SG.ACC fig:tree[M]   planted
'Somebody has planted a fig tree in his vineyard'

   c.   *Jemand   hat ein-en   in seinem Weingarten gepflanzt-en*
somebody has a-M.SG.ACC in his:DAT vineyard   planted-M.SG.ACC
Feigenbaum
fig:tree[M]
'Somebody has a fig tree planted in his vineyard'

Notice that in MSG it is still possible to use the participle in attributive function (4c): in this case, while the latter displays agreement with the object of HAVE, the structure cannot be considered a grammaticalized perfect.[2]

Decategorialization is the reason why grammaticalization has to be understood as constructionalization to the extent that the new grammatical(ized) morpheme becomes part of a broader construction (cf. Traugott, 2003). As a matter of fact, the process of constructionalization implies a decategorialization insofar as the construction as a whole prevails over the compositionality and the categorial properties of its parts. To make one concrete example, when the Latin NP *clarā mente* 'with a clear mind' was grammaticalized into an adverb formed with the suffix *-mente* in almost any Romance language (cf. French *clairement*, Italian *chiaramente*, etc.), the original inflectional marker *-ā* signaling the agreement of

---

2.   There is a more marginal option in which the attributive participle does not belong to the nominal phrase as in (4c) but is extraposed in appositive position: *Jemand hat einen Feigenbaum, in seinem Weingarten aber gepflanzt* 'Somebody has a fig tree, planted however in his vineyard'. Notice that in this case no agreement between the direct object and the participle appears (cf. Murelli, 2017 for a general description).

the adjective with its nominal head was decategorialized into a dysfunctional base allomorphy required by the suffix. In some cases, it also displays true properties of constructionalization, for instance with Italian adjectives like *cruento* 'bloody', *violento* 'violent', etc. where the adverb has the form *cruente-mente, violente-mente* instead of the expected **cruenta-mente, *violenta-mente* because of the analogical influence of the quite large family of adjectives ending with *-nte*, mostly going back to old present participles like *corrente-mente* 'currently', *sapiente-mente* 'wisely', etc. (cf. Gaeta, 2016: 85). In this regard, it has to be stressed that inflectional markers – especially in languages belonging to the fusional type – generally are a clear cue of a lexeme's categorial membership, whose change or loss indicates decategorialization (cf. Gaeta, 2014 for a discussion of the issue with regard to conversions).

In the same vein, the original biclausal structure witnessed by (4a) in which the single parts could be independently decomposed and analyzed gives rise after the grammaticalization to a partially arbitrary monoclausal construction in which the past participle loses its independence and its morphosyntactic property of agreement while the originally main verb of possession becomes an auxiliary (cf. Harris, 2003):

(5) $[\text{Subject}_i \; haben \; \text{Object}_j \; [\text{Subject}_{i,k} \; \text{Verb}_{\text{PSTPTCP[+AGR]}j} \; \text{Object}_j]]$
$$\Downarrow$$
$[\text{Subject} \; [haben \; \text{Verb}_{\text{PSTPTCP[-AGR]}}] \; \text{Object}]$

Notice incidentally that before the grammaticalization the subject of the main verb, i.e. the possessor of the fig tree, and the subject of the embedded verb, i.e. its planter, might be interpreted as coreferential although this was not obligatory, while in the subsequent stage coreferentiality has become mandatory.

Moreover, the decategorialization of the participle resulting from the loss of agreement has become in MSG a canonical trait characterizing any past participle which has entered a grammaticalization process in contrast to OHG, as for instance in the passive construction formed either with the auxiliary *werden* 'to become' (6a) or with the auxiliary *sein* 'to be' (6b):

(6) a. OHG *huuanda so dhin-e   dag-a   arfullid-e*
               when   so your-PL.NOM day-PL.NOM completed-PL.NOM
               uuerdhant   (Is. 9, 3)
               become:3PL
               'when your days will be completed in this way'

     MSG *wenn dein-e   Tag-e   so erfüllt   werden*
               when your-PL.NOM day-PL.NOM so completed become:3PL
               'when your days will be completed in this way'

b. OHG *sint thir   furlazan-o   sunt-a*   (Tat. 193, 24)
         are you:SG.DAT forgiven-PL.NOM sin-PL.NOM
         'your sins are forgiven you'

   MSG *dein-e   Sünde-n   sind dir   vergeben*
         your-PL.NOM sin-PL.NOM are you:SG.DAT forgiven
         'your sins are forgiven you'

Decategorialization is even more radical when the perfect construction is combined with other constructions, as for instance with the modal construction and the so-called AcI-construction (from *Accusativus cum Infinitivo*) in which a bare infinitive is governed either by a modal or by the causative verb *lassen* 'to let':

(7) a. *Hans hat einen Feigenbaum pflanzen wollen   / *gewollt.*
         Hans has a   fig:tree   plant:INF want:INF / *want:PSTPTCP
         'Hans wanted to plant a fig tree'.
    b. *Hans hat jemanden einen Feigenbaum pflanzen lassen / *gelassen.*
         Hans has somebody a   fig:tree   plant:INF let:INF / let:PSTPTCP
         'Hans made somebody plant a fig tree'.

In these cases, for reasons that will not be discussed here (but cf. Gaeta, 2008, 2010, 2013) a substitutive infinitive takes the place of the expected past participle, which can be interpreted as a kind of featureless, unmarked form of the verb.

Finally, decategorialization also characterizes MSG adjectives insofar as they have lost any inflectional agreement in predicative position, i.e. when they carry the main verbal function either accompanied by a copula or in a small clause:

(8) a. OHG *sie sint… wīsduam-es foll-e*   (O. I, 1, 111–112)
              they are   wisdom-GEN full-PL.NOM
              'They are full of wisdom'

        MSG *sie sind voll an Weisheit*
              they are full at wisdom
              'They are full of wisdom'

    b. OHG *uuanta her sin-az   folc   heil-az   tuot fon*
              because he his-N.SG.ACC people[N] saved-N.SG.ACC does of
              iro   sunton   (Tat. 5, 8)
              their sins
              'because he will make his people saved from their sins'

        MSG *denn   er wird   sein Volk   von seinen Sünden*
              because he becomes his   people of   their sins
              heil   machen
              saved make:INF
              'because he will make his people saved from their sins'

The decategorialization of adjectives in MSG goes even further to the point that they have become multifunctional insofar as they have lost any overt marking in adverbial position (9a) and can now carry the function of verbal modifiers:

(9) a. OHG *ubil* 'bad' / *ubilo* 'badly' > MSG *übel* 'bad / badly'
    b. OHG *min tohter ubilo fon themo tiuuale giuuegit ist*
          my daughter badly of the:SG.DAT devil:DAT tormented is
          'my daughter is terribly tormented by the devil' (Tat. 85, 2)
      MSG *meine Tochter wird schrecklich vom*
          my daughter becomes bad of:DEF.SG.DAT
          Teufel gequält
          devil tormented
          'my daughter is terribly tormented by the devil'

The development of a typical analytic feature like multifunctionality, which blurs the clear-cut distinction between adjectives and adverbs that were still in use in OHG, has a number of far-reaching consequences with regard to a synchronically productive process of decategorization in MSG (cf. Gaeta, 2010, 2014 for details and discussion).

3. The complexity mismatch in Walser German

As in other Walser German varieties, in Titsch the verb inflectional classes of OHG are quite well preserved insofar as the class I corresponding to the original strong (apophonic) macro-class is distinct from the original weak classes II and III, going back respectively to the Proto-Germanic *-ja-* or *-ē-* and *-ō-* classes:[3]

(10)   class I: *bisse* 'to itch, bite' / *bésset*, *biete* 'to offer' / *bottet*, *fénne* 'to find' / *gfònnet*
        class II: *fiere* 'to lead' / *gfiert*, *läbe* 'to live' / *gläbt*
        class III: *moalò* 'to paint' / *gmoalòt*, *rächnò* 'to calculate' / *grächnòt*

Notice that as a consequence of the so-called *Präteritumschwund* (preterite loss) commonly found in Southern German varieties, all that is left of the stem vowel alternations typical of the strong class is found in the past participle (cf. Russ, 1990, p. 377), although past subjunctive forms are also sporadically attested with

---

3. The Titsch forms are provided in compliance with the orthographic norms adopted by the dictionary compiled by the *Walser Kulturzentrum*. Note that <é>, <ä> and <ò> roughly correspond to [ɪ], [æ] and [ʊ], respectively, while vowel sequences like <ie>, <ée>, etc. correspond to true (falling) diphthongs: [iə̯], [ɪə̯], etc. It must be added that the texts acquired in our database do not always follow these orthographic norms, also because to a large extent they were written before their adoption (cf. Angster et al., 2017 for discussion).

highly frequent verbs like *éch gange* 'I would go'. Moreover, the original opposition between the nasal vs. the dental suffix illustrated by the Gothic participles in (2) above has been levelled in favor of the dental suffix, although the stem vowel alternation is preserved as shown by Titsch *biete / bottet*, *fénne / gfònnet*, etc. in (10).[4] However, in a significant number of cases a certain variation is observed between the nasal and the dental suffix:

(11)   *bräche* 'to break' / *brochen ~ brochet*, *schribe* 'to write' / *gschrében ~ gschrébet*, *usstéerbe* 'to die out' / *usgstòrben ~ usgstòrbet*, *vergässe* 'to forget' / *vergässen ~ vergässet*, *verliere* 'to lose' / *verlòren ~ verlòret*, etc.

While this might be apparently interpreted on a par with the levelling in favor of the weak class as shown by the MSG examples in (3) above,[5] in Titsch the situation is essentially more complex, as will be shown below. In fact, the variation depicted in (11) happens to be construction-sensitive as it has been remodeled according to well-defined grammatical contexts. Before going into the detail, however, it is necessary to have a brief look at the verbal complex in Titsch in order to understand the real place of the alternation observed in (11).

---

4.   It must be added that with (weak) class II verbs the phenomenon of so-called *Rückumlaut* 'regressive umlaut' shows up also in non-etymological cases such as Titsch *schenge* 'to donate' / *gschengt ~ gschangt*, *setze* 'to seat' / *gesetzt ~ gsatzt*, etc. (cf. Zürrer, 1982, p. 92).

5.   Actually, this is the common interpretation provided in the traditional literature, as for instance by Bohnenberger (1913, p. 232): "Eine Neigung zu Beseitigung der starken Konjugation zeigt sich in der häufigen Bildung schwacher Konjunktive Præt. und passiver Partizipien zu starken Verben. Der Stammvokal ist dabei teilweise wie beim regelrechten starken Verbum abgelautet, so daß diese Bildungen sich ohne weiteres als Mischformen darstellen, teilweise aber auch mit dem des Præsens identisch, so daß diese Formen für sich allein genommen wie reine schwache aussehen". [A tendency towards the elimination of the strong class manifests itself in the frequent formation of weak subjunctive preterites and of passive participles of strong verbs. In these cases, the stem vowel is partially derived according to the regular apophonic pattern of the corresponding strong verb, so that these formations readily look like mixed forms, but also partially identical with the present stem, so that these forms taken by themselves appear purely weak, my translation].

3.1   Conservative and innovative features in the Titsch verbal complex

While decategorialization largely affects MSG as shown in § 2.2 above, in the Walser German varieties the inflectional properties of participles and adjectives are largely preserved. In particular, Walser German varieties are fairly well known for preserving conservative features like the systematic agreement of the adjective in predicative position:[6]

(12)   *Of de Beärga   esch d'Loft   guet-e   on   ds'Wasser*
      on the mountains is   the:F.SG-air[F] good-F.SG and the:N.SG-water[N]
      escht gsond-s,   Der   Appetit   escht doa   gross-e,   der
      is   healthy-N.SG, the:M.SG appetite[M] is   there big-M.SG the:M.SG
      Mage   wie an   Bronz   (BEL_0019)
      stomach[M] as   a:M.SG cauldron[M]
      'On the mountains the air is good and the water is healthy, the appetite is big there and the stomach like a cauldron'

This trait also characterizes the other Walser German varieties spoken in Switzerland, as shown by the Example (13a) from Bosco Gurin (cf. Russ, 1990: 373), and more in general the Highest Alemannic dialects (13b) (cf. Fleischer, 2007: 221):

(13)   a.   *ds   weerch   isch gmachut-s.*
         the work[N] is   made-N.SG
         'The work is done'.
     b.   *Das   ischt nie luub-s.*
         this:N.SG is   not scary-N.SG
         'This is fishy'.

To account for this state-of-affairs, an internal explanation has traditionally been suggested consisting in the direct preservation of the OHG situation sketched in (8) above; alternatively, the effects of a more or less recent contact with Romance varieties have been emphasized (see Fleischer, 2007 for a detailed discussion and further references). At any rate, the effects of the constant contact with Italian and with the other Romance varieties are clearly tangible in the robust system of agreement found in Titsch.

    Reproducing strictly the constructional model of the predicative construction, participle agreement is also found in the perfect when the BE-auxiliary is selected by unaccusative verbs:

(14)   *Hilde   òn Cristina   sinn drobèr gsatzt-é*   (D_0010)
      Hilde[F] and Cristina[F] are thereon seated-F.PL
      'Hilde and Cristina are seated on that'

---

6.   The examples given in the text follow the numeration assigned in the corpus accessible online at www.archiwals.org. The latter presently amounts to about 41,000 tokens for the variety of Gressoney.

However, uninflected forms are normally found with weak verbs (15a–b), also verbs belonging to the class I, in which case they take the form of weak – i.e. dental-suffixed – participles, independently of the apophony of the stem vowel which is mostly retained (15c–d):[7]

(15) a. *Uf em   obr-e   Platz   sinn d'Medra*
     on the:M.SG.OBL upper-M.SG place[M] are the-mowers
     bim   Ronkreschtentsch-Hus gsetzt   (DOK-0086)
     at:DEF.N.SG.OBL R.-house[N]   seated
     'At the upper place the mowers are seated close to the Ronkreschtentsch-house'

   b. *wenn éscht-er eréfiert en Amerika*   (DOK_0002)
     when is-he   arrived in A.
     'when he arrived in America'

   c. *Laura òn Augusta sinn vòn ussna kchéemet als jòng-é*
     Laura and Ausgusta are of abroad come   as young-F.PL
     techtre   (DOK_0016)
     daughters[F]
     'Laura and Augusta came from abroad as young daughters'

   d. *Zumschtein escht anno 1783 z'Noversch (Gressoney) bored on*
     Zumschtein is   year 1783 to-Noversch (Gressoney) born and
     escht doa gschtorbet anno 1861   (DOK_0041)
     is   there died   year 1861
     'Zumstein was born in Noversch (Gressoney) in 1781 and died there in 1861'

On the other hand, the strong verbs of the class I also display either a form ending with the old nasal suffix which agrees with the subject (16a) or an uninflected form resulting from the deletion of the final nasal (16b–c):

(16) a. *Kréchterle hät glotzt   bés és sécher-s gsid dass de*
     K.[M]   has lain:in:wait until is sure-N.SG been that de
     Jònker-Ronker ésch rächt antschloafn-e   gsid   (DOK_0123)
     J.[M]   is   right fallen:asleep-M.SG been
     'Kréchterle lay in wait until he was sure that Jònker-Ronker had fallen completely asleep'

   b. *e chatzò éscht kchéeme én vòn der   fänschtrò*
     a cat[F] is   come in of the:F.SG.OBL window
     'a cat came in through the window'   (Angster, 2004–05, p. 66)

---

7. It must be added that the participle of auxiliary verbs is never inflected in neat contrast with the predicative adjective:

   (i) *de Greschoneyera woa en titsch-e   Gägente teteg-e   sin gsid*
     the Greschoneyera that in German-PL countries active-N.PL are been
     'the people from Gressoney who were active in German regions'   (DOK_0048)

c. *òn én ganz Europa sin vélle dra gschtòrbe* (DOK_0014)
   and in whole Europe are many there:at died
   'and in all of Europe many died because of it'

d. *aber all-e beuvisen dass éndsché verstòrbn-ò gangen nie* but all-PL attest that our:PL deceased-PL go:3.PL.SUBJ.IPF never
   vergässn-e (DOK_0190)
   forgotten-PL
   'but everybody attests that our deceased should never be forgotten'

The full inflected form containing the nasal is still found in cases of adjectival or nominal usage of the participle as shown by the nominalized participle found in (16d), going back to an unattested verb °*verschtéerbe* 'to pass away'. In a similar way, nasal-ending past participles of class I verbs normally displaying the dental suffix (*biete* 'to bid' / *bottet* 'bid', *vergoa* 'to pass away' / *verganget* 'gone', *glénge* 'to succeed' / *glònget*) are still found when they are used as adjectives:

(17) a. *Schtefantsch-Tag esch gsid bottn-e Fiertag* (DOK_0076)
       S.-day is been bidden-M.SG holiday[M]
       'Stephen's day was a holy day of obligation'

   b. *en dé vergangn-é zitte* (DOK_0015)
      in the:F.SG past-F.SG time[F]
      'in the past time'

   c. *En wònderbar glungn-e oabe* (DOK_0185)
      a:M.SG wonderful succeeded-M.SG evening[M]
      'A wonderfully successful evening'

This reminds us of the scenario depicted in § 2.1 above for Afrikaans in which strong participles are only retained in adjectival usage of the past participles.

Another conservative feature which is fairly robust in the Walser German varieties is the preservation of the BE-passive as the normal, unmarked way to express passive voice exemplified below with different tenses and moods, respectively present, perfect and infinitive:[8]

(18) a. *òn en der Chélchò em Lido vòn Venedig sinn*
       and in the:F.SG.OBL church in:M.SG.OBL Lido of Venice are
       ufbewart-é dschin Reliquie (DOK_0002)
       preserved-PL their remains
       'and in the church at the Lido of Venice their remains are preserved'

---

8. Even though also in this case the influence of Italian and of the Romance varieties which normally display a BE-passive as the unmarked form is tangible, it should not be forgotten that the BE-passive is also commonly found in Alemannic varieties, especially with a resultative value, as for example in Zürich: *De Stock isch verbroche* 'The stick is broken' (cf. Reese, 2007, p. 46).

b. *chammo erchenne wette d'Beldong   von de*
can:one recognize how   the:F.SG-culture of   the:PL
Greschoneyer-Walser set   1500 bes   1850 entwecklot-e   gsid escht
Greschoneyer-Walser from 1500 until 1850 developed-F.SG been is
'one can recognize how the culture of the Greschoneyer-Walser was developed from 1500 to 1850'

c. *hätt-er welle   si   vergrabt-e   em   eigen-e   land*
had-he want:INF be:INF buried-M.SG in:N.SG.OBL own-N.SG country
'he would have wanted to be buried in his own land'   (DOK_0016)

Notice that in all these cases the past participle of the main verb systematically agrees with the subject, reproducing once more the schema of the copula construction seen in (12) above while the past participle of the BE-auxiliary is uninflected (see fn. 7 above).

In addition to the BE-passive two other types of passives are found in Titsch in which respectively the verb COME – only in the present tense (19a) – and the verb GO – only in the past tense (19b) – are used as auxiliaries (cf. Gaeta, 2018 for a detailed discussion on the role played by Italian as contact language):

(19) a. *hitzòtag   sinn éndsche woalda   bschétzt-e   an*
nowadays are   our:M.PL forests[M] protected-M.PL and
d'Wiehnachtsboumiéné chéemen   kontròlliert-e   (DOK_0202)
the-Christmas:trees   come.3PL monitored-M.PL
'Nowadays our forests are protected and the Christmas trees are monitored'

b. *De   toufnoamna sin of franzesésch abkändret-e   kanget*
the:M.PL forenames[M] are up French   changed-M.PL gone'
'The forenames were changed into French'   (DOK_0014)

While the COME-passive is widespread throughout the whole Alpine area (cf. Wiesinger, 1989; Ramat, 1998; Wiemer, 2011), the GO-passive is a particular development of (some of) the Walser German dialects. Actually, the GO-passive is also used in the present, typically with a strong modal (deontic) value as shown in the following example:

(20) *al-z   was ésch kannet ufgchrében-z   geit vòn allerhand glòrt-e*
all-N.SG what is   gone up:written-N.SG goes of all:sorts learned-PL
litte   genou prueft-s   (DOK_0124)
people exact proved-N.SG
'everything which has been noted has to be exactly proved by all sorts of learned people'

Notice that in all the examples in (18), (19) and (20) the past participle of the main verb systematically agrees with the sentence subject basically reproducing the syntactic schema of the copula construction in (12).

In sharp contrast with these cases, the past participle is uninflected when the HAVE-auxiliary is selected in the perfect, which clearly shows the effects of the decategorialization discussed in § 2.2 above:[9]

(21) a. *wòa   éndschè Sèndég hät fèr éndsch artòat   / *artòat-s   /*
         where our   mayor has for us   opened / *opened-N.SG /
         *artòat-é   an paar butèllè wi   (DOK_0010)
         opened-PL a   pair bottles wine
         'where our mayor opened for us a couple of bottles of wine'

     b. *Wéwél   foto   héintschnè   gmacht / *gmacht-e !!* (DOK_0010)
         how:much picture have:they:them made   / *made-PL
         'How many pictures they took of them!!'

Notice that the HAVE-perfect is also found with unergative verbs such as *lache* 'to laugh' / *glachet*, *schloafe* 'to sleep' / *gschloafet*, etc., in which unsurprisingly only the uninflected weak form occurs:

(22) a. *aber héibèr   véll   glachet*   (DOK_0010)
         but   have:we much laughed
         'but we laughed a lot'

     b. *Noch én der   glichò nacht wenn alle   hein   hert*
         still   in the:F.SG.OBL same night when all.PL have:3PL hard
         gschloafet   (DOK_0199)
         slept
         'still in the same night when everybody was sleeping deeply'

Besides the development of the perfect shared by the whole German-speaking area, a couple of innovative features occur in Titsch which are generally found in the Southern dialects as well as in the West-Germanic languages, namely the diffusion of the DO periphrasis in the pro-verb function (see Weber, 2017 for a recent overview):

---

9. Notice that the only case of past participle agreement in the presence of the HAVE-perfect reflects an attributive construction like that seen in (17) above:

(i) *Mo häd die   vom   Heer kriegt, wenn mo d'Oschtre häd*
     one has those by:DEF.M.SG.DAT lord taken, when one the-Easter has
     gmacht-e   ghätt   (D_0047)
     made-PL.ACC had
     'One took those from the Lord, when one received those made at Easter'

(23) a. *va   woa   de noame tuet chéeme weiss mo ni rächt*
from where the name   does come   knows one not right
'One does not know precisely wherefrom the name comes' (DOK_0014)

b. *D'léck-é   zòcht   tin ganz   täl sproache lére*   (DOK_0016)
the-small-PL children do entirely easy languages learn
'The small children do quite easy learn languages'

The auxiliary verb *tin* 'to do' carries the agreement features and basically replaces the inflected forms of the main verb which appears in the infinitive. In the long run, this substitutive effect can be deemed to have the general impact of reducing the inflectional inventory of the verbs as the latter will usually appear in the infinitive form (cf. Angster & Gaeta, 2018 for a discussion on the basis of the so-called 'short verbs'), unless the pro-verb becomes an inflectional marker as in the case of the dental suffix of weak verbs discussed above.

Finally, Titsch also shares with Southern dialects the development of the so-called *Doppelperfekt* (cf. Zybatow & Weskott, 2018 for a recent overview), in which as a consequence of the generalized preterite loss a complex tense displaying a sequence of two past participles is found with a pluperfect reading as shown in Example (16a) above and in (24):

(24) *Aber sibber noch ni   gsit   arrivièrt*   (DOK_0010)
but are:we yet   not been arrived
'But we had not yet arrived'

Note that the past participle of the main verb swings between inflected forms agreeing with the subject as in Example (16a) above or uninflected forms as in this latter Example (24), as has been discussed above for the BE-perfect (see (15) above).

3.2 Remodeling the complexity mismatch

Given *Präteritumschwund*, the inflectional classes which result from the grammaticalization of DO in early Germanic and which are responsible for the complexity mismatch are only visible in the past participle. In fact, as shown in the previous paragraph, weak forms ending with *-et* – independently of the apophony of the stem vowel – are dominant and are extended to older strong forms ending with *-en*. In spite of the apparent parallel hinted at above with the situation in Afrikaans, in which strong participles are only found with lexicalized adjectives, strong participles in Titsch are not merely the relic of an earlier inflectional class. In fact, the strong/weak alternation has been completely remodeled according to the construction type in which the (strong) verb is involved. Accordingly, the doublets mentioned in (11) follow a clear distribution which relies upon the constructional type pertaining to that participle:

(25) a. *wenn heintsche brochet hein   d'bueba   de henne d'fädre*
         when have:they broken have:3PL the-children the hens   the-feathers
         uszochet   (DOK_0192)
         stripped:off
         'When they broke them, the children stripped off the feathers of the hens'

   b. *Dez   Hus   escht ends vom   letscht-e Joahrhondert*
         the.N.SG house[N] is   at:end of.N.SG.OBL last-N.SG century[N]
         abbrochen-s   canget   (DOK_0021)
         off:broken-N.SG gone
         'The house was demolished at the end of last century'

When the construction selects the HAVE-auxiliary (25a), the weak form *-et* appears, which does not display any agreement in accordance with its decategorialized character. On the other hand, the strong suffix *-en* shows up in the past participle when the construction selects other auxiliaries such as, for instance, the GO-passive in (25b), forcing the occurrence of agreement. It should be stressed that it is not possible to have an inflected form of the participle of the strong verb when the latter appears with the weak dental suffix: **Dez Hus escht abbrochets canget*. Such a state of affairs is regularly found also with the BE-passive and with the COME-passive (besides the BE-perfect seen in (16) above) in contrast to the HAVE-perfect, as shown by the following pairs of examples (26)–(29) reflecting the variation depicted in (11) above:

(26) *vergässe* 'to forget' / *vergässet ~ vergässen*
   a. *òn   hatt-er nie   vergässet dschis Titsch*   (DOK_0016)
         and has-he never forgotten his   Titsch
         'and he never forgot his Titsch'
   b. *ietza en Ireland d'oalt-ò   sproach,   daß ésht gsit schier*
         now in Ireland the-old-F.SG language[F] that is   been almost
         ganz   vergässn-é,   éscht nid nòmma läbén-é   (DOK_0016)
         entirely forgotten-F.SG is   not only   alive-F.SG
         'now in Ireland the old language which was almost completely forgotten
         is not only alive'

(27) *schribe* 'to write' / *gschrébet ~ gschrében*
   a. *wenn Benito Leopold Curtaz… hät éndsch gschrébet*   (DOK_0015)
         when B.   L.   C.   has us   written
         'when Benito Leopold Curtaz … wrote to us'
   b. *al-z   éscht kanget gschrében-z*   (DOK_0015)
         all-N.SG is   gone   written-N.SG
         'everything was written'

(28) *verliere* 'to lose' / *verlòret ~ verlòren*

    a.   *Doa häd de Tifol d'Wetté verlored*   (DOK_0015)

        there has the devil the-bet lost

        'there the devil lost the bet'

    b.   *Was   hein   éndsché Òaltò tòat òn gseit ésch nie   verlòren-z*

        what[N] have:3PL our   elders done and said is   never lost-N.SG

        'what our elders did and said is never lost'   (DOK_0011)

(29) *wässche* 'to wash' / *gwässchet ~ gwässchen*

    a.   *Kroa vor   dä Fiertaga hämmu z'Hus   al-z   subber*

        right before the holidays has:one the:N.SG-house[N] all-N.SG clean

        gwescht   (DOK_0076)

        washed

        'Immediately before the holidays one cleaned the whole house'

    b.   *Wiber un Techtre   hein   ou chät en der   Chelcho*

        wives and daughters have:3PL also had in the:F.OBL church[F]

        z'tue: di   hät musso si   ni numma subber gwäschn-e

        to-do: that:F.SG has must:INF be:INF not only   clean washed-F.SG

        'Wives and daughters also had to do in the church: it didn't only need to

        be cleaned'   (DOK_0083)

    c.   *di   hät mòssò chéeme suber gwässchn-e*   (fieldwork)

        that:F.SG has must:INF come:INF clean washed-F.SG

        'That had to be cleaned'.

In general, when agreement, i.e. inflection, is required by the syntactic environment, the strong participles regularly appear, as in the context of a small clause (30a) and of an appositive construction (30b), as well as in the attributive position already seen in (17) above:

(30) a.   *Fenné où ufgschrében-z   of enz notizbuech dass wier hättéber*

        find:1SG also up:written-F.SG on our notebook that we   had:we

        sollò   noasieché   (DOK_0015)

        should:INF investigate

        'I also find written down on our notebook what we should have had to

        investigate'

    b.   *mé dem   eigen-e   noame drònder gschrébn-e*

        with the:M.SG.OBL own:M.SG name[M] thereunder written-M.SG

        'with one's own name written thereunder'   (DOK_0002)

Accordingly, the lexically motivated distribution of verbs belonging to the strong classes in earlier stages of this Upper German variety (cf. Bohnenberger's quote in fn. 5) has been remotivated as construction-specific. The class I verbs display either the strong or the weak suffix depending on the construction in which they are employed:

(31) a. strong form *-n*
  i.   [Subj$_j$ [[AUX$_{BE/GO/COME}$] PastPart$_{+AGRj}$]
  ii. [Subj [[V] Obj$_j$ PastPart$_{+AGRj}$]
  iii. [NP$_j$ PastPart$_{+AGRj}$]
  iv. [PastPart$_{+AGRj}$ N$_j$]$_{NP}$
 b. weak form *-t*
  [Subj$_j$ [[AUX$_{HAVE}$] PastPart$_{-AGRj}$]

When agreement is operative, the strong form is selected (31a), while the uninflected participles select the weak suffix showing the decategorialization resulting from grammaticalization of the perfect construction (31b). Weak past participles of the verb classes II and III are also compatible with the constructions requiring agreement (31a), as shown in Examples (14), (18), and (19) above. Thus, it is not the case that in the constructional schemas of (31a) only inflected strong participles are allowed.

The crucial point is that only with the (strong) class I verbs has a refunctionalization taken place whereby morphological expression is given to the different syntactic environments in which the participles occur. In other words, the complexity mismatch characterizing the inflectional class distinction resulting from the grammaticalization of the DO periphrasis seen in (2) above was not levelled as happened in Afrikaans. Instead, the mismatch was remotivated by increasing the number of past participles found in the paradigm of class I verbs. Far from being a recessive trait like in Afrikaans, where it is only found with some inflected adjectives going back to older participles, the strong/weak distinction has been significantly refunctionalized to convey agreement features in specific constructional schemas.

How can we interpret this change from a general perspective? Should we consider this on a par with grammaticalization because of the rise of a new grammatical differentiation between strong and inflected vs. weak and uninflected participles? As shown above, the opposition between two inflectional classes in Gothic ultimately goes back to what in Gaeta (2016) has been called an adaptive change, namely the grammaticalization of the DO periphrasis illustrated in (1) above for Gothic. This is characterized by bleaching, constructionalization, and decategorialization. In our case, the new opposition between inflected and uninflected past participles displaying the strong and the weak suffix, respectively, is a means for strongly characterizing the participles, depending on the specific construction in

which they are involved. Far from displaying bleaching, as in typical grammaticalization, the refunctionalization of the strong/weak distinction can be treated as an exaptive change, the reuse of a morphological differentiation devoid of any evident function. In fact, the strong / weak inflectional classes of Gothic as well as of MSG – categories which used to have a fully lexical distribution – have been remotivated in Titsch insofar as the class I verbs display two participles distributed across different constructions.

On the other hand, this change cannot be treated as a degrammaticalization insofar as it cannot be taken to represent a step backwards on the grammaticalization cline (cf. Norde, 2009: 8). In fact, the new differentiation cannot be said to be a step towards restoring the original value of the DO periphrasis, but rather follows its own way, moving orthogonally with regard to the original diachronic development. This is typical of an exaptive change insofar as it contributes to the stabilization of a system which has undergone dramatic changes challenging its original order. Paraphrasing Sturtevant's famous paradox ('sound change is regular but causes irregularity; analogy is irregular but causes regularity'), we can say that adaptive changes – and in particular grammaticalization – follow general patterns but have the effect of destabilizing a system, introducing, for instance, a complexity mismatch, while exaptive changes follow quite specific patterns but have the effect of restoring a certain functionality within a system.

It must be added that this picture only holds for some Walser German varieties like Titsch. It is also found in Titschu, the variety spoken in Rimella (32a–b), for which a similar behavior is observed, but not in Pomattertitsch spoken in Formazza (32c–d), in which strong verbs preserve their original shape (data from personal fieldwork):

(32) a. *Där   Mario het-s   gwascht z'glosch*
          the:M.SG M.   has-N.SG washed the:N.SG-glass[N]
          'Mario washed the glass'

      b. *Z'glosch   isch schit gwaschu-s*
          the:N.SG-glass[N] is   been washed-N.SG
          'The glass was washed'

      c. *Der   Marjo het der   Chrög brochä*
          the:M.SG M.   has the:M.SG vase[M] broken
          'Marjo broke the vase'

      d. *Geschter éscht der   Chrög fom   Marjo*
          yesterday is   the:M.SG vase[M] of:DEF.M.SG.OBL M.
          prochn-ä   cho
          broken-M.SG come
          'Yesterday the vase was broken by Marjo'

It remains to be understood how far the reshaping of the complexity mismatch observed in Titsch and Titschu will also extend to the other varieties and, above all, whether it will also systematically affect the class II and III verbs, giving rise to a generalized association of the strong participles with the constructional schemas in (31a) in which agreement serves as a triggering factor. This would reflect a further step of the exaptive change towards a full remotivation of the purely lexically governed strong/weak class inflection found in the rest of the Germanic languages.

4. Conclusion

To sum up, in Titsch the past participle of (strong) class I verbs has been remodeled and remotivated ("exapted") depending on the construction in which it is involved. Namely, strong forms containing the nasal suffix are preserved when they are inflected, i.e. syntactically operational, namely in the following array of constructions:

1.	the BE-perfect construction	Example (16a–c)
2.	the passive construction	
	a. BE-passive	Examples (20), (26b), (28b), (29b)
	b. GO-passive	Examples (25b), (27b)
	c. COME-passive	Example (29c)
3.	the small clause construction	Example (30a)
4.	the appositive construction	Example (30b)
5.	the attributive construction	Example (17)

In all these cases, they resemble the behavior of adjectives in copula constructions seen in (12). Weak forms of the past participle displaying the dental suffix only appear when the latter is decategorialized (uninflected) in the HAVE-perfect construction and, apparently, only in this environment, as shown by examples (25a), (26a), (27a), (28a) and (29a) above. Furthermore, note that the past participles of all auxiliaries (*kanget*, *gsid*) are weak and decategorialized.

On the other hand, past participles of (weak) class II and III verbs do not display this construction-related inflectional split but share the same morphosyntactic behavior of inflected and uninflected forms. Therefore, the occurrence of strong past participles is not a specific property of the single constructions as a whole, but is only related to the (strong) class I verbs when they enter into such constructional schemas, as summarized in (31) above.

With regard to the picture of the complexity mismatch as it holds in Gothic and many other Germanic languages provided in (2) above, in Titsch it has been remodeled in the following terms:

(33)

$$|\text{PAST}|$$

| class I | *broche-t*[− **inflected**] | *broche-n*[+ **inflected**] |

class II   *gfier-t*[± **inflected**]

class III   *gmoalò-t*[± **inflected**]
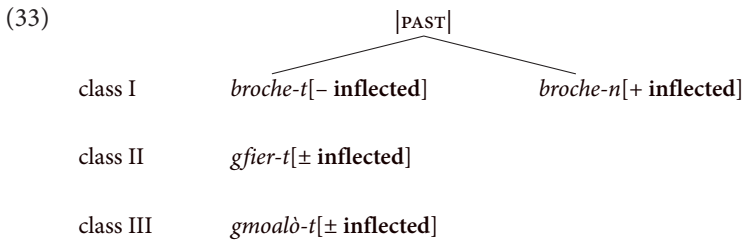
The strictly lexically-governed distribution of strong and weak markers requiring previous lexical knowledge found in Gothic as well as in MSG has been functionalized depending on specific constructional schemas connected with the activation of inflectional agreement: when the latter is activated, the strong nasal suffix is selected, while the weak dental suffix appears elsewhere. It remains to be seen whether the exaptive change underlying the remodeling of the complexity mismatch can be further generalized to other verb classes beyond the original domain of the strong verbs in Titsch, as well as in other varieties in which the exaptive change has taken place. This would remove the need for previous lexical knowledge on the part of the speakers, but this must be left for future research.

Acknowledgements

Abbreviations

ACC	accusative	N	neuter
Afr	Afrikaans	NOM	nominative
AGR	agreement	NP	noun phrase
AUX	auxiliary	OBJ	object
DAT	dative	OBL	oblique
DEF	definite	OHG	Old High German
F	feminine	PL	plural
GEN	genitive	PST	past
INF	infinitive	PTCP	participle
IPF	imperfect	SG	singular
M	masculine	SUBJ	subject
MHG	Modern High German	3	third person
MSG	Modern Standard German		

Bibliography

Angster, M. (2004–05). La perifrasi *tue* + infinito nel titsch di Gressoney. BA-Thesis, University of Turin.

Angster, M., Bellante, M., Cioffi, R. & Gaeta, L. (2017). I progetti DiWac e ArchiWals. in L. Gaeta (Ed.), *Le isole linguistiche tedescofone in Italia. Special issue of Bollettino dell'Atlante Linguistico Italiano*, 41, pp. 83–94.

Angster, L. & Gaeta, L. (2018). Wie kurz sind die Kurzverben? Morphologische Merkmale in Gressoney und Issime. In S. Rabanus (Ed.), *Deutsch als Minderheitensprache in Italien. Theorie und Empirie kontaktinduzierten Sprachwandels.* Special issue of *Germanistische Linguistik*, pp. 239–240, 211–237.

Bohnenberger, K. (1913). *Die Mundart der deutschen Walliser im Heimatland und in den Außenorten.* Frauenfeld: Huber.

Del Negro, S. (2004). *The Decay of a Language: The Case of a German Dialect in the Italian Alps.* Bern: Peter Lang.

Donaldson, B. C. (1993). *A Grammar of Afrikaans.* Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110863154

Fleischer, J. (2007). Zur Herkunft des flektierten prädikativen Adjektivs im Höchstalemannischen. *Zeitschrift für Dialektologie und Linguistik*, 74(2/3), 196–240.

Francis, E. J. & Michaelis, L. A. (2003). Mismatch: A crucible for linguistic theory. In E. J. Francis & L. A. Michaelis (Eds.), *Mismatch: Form-function incongruity and the architecture of grammar* (pp. 1–27). Stanford, CA: CSLI.

Gaeta, L. (2008). Mismatch: Grammar distortion and grammaticalization. In M. J. López-Couso & E. Seoane (Eds.), *Rethinking Grammaticalization. New perspectives* (pp. 103–127). Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.76.07gae

Gaeta, L. (2010). The invisible hand of grammaticalization: West-Germanic substitutive infinitive and the prefix *ge-*. In W. U. Dressler, D. Kastovsky, H. Ch. Luschützky & F. Rainer (Eds.), *Variation and change in morphology* (pp. 89–105). Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.310.05gae

Gaeta, L. (2013). Multiple sources for the German scandal construction. *Studies in Language*, 37(3), 566–598. https://doi.org/10.1075/sl.37.3.05gae

Gaeta, L. (2014). On decategorization and its relevance in German. In R. Simone & F. Masini (Eds.), *Word classes: nature, typology and representations* (pp. 227–241). Amsterdam: John Benjamins.

Gaeta, L. (2016). Co-opting exaptation in a theory of language change. In M. Norde & F. Van de Velde (Eds.), *Exaptation in language change* (pp. 57–92). Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.336.03gae

Gaeta, L. (2018). Im Passiv sprechen in den Alpen. *Sprachwissenschaft*, 43(2), 221–250.

Harris, A. C. (2003). Cross-linguistic perspectives on syntactic change. In B. D. Joseph & R. Janda (Eds.), *The handbook of historical linguistics* (pp. 529–51). Oxford: Oxford University Press. https://doi.org/10.1002/9780470756393.ch16

Heine, B. & Kuteva, T. (2004). *World lexicon of grammaticalization.* Cambridge: Cambridge University Press.

Hill, E. (2010). A case study in grammaticalized inflectional morphology. Origin and development of the Germanic weak preterite. *Diachronica*, 27(3), 411–458. https://doi.org/10.1075/dia.27.3.02hil

Hopper, P. J. & Traugott, E. C. (2003). *Grammaticalization. Second Edition*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139165525

Lass, R. (1990). How to do things with junk: Exaptation in language evolution. *Journal of Linguistics*, 26, 79–102. https://doi.org/10.1017/S0022226700014432

Lehmann, Ch. (1989). Markedness and grammaticalization. In O. Tomić (Ed.), *Markedness in synchrony and diachrony* (pp. 175–190). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110862010.175

Murelli, A. (2017). Partizipialattribute. In L. Gunkel, A. Murelli, S. Schlotthauer, B. Wiese & G. Zifonun (Eds.), *Grammatik des Deutschen im europäischen Vergleich. Das Nominal* (pp. 1807–1838). Berlin: de Gruyter. https://doi.org/10.1515/9783110341461-023

Norde, M. (2009). *Degrammaticalization*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199207923.001.0001

Norde, M. & Van de Velde, F. (Eds.) (2016). *Exaptation in language change*. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.336

Ramat, P. (1998). Typological comparison and linguistic areas: Some introductory remarks. *Language Sciences*, 20(3), 227–240. https://doi.org/10.1016/S0388-0001(98)00001-1

Reese, J. (2007). *Swiss German. The Modern Alemannic Vernacular in and around Zurich*. München: Lincom Europa.

Ringe, D. (2006). *From Proto-Indo-European to Proto-Germanic*. Oxford: Oxford University Press.

Russ, Ch. V. J. (1990). High Alemannic. In Ch. V. J. Russ (Ed.), *The Dialects of Modern German. A Linguistic Survey* (pp. 364–393). London: Routledge.

Smith, J. Ch. (2011). Change and continuity in form–function relationships. In M. Maiden, J. Ch. Smith, & A. Ledgeway (Eds.), *The Cambridge history of the Romance languages. Volume 1: Structures* (pp. 268–317). Cambridge: Cambridge University Press.

Szczepaniak, R. (2009). *Grammatikalisierung im Deutschen*. Tübingen: Narr.

Traugott, E. C. (2003). Constructions in grammaticalization. In B. D. Joseph & R. Janda (Eds.), *The handbook of historical linguistics* (pp. 624–647). Oxford: Oxford University Press. https://doi.org/10.1002/9780470756393.ch20

Weber, Th. (2017). *Die TUN-Periphrase im Niederdeutschen. Funktionale und formale Aspekte*. Tübingen: Stauffenburg.

Wiemer, B. (2011). The grammaticalization of passives. In H. Narrog & B. Heine (Eds.), *The Oxford Handbook of Grammaticalization* (pp. 535–546). Oxford: Oxford University Press.

Wiesinger, P. (1989). Zur Passivbildung mit *kommen* im Südbairischen. In W. Putschke, W. Veith & P. Wiesinger (Eds.), *Dialektgeographie und Dialektologie* (pp. 256–268). Marburg: Elwert.

Zürrer, P. (1982). *Wörterbuch der Mundart von Gressoney*. Frauenfeld: Huber.

Zybatow, T. & Weskott, Th. (2018). Das Doppelperfekt: Theorie und Empirie. *Zeitschrift für Sprachwissenschaft*, 37(1), 83–124. https://doi.org/10.1515/zfs-2018-0003

From noun to quantifier

Pseudo-partitives and language change

Johanna L. Wood
University of Aarhus

Micro-variation in Germanic pseudo-partitives is investigated, focusing on the syntax and morphology of expressions where two nouns are juxtaposed without a linking preposition, attested in Middle English. A standard measure (*pound*), containers (*gallon*, *ton*, *pipe*, *barrel*, and *sack*) and a collective (*pair*) are examined in detail. Few similarities are found between the Middle English expressions and other Germanic languages for standard measures and containers, whereas the noun *pair* has syntactic, morphological and semantic similarities.

Keywords: pseudo-partitives, container nouns, measure nouns, collective nouns, Germanic, Middle English, Early Modern English, singular/plural

1. Introduction

As is frequently pointed out, although English groups genetically with the Germanic language family, it is, in many ways, morphologically and syntactically a typological outlier of that family (Harbert, 2007, p. 13; Lass, 1987, p. 317–32; McWhorter, 2002, p. 265). Ways in which English differs from its sister languages include development of *do*-support and loss of grammatical gender. The expression investigated in this paper, the pseudo-partitive, is another way in which English differs from its Germanic sisters. Pseudo-partitives are nominal expressions containing two nouns, designated N_1 and N_2, where N_2 is either non-count or plural. Although most modern Germanic languages optionally juxtapose N_1 and N_2 to form pseudo-partitives in the Direct Partitive Construction (DPC), present-day English requires a preposition, termed the Indirect Partitive Construction (IPC) (van Riemsdijk, 1998, p. 11).

(1) Da.
    a. *en spand vand*
       a bucket water
       'a bucket of water'

(2)   En.
    a.   a bucket of water
    b. *a bucket water

This, however, was not always the case. As Mustanoja (2016/1960, p. 84) observes, the DPC construction is "not uncommon in the 13th and 14th centuries" and is found in Chaucer's verse (*no morsel breed*; 'no morsel of bread') and in Gower (*a paire scheres*; 'a pair of shears').

Although the presence of the DPC in Middle English has long been known, there seems to be little detailed diachronic work on its development (and subsequent demise). As a first step, this paper attempts to ascertain how established the expression was in Middle English (ME). As a starting point I take claims made by Grestenberger (2015) that ME developed both DPCs and IPCs, although only the IPCs were "generalised in modern Standard English" (Grestenberger, 2015, p. 128). The same article makes additional claims about similarities and differences between English and German as regards singular/plural morphology. The significance of these claims is addressed briefly in Section 2 with regard to German and present-day English; further data from ME is needed if such claims are to be supported.

It appears, then, that English started to develop a DPC system which did not last into the modern period. Granted, changes do not necessarily have to go to "completion". There can be actuation (innovation) and partial transmission of a variant (cf. Weinreich, Labov & Herzog, 1968) which does not spread. A well-known example in English is the failure of auxiliary *do* to extend to all clauses, not just negatives and questions, despite the finding that between 1400 and 1600, it was possible to use auxiliary *do* in non-emphatic contexts, known as "positive affirmative *do*" (Ellegård, 1953, p. 162). Nevertheless, given that the sister languages all developed the DPC, and assuming similar constraints were in place as in the sister languages, it is noteworthy that present-day English lacks the construction.

There are, of course, numerous candidates for N_1 in [N_1 (of) N_2] expressions. This investigation is confined to seven N_1 nouns: *pound*, *gallon*, *ton*, *pipe*, *sack*, *barrel* and *pair.* The first six were selected because they are standard measures or are containers used as measures (the same group as those nouns analysed by Grestenberger, 2015). The noun *pair* is not a container/measure but *pair* is particularly interesting as it occurs much more frequently as a DPC than the container/measures.

The organization of the paper is a follows: Section 2 gives some background with respect to the morphology and syntax of the DPC, and Section 3 reports on method and limitations. Section 4 analyses six of the seven nouns mentioned above with respect to both syntax and morphology, and Section 5 focuses on *pair.*

In general, little evidence is found that that the DPC construction was widespread in ME with the measure noun *pound* or container nouns, although this is not the case for *pair*.

2.   Background

This investigation is concerned only with nominals that are pseudo-partitives, and in this section I describe the characteristics of pseudo-partitives, particularly those that are container/measures. Despite some cross-linguistic micro variation, there are some general trends regarding which nouns occur only in the DPC (and not in the IPC) and regarding the presence or absence of plural morphology on N_1.

2.1   Pseudo-partitives

Partitives consist of two nominals, which are usually designated N_1 (a portion or container) and N_2 (that which is contained or portioned). N_1 measures a proportion of N_2. The distinction between true partitives and pseudo-partitives is that ordinary partitives involve restricted or contextually bound sets, whereas pseudo-partitives involve unrestricted or unbounded entities (Selkirk, 1977; Jackendoff, 1977).

(3)   a.   a bunch of those flowers/a group of our students (restricted set: partitive)
      b.   a bunch of flowers/a group of students (unrestricted set: pseudo-partitive)

Essentially, in a true partitive N_2 must be definite, sometimes known as the "Partitive Constraint" (Jackendoff, 1977); in the pseudo-partitive N_2 is either non-count or plural. Cross-linguistically, in Germanic, there are three strategies for forming pseudo-partitives:

i.   surface structure $[N_1\ N_{2GEN}]$
                       (Juxtaposition of N_1 and N_2 and special case marking)[1]
ii.   surface structure $[N_1\ N_2]$
                       (Juxtaposition of N_1 and N_2 and no special case marking)
iii. surface structure $[N_1\ of\ N_2]$
(Preposition preceding N_2)

The first strategy, using the genitive case, is found with nouns denoting "mass, measure, quantity and extent" in Old English (Mitchell 1985: 546) See also Allen (2008: 85–87).

---

1.   The case marking is genitive in Germanic but not necessarily in other languages.

(4) *lc   wifmon hæfde <u>ane</u> <u>yndsan</u> <u>goldes</u>   &   <u>an   pund   seolfres</u>.*
each woman had   one ounce gold-_{GEN} and one pound silver-_{GEN}
'Each woman had one ounce of gold and one pound of silver'   (OED: 800)

Generally, as case was lost, prepositions were introduced to compensate for the functions, the preposition *of* commonly replacing genitive case (Los, 2015: 44–46; Traugott, 2007: 532). The two other strategies for forming pseudo-partitives are either with a preposition (IPC), as in the a. examples, or without (DPC), as in the b. examples.

(5) Da.
  a. *en   spand   vand*   (pseudopartitive, DPC)
      a.COM bucket.COM water.NEU
      'a bucket of water'
  b. *en   spand   med   vand*   (pseudopartitive, IPC)
      a.COM bucket.COM with/of water.NEU
      'a bucket of/with water'
                    (Hankamer & Mikkelsen, 2008 example (29) & (61))

(6) Du.
  a. *drie   kisten sigaren*   (pseudopartitive, DPC)
      three boxes cigars
      'three boxes of cigars'
  b. *drie   kisten met   sigaren*   (pseudopartitive, IPC)
      three boxes with cigars
      'three boxes of/with cigars'
                    (van Riemsdijk, 1998, p. 15, example (18) &(19))

German uses all three strategies for pseudopartitives: DPC and IPC and also case (in (formal) modern Standard German). Whereas the caseless West Germanic and mainland Scandinavian languages use both the DPC and the IPC, modern English always requires a preposition.[2]

(7) En.
  a. a box of cigars
  b. *a box cigars

This investigation takes a detailed look at the Middle English DPC.

---

2. Icelandic also has only the IPC.

2.2 N_1 and characteristics of DPCs

As was mentioned, there are many noun candidates for N_1. Table 1 summarises the different types.

Table 1. Classification of partitive nouns

	English (En)	Dutch (Du)	Danish (Da)	German (Ge)
Quantity	*number*	*aantal, paar*	*antal, par*	*Anzahl, Paar*
Measure	*pint, pound*	*liter*	*liter, kilo*	*Liter, Kilo*
Cardinal	*dozen, million*	*milljoen*	*dusin*	*Dutzend, Millionen*
Partitive	*slice, piece*	*snee*	*stykke*	*Stück*
Container	*bottle, box*	*krat*	*kasse*	*Schachtel, Kiste*
Collective (for count nouns)	*swarm, herd*	*kudde*	*sværm klynge*	*Schwarm, Herde*
Quantums (for non-count nouns)	*lump, drop*	*druppel*	*dråbe*	*Klumpen, Tropfen*
Forms (for non-count and count nouns)	*pile, bunch*	*bos*	*bunke*	*Haufen*

The most important detail to note is that the nouns generally divide into two main groups. Quantity and measure nouns, generally the top three rows in Table 1, may be distinguished from the others (Grimshaw, 2007: 202 for English; van Riemsdjk 1998, p. 13–17 for Dutch; Delsing (1993, p. 203) for Swedish). Concerning the nouns in this investigation, *pound* is a measure, *pipe*, *sack* and *basket* are containers and *gallon* and *ton* are containers that have grammaticalized into standard measures. The cognates of English *pair* are quantity nouns in the languages other than English. The terms in the literature are varied. I use "genuine measure noun" and "quantifier noun" to distinguish standard measures like *pound* (English) and measure nouns like *aantal/antal* (Dutch & Danish/Swedish) and "temporary measure noun" for the others.

There are morphological and syntactic reasons for distinguishing these two groups. First, measure nouns like *kilo* and *litre* and quantifier nouns like *masse* 'a lot' do not occur at all in the IPC, as shown for Danish:

(8)   Da.
  a.   *et kilo (*af) smør*
       'a kilo of butter'
  b.   *en liter (*af) vand*
       'a litre (of) water'
  c.   *en masse (*af) penge/mennesker*
       'a lot of money/people'

Secondly, although there is some micro-variation, there is a general tendency in many Germanic languages, including Standard German and mainland Scandinavian (but not Dutch), for genuine measure nouns (*kilo, liter*) to lack plural marking when preceded by a numeral (Delsing, 1993, p. 204; van Riemsdijk, 1998; Hankamer & Mikkelsen, 2008, p. 318; Grestenberger, 2015).

(9)   Ge.
    a.   *zwei Kilo/*Kilos Äpfel*
        two   kilo/kilos   apples.$_{PL}$
        'two kilos of apples'
    b.   *zwei Gramm/ *Gramme Zucker*
        two   gram/grams   sugar
        'two grams of sugar'

(10) Da.
    a.   *to   snes (*snese) æg*
        two   score   eggs
        'two score of eggs'
    b.   *to fod (*fødder) elastik*   (archaic)
        two foot   elastic
        'two feet of elastic'

Therefore, if Middle English does have the DPC construction, the genuine measure noun *pound* is the one most likely to occur as a DPC and most likely to occur without plural morphology.

When it comes to the container noun group, these can occur as either IPC or DPC, as can be seen in (5) and (6) above. However, as initially pointed out by Selkirk (1977), nouns such as *cup* and *basket* often have two readings, a container reading, where the focus is on the container and where N_1 is referential, and a quantity/measure reading where N_1 is non-referential and measures the quantity contained. The two readings can be disambiguated in English by predicate selection as seen in (11a) where N_1 has a quantity reading and (11b) a container reading. In (11a), N_2 is the head and *sugar* selects the appropriate predicate with *strewn*, while *cup* is a measure term. In (11b), *cup*, is referential and goes with the predicate *smashed on the floor*.

(11)   a.   A cup of sugar was strewn on the floor.
       b.   A cup of sugar smashed on the floor. (Selkirk, 1977, example (93) & (94))

The two readings can be distinguished morphologically and syntactically, but this distinction varies according to the language. Comparison within the Germanic languages reveals that generally, for those nouns that can be used as measures, plural morphology is obligatory on N_1 regardless of whether the reading is container or

quantity. This is so for Standard German, Danish and Dutch,[3] but not for Viennese German (VGe.) where lack of plural morphology on N_1 can be used to distinguish between the two readings.[4]

As can be seen in (12) and (13), in Standard German the DPC is used for both container and measure readings, whereas in Danish, the DPC/IPC difference correlates with a quantity/container reading. (Hankamer & Mikkelsen, 2008, p. 324–5).

(12) Ge.
    a.   *drei   *Flasche$_{SING}$/ Flaschen$_{PL}$ Wasser*   (DPC: ambiguous)
        three bottle/bottles   water
        'three bottles of water'
    b.   *zwei *Korb$_{SING}$/Körbe$_{PL}$ Äpfel*   (DPC: ambiguous)
        two   basket/baskets   apples
        'two baskets of apples'

(13) Da.
    a.   *Hunden havde det godt da de kom ud fra biografen, men <u>tre poser mel</u> lå strøet ud over bagsædet.*   (DPC: measures the quantity)
        'Dog-the was OK when they came out from cinema-the, but three bags flour lay strewn out over backseat-the'.
    b.   De fleste poser i supermarkedet var lavet af plastic, men <u>tre poser med mel</u> var lavet af genbrugspapir.   (IPC: counts the bags)
        'The most bags in supermarket-the were made of plastic, but three bags with flour were made from recycled-paper'.

Viennese German (VGe.), on the other hand, like Standard German, uses the DPC for both readings but, unlike Standard German, number inflection (or lack thereof) on N_1 can be used to distinguish between the two readings (14).

(14) VGe.
    a.   *zwei Glas   Wasser*
        two   glass.SG water
        'two glasses of water'   (DPC: quantity)
    b.   *zwei Gläs-er Wasser*   (DPC: container)
        two   glass-PL water
        'two glasses of water'   (Grestenberger, 2015, p. 98, example (14))

Therefore, when it comes to the possible DPC occurrences in earlier English for the containers *gallon*, *ton*, *sack*, *barrel* and *pipe*, we could find behavior more like

---

3.   Although Dutch marginally has a plural to distinguish collective vs distributive readings.

4.   This holds for *Sack*, *Korb* and *Glas* but not for the *n*-plurals, *Tonne-n*, 'barrel', or morphologically inert nouns, *Kubel*, 'bucket'.

modern Danish, which distinguishes the two readings syntactically, with an IPC/DPC, or more like Viennese German, which distinguishes the two readings morphologically. However, the data in Section 4 will show that neither of these hypotheses can be supported.

Regarding the DPC examples in ME, Grestenberger's (2015, p. 128) claim is that ME developed a DPC "during the stage of losing the morphologically distinct post-nominal genitive and before *of* became obligatory". She gives the following examples as support.

(15) *To the same place iij sakkes lym*
      'To the same place 3 sacks (of) lime'
                               (1465, MED, Grestenberger, 2015, p. 128, example (80))

(16) *Me were leuere slepe, Than the beste galon wyn in Chepe.*
      'I would prefer sleep to the best gallon (of) wine in Cheapside'   (Chaucer,
      Canterbury Tales, late 14th century, Grestenberger, 2015, p. 128, example (79)

Hovever, as I show in Section 4.1, although DPCs occur, they are not frequent. She also claims that, unlike Viennese German, container N_1s do not omit plural morphology when referring to quantities in OE (cf. *sakkes* above) although genuine measures like *pound* lack plural forms. However, as I show in Section 4.2, it is possible to find similar sporadic examples of all the nouns without plural morphology. It seems that the plurals were absent also in earlier German, but this is not so in modern Standard German, as seen in (9). "[B]oth English and German went through a stage in which standard units of measure were inert with respect to plural marking and then apparently reintroduced plural morphology on such units" (Grestenberger, 2015, p. 130). She suggests that these changes were due to prescriptivism, although this would not explain why Standard German subsequently lost the morphology.

The grammaticalization of certain nouns in $[N_1$ (of) $N_2]$ expressions, for example *bit*, *shred*, *bunch*, *piece* and *lot*, has been examined at length (e.g. Brems, 2003, 2011, 2012; Traugott, 2007, 2008a, 2008b), although diachronic comparative issues have received little attention. In general, when these expressions grammaticalize, the change is from $[N_1\ [_{PP}\ \text{of } N_2]]$, where the prepositional phrase is a complement of the head N_1, to $[[N_1\ \text{of}]\ N_2]$ where N_2 heads the phrase and N_1 is a quantifier. There are accompanying semantic and collocation changes, e.g. *a bit(e) of an apple* vs. *a bit of a hypocrite*. (See Traugott, 2007, p. 531 for a list of criteria for distinguishing the two.) The literature has extensively debated the status of N_1 from various theoretical perspectives without any real consensus, apart from agreeing that N_1 is not always fully lexical. (See Corver, 1998; van Riemsdijk, 1998; Corver & van Riemsdijk, 2001; Schwartzchild, 2006; Alexiadou et al. 2007; Stickney 2009.)

Summarising, the consensus is that in pseudo-partitives N_1 can be referential or non-referential, where the non-referential reading correlates with a measure reading. Nouns that are genuine measures, such as *pound* and *kilo*, only have a measure reading. Genuine measure nouns are associated with reduced morphology, but this is variable depending on the language. In many Germanic languages, but not English, for both temporary quantifier nouns and genuine measure nouns, measure readings are associated with the DPC construction. In present-day Standard English, there is no morphological or syntactic way of distinguishing a referential vs a non-referential N_1, as plural morphology is required on N_1 and the preposition is always used.

3.   Data and method

Assuming that the absence of *of* indicates a DPC (Hankamer & Mikkelsen, 2008, p. 319), I initially searched the Helsinki Corpus (HC) for examples. Apart from *pound* there were few examples of the nouns in question in pseudo-partitives and even fewer pseudo-partitives preceded by a numeral greater than one. I have therefore adopted a qualitative approach, giving representative examples from the *Oxford English Dictionary* (OED), the *Middle English Dictionary* (MED), and the *Helsinki Corpus* (HC). Occasionally, additional texts are included as noted in the results. The *British National Corpus* (BNC) was used to establish whether any of the nouns occur without plural morphology in modern English. The searches found seven nouns used as N_1 in what are potentially DPC constructions: *pound*, *gallon*, *ton*, *sack*, *barrel*, *pipe* and *pair*. These nouns fall into three groups: (i) *pound*, a genuine measure; (ii) five container nouns which were used as both measures and containers at that time (*gallon*, *ton*, *barrel*, *sack*, *pipe*); (iii) a collective noun, *pair*. Particularly interesting in terms of language change are *gallon* and *ton*, which are now standard measures, no longer considered containers.

The search for DPC vs. IPC constructions is straightforward; the criterion is the absence of a preposition. When it comes to the morphology, the situation is more complex, as morphological number marking is settling down in ME after the reduction of OE inflectional morphology. However, in Section 4.2, I will attempt to predict the expectations for each noun. In the results, I first discuss *pound*, *gallon*, *ton*, *sack*, *barrel* and *pipe* and whether they occur as DPCs (4.1), then the possible occurrence of unmarked plurals (4.2). In Section 5, I discuss, the collective *pair*, which occurs quite frequently as a DPC in ME although the noun itself has followed a different grammaticalization path in English compared with the other Germanic languages.

4. Results

In the discussion that follows, it is relevant to note whether the word was used in OE or borrowed during the ME period. The use in French of the preposition *de* with pseudo-partitives might encourage preposition use in English with the borrowed words. Also, there are different expectations for plural morphology for native words vs borrowed words. The nouns *pound*, *sack* and *ton* were used in OE, whereas *pipe* (1348) *barrel* (1300) and *gallon* (1225) are borrowed from French. These dates are the first citation in the OED.

4.1 Occurrences of the DPC

As mentioned earlier, *pound* is the noun most likely to occur in DPC constructions in earlier English, since genuine measures do not occur in IPCs in the other Germanic languages. However, although *pound* without a following *of* may be found in the HC, there are only three examples, as compared with 17 examples with *of*. Additionally, all three are from the same document, the statutes of the Carpenter's Guild (17).

(17)   *& who so is absent at þilk masses… schal paie to þe brotherede <u>a pound wex</u>.*
'And whosoever is absent at those masses…shall pay to the brotherhood a pound (of) wax'   (HC: 1389)

The only examples I was able to find without *of* after1400 are in lists and recipes, a usage still found in present-day English with various measures (18).

(18)   <u>*Six pund musk almonds*</u> *at 12s. per pund*.   (OED: 1675)

The conclusion for *pound* is that the DPC is not prevalent and is short-lived.

Five of the remaining N_1s found in DPC constructions *gallon*, *barrel*, *sack*, *ton* and *pipe* are containers used as measures in ME. Although all the container nouns have the potential to grammaticalize into standard measures, *gallon* and *ton* have done so to the extent that they have lost their container reading altogether. For this reason they are the next most likely group after *pound* to occur as a DPC. However, I found only one DPC example of *ton*, as in (19), a measure, but also from a list, as with (18).

(19)   *ij <u>ton tymby</u>*
'two ton (of) timber'   (MED: 1423)

The noun *tunne* in OE was generally used to mean a container, the usage also shown in (20). Earlier, the two spellings *ton(ne)* and *tun(ne)* are used somewhat interchangeably.

(20)   *He hadde a vyneȝerd, þe whiche . . bare hym x tunne of wyn.*
      he had   a vineyard the which . . bore him ten tun(s) of wine.
      'He had a vineyard which produced ten tuns of wine'   (OED: 1450)

In the late ME period there is a split between the container and the measure, with the spelling *tun* used for the container, and *ton* for the measure (volume or weight). One of the issues that these data reveal, particularly with respect to *tun/ton*, is a third possible reading, when the preposition *of* denotes that the material and not the contents are meant, as in (21).

(21)   *Heo makeden ane tunne of golde. Þene king heo duden þer-inne.*
      'She made   a   cask of gold. The king she put   therin'.
                                         (MED: 1275)

Although this is often disambiguated by world knowledge, there seems a tendency to specify the container reading with *full*, as in (22).

(22)   *Jupiter..Hath in his celier..Tuo tonnes fulle of love drinke.*
      'Jupiter has in his cellar two casks full of love drink'.   (MED: 1393)

For *gallon*, both readings are also available. In (23) *galoun* glosses the Latin word *lagenam* 'bottle' indicating a container and (24) shows a measure. The container reading is last cited in 1459.

(23)   *And another berynge a galoun of wyn*
      'And another bearing a gallon [Latin: lagenam. 'bottle'] of wine.'
                                         (MED: 1425)

(24)   *Þere were at þe feste sixe water pottis sett, and ech of hem held a galoun or more.*
      'There were at the feast six water pots set, and each of them held a gallon or
      more'   (MED: 1425)

As with *ton*, DPC examples are difficult to find, although in (25), from *Piers Plowman*, alternative manuscript readings with and without *of* are given showing both were an option.

(25)   *And greten Sir gloten with a galun of ale*
      And greten Sir gloten with a galun ale [B.ms.]
      'And greet Sir Glutton with a gallon of ale'   (OED: 1390)

Example (16) above, from Grestenberger (2015: 128), is one of the rare examples of DPC with *gallon*.

The semantic split of *ton/tun* and the adoption of *gallon* and *ton* as a standard measure are interesting developments, and merit further analysis, beyond the scope of this investgation. As far as the DPC is concerned, there is little evidence that it was prevalent with *pound*, *ton*, and *gallon*.

The three remaining container nouns, *sack*, *pipe*, and *barrel*[5] are not used as standardized measures in modern English, but were everyday containers and measures in ME. In the ME section of the HC there is only one unremarkable example of *sack* as an IPC.

(26) *that the seid Beseker may haue <u>xx sakes of wolle</u>*
      'that the said Beseker may have 20 sackes of wool'.   (HC: 1437)

However, the DPC can be found occasionally in inventory lists seen in (27), as well as in (15).

(27) *For <u>iij sak lyme</u> to þe same mason*
      'For 3 sacks (of) lime to the same mason'   (MED: 1427–8)

Similarly, DPC examples with *barrel* and *pipe* are sparse, although they do occur with *barrel*. As mentioned above, French uses the preposition *de*, as shown in (28), so we might expect French usage to inhibit the DPC, especially if the construction as a whole (*barells de picch*) has had an influence. However, this does not prevent scattered early examples of *barrel* from occurring without the preposition as in (29) and (30). This fact could indicate that the single lexical item is borrowed, not the construction.

(28) <u>*xxii barells de picch, xxiii barrell de tarre.*</u>
      '12 barrels of pitch, 13 barrels of tar'.   (OED: 1338)

(29) *I hadde leuere than <u>a barel ale</u>*
      That Goodelief my wyf hadde herd this tale.
      'I would prefer my wife hearing this tale than to have a barrel of ale'.
                                                              (OED: 1390)

(30) *Item <u>ix barellis beere</u> val. xx s*
      'Item ten barrels beer value 20 shillings'   (OED: 1396)

Note again that (28) and (30) are from lists as is (32). Finally, like *barrel*, *pipe* is a container used as a liquid measure, again borrowed from French and found in

---

5.   Although *barrel* is still used for oil and alcoholic liquids, it is not an everyday measure.

straightforward IPC examples such as (31); it should be noted, however, that I did find a late DPC example as in (32).

(31)   *Y wyt to Iohan Whyte the yongger, & to hys wyfe <u>a pipe of wyne,</u> prys of xl s*
      'I will to JW the younger and to his wife a pipe of wine, price of 40 s'.

<div align="right">(MED: 1406)</div>

(32)   *Ther was gevon to the Duke of Gloucestre..<u>iiij pype wyn</u>*
      'There was given to the Duke of Gloucester 3 pipes (of) wine.   (OED:1525)

It can be concluded that the DPC does not occur frequently, at least with containers used as measures or, where it is expected most in comparison to other Germanic languages, with the genuine measure, *pound.* Limited examples are found with the surface structure of a DPC, particularly in lists, but such usages are not confined to ME (cf. Abney, 1987, p. 290). Although we might expect omission of *of* to be discouraged with French origin words, DPC examples with *barrel*, *pipe* and *gallon* are found, indicating that the individual lexical item was borrowed, not the construction itself.

4.2   Singular/plural morphology

In this section, I focus on the presence or absence of plural morphology, and examine whether English behaves like its sister languages in associating unmarked plurals with the DPC. First, the expected plural morphology for each noun is assessed. As pointed out by Mustanoja (2016: 57), identical singular and plural forms are expected in ME for those nouns that had an unchanged plural in OE and for nouns ending in *-u*, *-e* and *-an*. This is because final vowels are much less salient than sibilants and more easily lost. For those nouns found in OE, *pound*, *ton* and *sack*, unmarked plurals are expected for *pound* and *ton*. As a neuter noun in OE, *pound* had an unchanged neuter nominative and accusative plural and a genitive plural in *-a*, and it is not difficult to find examples of the unmarked plural:

(33)   *After for <u>a þousend pound of golde,</u> Þe prest no more synne wolde.*
      'For a thousand pounds of gold the priest no more would sin'   (OED: 1390)

However, scattered examples of singular *pound* with a numeral can be found in later centuries. I found 17 modern examples in the BNC, all spoken, as in (34).

(34)   There's <u>three pound of sugar</u> to every <u>two pound of fruit</u> to make jam.

<div align="right">(BNC: spoken)</div>

Similarly, plurals of *ton* without *-s* are not difficult to find. But, as with *pound*, this outcome is expected, considering that *tunne* is a weak feminine noun in OE, with

syncretism of accusative singular and plural as -*an*. Phonological reduction of the -*an* affix to -*e* occurs in later ME as seen in (36). Therefore, *tunnan* and *tune*, as seen in (35) and (36) (= (20)), are both ambiguous between singular and plural.

(35) *He scolde gife ilca gear into þe minstre…twa tunnan fulle hlutres aloð*
      'He should give each year to the minster, two tuns full (of) pure aloe'

      (OED: 1121)

(36) *He hadde a vyneȝerd, þe whiche…bare hym x tunne of wyn.*
      'He had a vineyard the which bore him ten tun(s) of wine'.   (OED: 1450)

Nevertheless, examples with an-*s* plural (37) and without, (38), (cf. also (19)) are also found.

(37) *Jupiter..Hath in his celier..Tuo tonnes fulle of love drinke*
      'Jupiter has in his cellar two casks full of love drink'   (MED: 1393)

(38) *take the a yerely fee..ii ton of wyne*
      'take you a yearly fee. . 2 tun(s) of wine   (OED:1475)

However, as with *pound*, unmarked plurals occur colloquially in present-day English (39), here in a true partitive (N_2 is definite):

(39) Put ten ton of that stuff down you!   (BNC: spoken)

Conversely, when it comes to *sack*, a strong masculine noun in OE, a plural in -*s* is expected, as seen in (15) and (26) above. However, one or two unmarked plurals may be found in the dictionaries as in (27) repeated here as (40).

(40) *For iij sak lyme to þe same mason*
      'For 3 sack(s) (of) lime to the same mason'   (MED: 1427–8)

There are different expectations for *barrel*, *pipe*, and *gallon*, borrowed from French. These would adopt the plural in -*s*. Examples in the *Anglo-Norman Dictionary* (AND-online) show that all three take an -*s* plural in Anglo-Norman also, and this is seen in (28) for *barrel*, although note (30) where *barrel* occurs both with and without a plural. Occasional unmarked plurals of *pipe* occur following numerals, as in (32). For *gallon* I could find no pseudo-partitive examples with numerals greater than one, although the -*s* plural can be seen in (41).

(41) *Oyle to ye lompe iii galonys and half a pynt*   (MED: 1442)
      'Oil to the lamp 3 gallons and half a pint'

As with *ton*, present-day examples with unmarked plurals of *gallon* may be found colloquially in the BNC:

(42)   We've er, just give <u>two gallon</u> away.   (BNC: spoken)

In this section, it has been argued that unmarked plurals of *pound* and *ton* are expected in ME as a result of their OE morphology, but *-s* is expected for *sack*. French-borrowed words would be expected to adopt the *-s* (although there could be a period of uncertainly). However, it is possible to find examples of all six without plural morphology. The conclusion that can be drawn from this is, first, that there is little or no correlation between lack of plural and the DPC/IPC difference. Secondly, although the OED entry for *pound* says that "the unchanged plural was long retained following a cardinal number.… and [is] still common in colloquial and regional English", the presence of singular *gallon* in examples like (42) and *pipe* in (32) shows that it is not "retained" from OE, but that it is characteristic of measure terms regardless of their earlier history.

5.   A collective: *Pair*

So far in this investigation of possible DPC constructions in English, little evidence has been found that the measure nouns occur frequently as DPCs. However, in this section I show that *pair* occurs as a DPC quite frequently in ME. On the surface, *pair* has a similar pseudo-partitive form to the container nouns, [N$_1$ (*of*) N$_2$] although, unlike the container nouns, N$_2$ is limited to count nouns. As has been shown in Section 2, certain measure nouns can occur only in the DPC, and this includes cognates of *pair*, e.g. Dutch *paar* and Danish *par*. As seen in (43), *par* behaves like a genuine measure noun, taking singular morphology only, like *liter* and *masse*. Plural morphology on N$_1$ and a linking preposition are both ungrammatical.

(43)   Da.
     a.   *et par bønner*
        a   pair bean-PL
        'a pair/a few beans'
     b.   *tre   par bønner*
        three pair bean-PL
        'three pairs of beans'
     c.   **et par af bønner*
     d.   **tre par af bønner*

In addition, as seen in (43) and (45a), in Germanic languages other than English, *pair* can mean 'few' as well as 'two of a kind'.

(44) Da.

så lod hun hænderne glide søgende gennem de sidste par jordbærplanter

'as let she (her) hands glide, searching through the last few strawberry plants'

<div align="right">(Korpusdk; Familie-Journalen)</div>

This *pair/few* ambiguity can be disambiguated in e.g. Dutch with verbal agreement.

(45) Du.

    a.   *Er staan een paar schoenen op de tafel*   (quantification reading)

        'there stand.PL a few shoes on the table'

    b.   *Er staat een paar schoenen op de tafel*   (collective reading)

        'there stands.SG a pair (of) shoes on the table'   (van Riemsdijk, 1998: 17)

Note that when *par* is used with a numeral, the meaning 'two' is forced, just as the lexical meaning can be forced with English *lot*: (*lots of trailers* = many; *three lots of trailers* = 3 groups).

In modern English, *pair* is restricted to meanings associated with 'two' or with 'sets'. It is usually followed by one of two types of noun, certain *plurale tantum* nouns such as *jeans*, *pants*, as well as plural count nouns that have a morphological singular but often come in sets of two, such as *eyes*, or *shoes*. Historically, however, English *pair* could be used without a preposition and tended to have an unmarked plural – both features associated with DPCs. What is more, it could also mean 'a few'.

5.1 *Pair*: Morphology and syntax

Pair is borrowed into English through Anglo-Norman around 1300 with the meaning: 'A couple; a set of two' (OED). In the sister languages it comes directly from Latin plural *paria*, (Falk & Torp, 1911: 815) into Middle Low German and it is also in Icelandic (*pár*), Swedish (*par*), Danish (*par*) and German (*Paar*). As was mentionned, French uses the preposition *de* as in the Anglo-Norman example, (46), which might discourage the DPC in the earliest borrowings. This seems not to be the case, as the early entries (first mention in 1300) tend to have no preposition as in (47) and (48):

(46) *Deux peir de plates coverez de velvet noir.*

    'Two pair of plates covered with velvet black'   (MED: 1354)

(47) *Is chamberlein him bro3te. . a peire hosen.*

    'His chamberlain him brought a pair hose'   (OED: 1325)

(48) *Euere he hadde ane peire feteres faste him up-on.*

    'Always he had a pair fetters fastened him upon'   (OED: c1300)

The use without a preposition is fairly frequent, with more entries in the MED without than with, continuing up until the beginning of the 16th century and beyond, as illustrated in (49)–(52).

(49) *From þe forparte of the hede in whiche bene vij payr sennes comythe <u>a payr senewys</u> out of the brayn.*
      'From the front part of the head in which are 7 pair (of) sinews comes a pair (of) sinews out of the brain'   (MED: 1500)

(50) <u>*Tua pair taingis*</u> *with purringe ironis*
      'Two pair tongs with pokers'   (OED: 1627)

(51) <u>Two pair black satin slippers</u>   (OED: 1826)

(52) To equip one soldier, – item, – <u>two pair shoes</u>, five dollars each
                                                                              (COHA: 1929)

The Helsinki Corpus is too limited to have a large number of examples but the *Paston Letters* (15th century) have 41 DPC examples, as in (53) and (54), and only 8 of the IPC, as in (55).

(53) *I bequethe to Marie Tendall, my goddoughter, <u>my peir bedys</u> of calcidenys*
      'I bequeath to Marie Tendall, my goddaughter, my pair beds of chalcedony'
                                                  (1482: copy of will of Margaret Paston)

(54) *Item, he hathe <u>a peyre botys</u> off Edmond Reedys þe shomakere.*
      'Item he has a pair (of) boots off ER the shomaker'
                                                     (1474: John II to Margaret Paston)

(55) *<u>and iij payer of newe shetys</u>*
      'and three pair of new sheets'
                                     (1487: will of Elizabeth Poynings or Browne, née Paston)

Such constructions continue in lists and colloquial usage today, as in (56) and (57).

(56) My finishing-school clothes list: <u>One pair plimsolls</u>   (BNC: written)

(57) She went, he's pretty good with <u>a pair clippers</u>, ain't he, old Victor.
                                                                              (BNC: spoken)

As for the morphology, as a noun borrowed from French, *pair* has an -*s* plural, as it does in the Anglo-Norman dictionary, although there are dictionary entries where it is not inflected (58).

(58) *deux peire de trestels*   (AND: s.v. pair)
       two   pair of trestles

This example is evidence that reduced morphology is likely to be affected by the proximity of a numeral and not the DPC. While considered 'non-standard', singular *pair* is still quite frequent, at least in colloquial speech, and may be found in representative examples in all centuries since its first appearance, as in (59)–(61).

(59)   Behind my saddle I carried <u>three pair</u> of camping blankets and a quilt.
                                                              (COHA: 1879, non-fiction)

(60)   At every levee, appeared <u>eighteen or twenty pair of lawn sleeves</u>.
                                                      (OED: 1887 Macaulay *Earl of Chatham*)

(61)   I mean I bought <u>one or two pair of clogs</u> when I was on the dredger
                                                              (BNC: 1987, oral history)

Of all the nouns covered in this investigation, *pair* is the only one that occurs frequently in the DPC construction and also regularly occurs without a plural. In this way it is the most similar syntactically and morphologically to the quantificational nouns in the sister languages.

5.2   Semantic development of *pair*

In French, *pair* meant 'a set'. The OED gives *paire d'armes*, 'set of military equipment'; *paire de cartes*, 'pack of cards'; *paire d'habits* 'outfit of clothing' often meaning 'a set of two', but not necessarily, as in (62):

(62)   *Item, <u>a pair of keruynge knyues, iiij in a sheth</u>*
        'Item a pair (set) of carving knives, 4 in a sheath'   (OED: 1415)

One of the obsolete meanings, first attested around 1400 in the OED (s.v. *pair*), is similar to that found in the other Germanic languages: "a few, two or three". Interesting evidence of this comes from two manuscripts of *Cursor Mundi*, one of which uses the phrase 'a pair of letters', the other 'some letters', showing that the scribes considered the phrases equivalent:

(63)   a.   *Þe king a pair o letters writte Did*   (*Vespasian* ms)
        b.   *Þe king did sone lettris to write*   (*Göttingen* ms)
             'The king had some letters written'   (MED: *Cursor Mundi* c1400)

Other examples may be found in dictionaries and literature. In order for reanalysis (from a collective meaning 'two' to a measure term) to take place, we would expect to find potentially ambiguous bridging contexts. This is seen in (64) and (65) where the meaning could be 'two' or 'a few'. *Years* and *minutes* do not typically come in sets of two like *eyes*, although they could. Notice also, the addition of the terms *or so* and *thereabout* giving the sense of 'approximation'.

(64) EModE What ist to bide A little hardness for <u>a pair of years</u> or so.

<div align="right">(OED: 1627)</div>

(65) EModE and what's a sonne? A thing begot within <u>a paire of minutes</u>, there about.

<div align="right">(OED: 1602)</div>

Temporal expressions, difficult to measure with precision, are contexts conducive for semantic change. The smaller the unit of time, the more difficult it is to estimate exactly, making *minutes* more felicitous than *hours* or *days*.

(66) To entertaine you for <u>a payre of howres</u>.

<div align="right">(OED: 1631)</div>

(67) Some say they will wait <u>a pair of days more.</u>

<div align="right">(OED: 1889 'M. Twain' *Connecticut Yankee*)</div>

Prompted by example (68) from James Shirley's play *Wedding*, I searched other plays and found several similar examples, such as (69) and (70):

(68) I may be compeld within <u>A pair of minutes</u> to turn ashes

<div align="right">(OED: 1629)</div>

(69) Since I arriu'd, Tis but <u>a paire of minutes</u>

<div align="right">(1640: James Shirley *The Gratefull Servant*)</div>

(70) Who would trust woman? Lost in <u>a paire of minutes</u>

<div align="right">(1640? James Shirley *The Coronation: a comedy*)</div>

These data show the collocation restrictions for *pair* loosening so that N_2 no longer has to be a set of two. However, as far as the data shown here are concerned, the change appears to have stopped. The type of evidence that would show a further stage would be, for example, more frequent use of abstract N_2.

6. Conclusion

This investigation has found little evidence that the DPC construction was established to any significant extent in ME with measure and container nouns. Although the measure noun *pound* is the one expected to occur most frequently without a preposition, this was not the case. It is notable that, when expressions like "sacks of x" occur, they tend to be in inventories and lists, which may leave out the preposition as shorthand. Although we might expect the French *de* to promote preposition use, even in lists of recipe ingredients in present-day English, e.g. *3 cups flour*, the preposition is often omitted. The data suggest that the English DPC is a register specific construction that is present in all historical periods.

When it comes to morphology, the data are more solid, with a tendency for measures to occur as unmarked plurals when following a numeral. However, there is no correlation with the DPC pseudo-partitive. It is more likely to be a feature

of measures adjacent to numerals. Example (58) in particular provides evidence. Since the numeral already shows the nominal is plural, inflectional morphology is reduntant. This is so in colloquial modern English as well.

The collective *pair* is unlike the other nouns investigated. Morphologically, syntactively, and semantically, it appears to have started to change in the same direction as the other Germanic languages. It is not immediately obvious why *pair* did not go on to develop the meaning 'approximately' in English. It appears that it did briefly mean, 'a few, two or three' in Early Modern English, but only in a very limited way. A language-internal explanation may perhaps be found in the etymology (Falk & Torp, 1911). In Germanic languages other than English, the word was borrowed directly from Latin from the adjective use of *par* meaning 'same, equal' (cf. English *par*, *parity*). English *pair* came through French, from Latin plural *paria* and so it was borrowed into English already with slightly different semantic features. Supporting this intriguing possibility would require further research.

This investigation has focused on only a limited group of nouns. The full inventory of possible DPCs, e.g. those like *a morsel bread* mentioned in the introduction, should be investigated in detail. Further research of a quantitive nature as well as more comparative work with Anglo-Norman should yield significant results. Yet, this qualitative analysis was able to reveal several important findings: English, for the most part, does not follow the patterns found in other Germanic languages for standard measures and containers, whereas the noun *pair*, with its unique history, shows several syntactic, morphological and semantic similarities.

Acknowledgements

Sources

British National Corpus (BNC) http://corpus.byu.edu/bnc
Corpus of Contemporary American English (COCA) http://corpus.byu.edu/coca
Corpus of Historical American English (COHA) http://corpus.byu.edu/coha
Davis, N. (1971) (Ed.), *Paston letters and papers of the fifteenth century*. Oxford: Oxford UP.
KorpusDK http://ordnet.dk/korpusdk
Ordbog over det danske sprog (ODS, 28 vols., 1919–1956). Copenhagen: Gyldendal. http://ordnet.dk/ods/
Oxford English Dictionary (OED) on-line www.oed.com
Helsinki Corpus of Middle English texts
Middle English Dictionary http://quod.lib.umich.edu/m/med/

References

Abney, St. P. (1987). *The English noun phrase in its sentential aspect*. Cambridge, MA: MIT dissertation.

Anglo-Norman Dictionary, AND² Online edition. <http://www.anglo-norman.net/>; Accessed 25 Sep 2018

Alexiadou, A., Haegeman, L. & Stavrou, M. (2007). *Noun phrase in the generative perspective*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110207491

Allen, C. (2008). *Genitives in early English*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199216680.001.0001

Brems, L. (2003). Measure noun constructions: An instance of semantically-driven grammaticalization. *International Journal of Corpus Linguistics* 8.2: 283–312. https://doi.org/10.1075/ijcl.8.2.05bre

Brems, L. (2011). *Layering of size and type noun constructions in English*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110252927

Brems, L. (2012). The establishment of quantifier constructions for size nouns: A diachronic study of *heap(s)* and *lot(s)*. *Journal of Historical Pragmatics* 13: 202–231. https://doi.org/10.1075/jhp.13.2.02bre

Corver, N. (1989). Left branch extractions and DP. In Bennis, H. & van Kemenade, A. (Eds.). *Linguistics in the Netherlands*. (pp. 31–40). Dordrecht: Foris.

Corver, N. & van Riemsdijk, H. (2001). *Semi-lexical categories: The function of content words and the content of function words*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110874006

Delsing, L. (1993). The internal structure of noun phrases in the Scandinavian languages. A comparative study. University of Lund dissertation.

Ellegård, A. (1953). *The auxiliary 'do': The establishment and regulation of its use in English*. Goeteborg, Sweden: Elanders Boktryckeri Aktiebolag.

Falk, H. & Torp, A. (1911). *Etymologisk ordbog over det norske og det danske sprog* [Etymological dictionary of the Norwegian and Danish languages]. Heidelberg: Carl Winter's Universitätsbuchhandlung.

Grestenberger, L. (2015). Number marking in German measure phrases and the structure of pseudopartitives. *Journal of Comparative Germanic linguistics*. 18, 93–138. https://doi.org/10.1007/s10828-015-9074-1

Grimshaw, J. (2007). Boxes and piles and what's in them: Two extended projections or one? In A. Zaenen, J. Grimshaw, J. Maling, C. Manning and J. Simpson (Eds), *Architectures, rules and preferences: variations on themes by Joan W. Bresnan* (pp. 245–252). Center for the Study of Language and Information, Stanford, California.

Hankamer, J. & Mikkelsen, L. (2008). Definiteness marking and the structure of Danish pseudo-partitives. *Journal of Linguistics*. 44, 317–346. https://doi.org/10.1017/S0022226708005148

Harbert, W. (2007). *The Germanic languages*. Cambridge: Cambridge University Press.

Jackendoff, R. (1977). *X-bar syntax: A study of phrase structure*. Cambridge, Mass: MIT Press.

Lass, R. (1987). *The shape of English*. London: Dent.

Los, B. (2015). *An historical syntax of English*. Edinburgh: EUP.

McWhorter, J. (2002). What happened to English? *Diachronica* 19, 217–272. https://doi.org/10.1075/dia.19.2.02wha

Mitchell, B. 1985. *Old English syntax*, 2 vols. Oxford: Clarendon

Mustanoja, T. (2016). *A Middle English syntax*. Amsterdam: Benjamins. (Originally published 1960). https://doi.org/10.1075/z.207

Riemsdijk, H. van. (1998). Categorial feature magnetism: The endrocentricity and distribution of projections. *Journal of Comparative Germanic linguistics*. 1, 1–48. https://doi.org/10.1023/A:1009763305416

Schwarzschild, R. (2006). The role of dimensions in the syntax of noun phrases. *Syntax*. 9.1, 67–110. https://doi.org/10.1111/j.1467-9612.2006.00083.x

Selkirk, E. (1977). Some remarks on noun phrase structure. in Culicover, P. et al. (Eds.). *Formal syntax*. New York: Academic Press.

Stickney, H. (2009). The emergence of DP in the Partitive Structure. University of Massachusets dissertations. http://scholarworks.umass.edu/open_access_dissertations/

Toyota, J. (2009). When mass was counted: English as a classifier and non-classifier language. *SKASE Journal of Theoretical Linguistics* [online]. 6.1. <http://www.skase.sk/Volumes/JTL13/pdf_doc/07.pdf>.

Traugott, E. (2007). The concepts of constructional mismatch and type-shifting from the perspectiveof grammaticalization. *Cognitive Linguistics* 18–4, 523–557.

Traugott, E. (2008a). Grammaticalization, constructions and the incremental development of language: Suggestions from the development of Degree Modifiers in English. In R. Eckardt, J. Gerhard & T. Veenstra (Eds) *Variation, selection, development: Probing the evolutionary model of language change*. (221–250). Mouton de Gruyter.

Traugott, E. (2008b). The grammaticalization of NP of NP construction. In A. Bergs & G. Diewald (Eds), *Constructions and Language Change* Mouton de Gruyter.

Weinreich, U., Labov, W. & Herzog, M. (1968) "Empirical foundations for a Theory of language change". In W. P. Lehmann & J. Malkiel (Eds.) *Directions for historical linguistics: A symposium*. Austin & London: University of Texas Press, 96–195.

Grammaticalization & construction grammar

Old French sɪ, grammaticalisation, and the interconnectedness of change

Sam Wolfe

University of Oxford

The particle sɪ is ubiquitous across the early French textual record, yet receives no uniform analysis, with numerous competing and often contradictory claims in the literature about its distribution and formal status. This study draws on a novel diachronic corpus analysis to put forward an original analysis of sɪ, under which it is not the homogeneous entity often assumed but rather an adverbial which grammaticalises as a topic-continuity marker and, later, two forms of V2-related expletive. sɪ loses its previously specific temporal and discourse-pragmatic meaning, shows a widening of distribution and occupies an increasingly high position within the left periphery. In these respects it is shown to instantiate a classic form of upwards grammaticalisation pathway.

Keywords: left periphery, Old French, V2, Topic, grammaticalisation, reanalysis, historical pragmatics

1. Introduction and background

1.1 Old French and syntactic change

Scholars of all persuasions acknowledge that French and other Romance varieties undergo a raft of significant syntactic changes from the earliest attestations in the 9th century throughout the medieval period (Harris, 1976, 1978; Bauer, 1995, Chapters 4, 5; Ledgeway, 2011b, pp. 389–433, 2012; Schwegler, 2013, Chapters 3, 4). A notable proportion of these changes have been dealt with as instances of grammaticalisation phenomena, which are argued to contribute to the large-scale typological changes taking place during the transition from Latin to Romance (Fleischman, 1982; Hopper & Traugott, 2003, pp. 55–64; Klausenburger, 2000; Ledgeway, 2011a; Roberts, 2012; Roberts & Roussou, 2002).

The observation by Malkiel (1988, p. 19, cited in Sornicola, 2011, p. 3) that Romance linguists have a "reservoir of priceless data" is particularly pertinent to the study of syntactic changes of these types, despite the non-trivial limitations of the Old French textual data (on which see Fleischman, 2000). In particular, given the relatively wide sample of primary texts available and ever-growing body of secondary literature, Romance data, in general, and French data, in particular, provide a fitting domain to investigate the interrelatedness of change.

In this chapter I focus on the changes affecting what seems, at first glance, to be a very small area of the syntax of Old French: the particle sɪ, which derives from Latin sɪc 'thus'. I argue firstly for a reappraisal of the available evidence, so that competing claims in the literature regarding the syntax of sɪ can, in certain ways, be reconciled if we acknowledge that different Old French texts are indicative of different stages of a grammaticalisation pathway for sɪ. Secondly, I argue that the changes affecting sɪ are a case-study of a grammaticalisation phenomenon, where each stage of the process of changes is both connected to, and dependent on, other changes in the clausal structure of French and Romance occurring during the medieval period. The grammaticalisation cline is argued to be a clear instance of upwards grammaticalisation throughout the clausal spine in the terms of Roberts & Roussou (2002).

1.2   Background on Old French

There are several characteristics of Old French grammar which are necessary background for much of the discussion that follows. Firstly, as noted since Thurneysen (1892), Old French went through a Verb Second (V2) stage (Skårup, 1975, p. 290; Harris, 1984, pp. 189–198; Adams, 1987, pp. 2–3; Roberts, 1993, Chapter 3; Vance, 1997; Vance, Donaldson, & Steiner, 2009, pp. 302–307; Steiner, 2014). Since the seminal work of Den Besten (1983), the V2 constraint has been understood in formal syntactic work as a syntactic constraint that requires (a) that a constituent be moved/merged in the left periphery of the clause and (b) that the finite verb move to the head position normally occupied by the complementiser, C°. Two of the most notable effects of the V2 constraint throughout the history of French are the occurrence of so-called Germanic-inversion, where postverbal subjects occur sandwiched between the finite auxiliary and past participles, infinitives or VP-adverbs (Adams, 1987, p. 4; Roberts, 1993, Section 2.2; Vance, 1997, pp. 78–79) (1), and the fact the preverbal field can host a far wider variety of constituents than is possible in the modern language (Foulet, 1928, pp. 306–322; Skårup, 1975, pp. 9–69; Vanelli, Renzi, & Benincà, 1986, Section 4.1; Mathieu, 2012, p. 327) (2):

(1) *Par tantes   teres **ad**   <u>sun</u> <u>cors</u> traveillet*
     over so-many lands have.3SG his body suffer.PTCP
     'His body has suffered across so many lands'   (*Roland*, 540)

(2) *Ce **oïrent**   el   pales <u>maint</u>*
     that hear.3PL.PST in-the palace many
     'Many in the palace heard this'   (*Charrette*, 80)

Secondly, in contrast to contemporary French, Old French is a form of Null Subject Language. Generally speaking, until approximately 1200, null subjects are licensed both in main clauses, yielding verb-initial orders, and embedded clauses. However, around the turn of the 13th century the null argument system changes such that verb-initial orders such as (3) become very rare (Skårup, 1975, p. 291), and null subjects in embedded clauses are far more restricted (Adams, 1987, p. 3; Roberts, 1993, p. 139; Vance, 1997, Chapter 5).

(3) ***Getet***   le   a terre
     throw.3SG.PST= it.CL to ground
     'He throws it to the ground'   (*Roland* 464)

Thirdly, consider the syntax-pragmatics mapping. Data such as (1) suggest that a wide range of constituents can be moved into the Old French left periphery (see in particular Mathieu, 2012 and Salvesen, 2013). There are two particularly important differences in this domain which differentiate Old French from the modern language. Firstly, as has long been noted, Old French does not require a resumptive clitic within the clausal core when topicalisation obtains, in contrast to contemporary French varieties, where this is strongly preferred (Rowlett, 2007, pp. 172–188). Secondly, Early Old French texts, generally within the 9th–12th centuries, license preverbal new Information Focus (cf. Cruschina, 2006), where a constituent encoding brand new information occurs in the preverbal field. Preverbal placement for a new Information Focus is ruled out in Standard Modern French (Cruschina & Ledgeway, 2016, section 31.2), in contrast to the 12th-century example in (4):

(4) *Plein **est**   de figure é   signefiance*
     clear be.3SG of figure and meaning
     'This is clear with respect to form and meaning'   (*QLR* I, 4)

2. Previous approaches to the syntax of sɪ

Studies on the syntax and interpretative properties of sɪ represent a truly vast literature.[1] There are three broad types of accounts that have been advanced, all of which have certain strengths and weaknesses.

Many traditional treatments of Old French treat sɪ as a simple adverb (Foulet, 1928, pp. 301–303; Skårup, 1975, pp. 238–239; Jensen, 1990, pp. 472–473). Whilst sɪ itself derives from a Latin temporal deictic adverb sɪc (Salvi, 2004, p. 15; Ledgeway, In Press a) and patterns with Old French adverbs in triggering verb-subject inversion (5) (Foulet, 1928, pp. 301–303), this account has certain shortcomings. Specifically, classing sɪ as a simple adverbial fails to account for its sheer abundance across the Old French textual records. Furthermore, it is not clear why an adverb sɪ roughly meaning 'thus, so' would need to occur near-systematically after an initial clause in *La Queste* (6), where it does so 177 times. Finally, even more surprising under such an analysis would be the cases such as (7), discussed by Fleischman (1991, p. 263). Here, other adverbs encoding the very values of temporal continuity or succession which sɪ itself is meant to encode co-occur with sɪ (see also Schøsler & Van Reenen, 1996, pp. 640–641).

(5) *Si **luisoient** ja les estoiles*
sɪ shine.3ᴘʟ.ᴘsᴛ already the stars
'The stars were already shining' (*Charrette* 31a)

(6) *Quant il vint en la valee, si **comença** a penser…*
when he come.3sɢ.ᴘsᴛ in the valley sɪ begin.3sɢ.ᴘsᴛ to think.ɪɴf
'When he came to the valley, he began to think….' (*Graal*, 149d, 8)

(7) *En la cite puis si l'en **meinent***
in the city then sɪ him.cʟ-ʟoc.cʟ= lead.3ᴘʟ
'They then lead him into the city' (*Thèbes1*, 384)

An alternative account, which has certain properties in common with the adverbial analysis, views sɪ as a particle encoding Topic-continuity (Diez, 1882, p. 2060; Reenen & Schøsler, 2000, p. 84; Buridant, 2000, p. 508). This analysis, set out in the most convincing detail by Fleischman (1991), accounts for a wide range of the data. Consider the example in (8) here where sɪ acts as a form of resumption device for the subject *Li vaslés* 'the vassal':

---

1. For detailed overview of the literature and relevant discussion and analysis, see Marchello-Nizia (1985), Fleischman (1991, 1992) and Ledgeway (2008).

(8) _Li   vaslés_ **entendi**   bien que li   empereres li
     the vassal understand.3SG.PST well that the emperor him.CL=
     donnoit   _boin consel; si s'atorna_   _au   plus_
     give.3SG.PST good counsel SI REFL.CL=prepare.3SG.PST at-the most
     belement qu'il   peut,   si s'en   **vint**…
     best   that-he can.3SG SI REFL.CL=LOC.CL= come.3SG.PST
     'The servant understood clearly that the emperor had given him good advice.
     He prepared the best he could and came…'   (_Clari_ 30, 31)

Further evidence that SI can encode Topic-continuity comes from a context where
SI is seemingly absent across the Old French textual records. There are no reported
cases where SI occurs at the beginning of a portion of text sharing the same theme or
Topic (Marchello-Nizia, 1985, p. 25; Fleischman, 1991, p. 256; Lemieux & Dupuis,
1995, p. 96; Reenen & Schøsler, 2000, p. 86).

   However, data are relatively widespread which do not fit so neatly with an
account of this kind. Consider (9), where SI co-occurs with thematic nominal ex-
pressions, _souffrance_ 'suffering', _ceste ville_ 'this town', which are readily identifiable
from the discourse and in no need of a disambiguating particle encoding Topic-
continuity. Equally problematic are cases such as (10), where SI co-occurs with an
overt postverbal subject which is neither active (Prince, 1981, p. 243) nor accessible
(Ariel, 1988, p. 66) and thus highly unlikely to be a topical referent, indicating that
the sentence does not encode Topic-continuity:

(9) a. _Souffrance_ si **est**   _semblable a esmeraude…_
         suffering   SI be.3SG similar   to emerald
         'Suffering is similar to emerald…'   (_Graal_ 189c, 18)
     b. _et   ceste ville_ si **est**   _mult riche…_
         and this town SI be.3SG very rich
         'and this town is very rich…'   (_Villehardouin1_, 86, 4)

(10) … _si m'_**avint**   _une avision merveilleuse_
         SI me.CL=come.3SG.PST a   vision marvellous
         '[When I was asleep] a marvellous vision came to me'   (_Graal_ 197a, 23–24)

A final approach to the syntax of SI, sees the particle as intimately linked with the
V2 syntax of Old French and Medieval Romance.[2] Here, SI is either viewed as a
phrasal expletive category, which acts as an alternative to V2-related XP merger
(Benincà, 1995, p. 333; Rouveret, 2004, pp. 193–195; Salvesen, 2013, p. 143; Wolfe,
2016, pp. 469–470), or as a Head, whose merger is either an alternative (Ledgeway,

---

2. For a detailed comparative treatment of Medieval Romance's V2 syntax see Benincà (2004:
2006, 2013).

2008, pp. 452–465) or complementary (Ferraresi & Goldbach, 2002, pp. 18–23) strategy to V-to-C movement.

Although I go on to reformulate and refine the analyses offered in this vein, they are appealing on a number of grounds. Firstly, the expletive analysis accounts for the numerous observations in the literature that sɪ often seems to be semantically vacuous (Foulet, 1928, section 300; Roberts, 1993, p. 330; Vance, 1995, p. 185, 1997, p. 53). Secondly, sɪ's near-absence in embedded contexts is accounted for (Marchello-Nizia, 1985, p. 15; Lemieux & Dupuis, 1995, p. 96; Ferraresi & Goldbach, 2003, p. 113), as V2 is generally not operative in embedded clauses. Thirdly, the decline of the particle at the end of the 15th century (Marchello-Nizia, 1985, p. 200) is also predicted, as this is when the V2 constraint is generally viewed as no longer operative (Vance, 1995).

In what follows, I seek to show that whilst the literature on sɪ, and indeed the evaluation offered so far of that literature, may seem contradictory, this is not always because the literature is wrong. Rather, there is synchronic and diachronic variation in the syntax of sɪ which is indicative of an ongoing grammaticalisation process, where sɪ undergoes both pragmatico-semantic and syntactic changes.[3]

3. A new approach

A close examination of a range of Old French texts provides ample evidence of rich synchronic and diachronic variation in the grammar of sɪ. Here, I draw on the major descriptive findings of the corpus analysis in Wolfe (2018),[4] with a particular focus on the diachronic generalisations that emerge.

3.1   Latin and the Strasbourg Oaths

Although this chapter does not focus principally on Latin, a few preliminaries are helpful for a discussion of the grammaticalisation of sɪ throughout the history of Old French. As noted above, Latin sɪc is a temporal deictic adverb, generally translated as 'so, thus' and is extremely common in later Latin texts in particular (Salvi, 2004, p. 15; Pinkster, 2006, pp. 65, 107–107; Ledgeway, In Press a). Fleischman (1992, p. 436) draws particular attention to cases such as the following, where sɪc follows a temporal clause and clearly encodes temporal succession:

---

3. Fleischman (1992) also posits a grammaticalisation pathway for sɪ, but the majority of the data and analyses offered here are new.

4. Texts cited here are taken from the *Base de Français Médiéval. See http://bfm.ens-lyon.fr*

(11) *At   ubi   autem   sexta   hora   se   **fecerit,**   sic*
at   where however sixth.ɴᴏᴍ hour.ɴᴏᴍ ʀᴇꜰʟ.ᴄʟ do.3sɢ.ᴘʀꜰ.sʙᴊᴠ then
itur   ante   Crucem
walk.3sɢ.ᴘᴀss before cross.ᴀᴄᴄ
'However, when the sixth hour struck, one then goes before the Cross'

(*Peregrinatio* 37, 4)

However, sɪᴄ is very frequent in different contexts across a range of later Latin texts
and can also occupy initial position of the clause:

(12) *et   sic   **plecaremus**   nos ad montem   Dei*
and thus arrive.1ᴘʟ.ᴘsᴛ.sʙᴊᴠ we at mountain.ᴀᴄᴄ God.ɢᴇɴ
'And we thus approached the mountain of God'   (*Peregrinatio* 2, 4)

Importantly for the analysis of grammaticalisation that follows, sɪᴄ, when used as
a manner adverb meaning 'thus, like this', is always in postverbal position in the
Peregrinatio, the text from which the other examples above come:

(13) *Item   dies   paschales   cum   uenerint,*
likewise day.ɴᴏᴍ.ᴘʟ paschal.ɴᴏᴍ.ᴘʟ when come.sʙᴊᴠ.ᴘᴇʀꜰ.3ᴘʟ
celebrantur   *sic*
celebrate.ᴘᴀss.3ᴘʟ thus
'And when the Paschal days come, they are celebrated thus'

(*Peregrinatio*, XXVII 1)

Indeed, although multiple constituents precede sɪ in the text, consider the earliest
attestation of sɪ in 'Gallo-Romance', where it occurs after a temporal clause, akin to
that found in Example (11). Although this one example is too fragmentary to offer
a full analysis, we will see that the ability for multiple constituents to occur before
sɪ is typical of the earliest French texts.

(14) *d'ist   di in avant, in quant Deus savir   et   podir me*
from-this day in forward in so-far-as God knowledge and power me.ᴄʟ
*dunat,   si **salvarai**   eo cist meon fradre   Karlo*
give.3sɢ sɪ support.1sɢ.ꜰᴜᴛ I this my   brother Charles
'From this day forward, insofar as God gives me knowledge and power, I will
support this my brother Charles'   (*Strasbourg Oaths*, as in Price, 1971, p. 8)

3.2   Evidence for diachronic change

Despite the relatively wide range of textual records available for the history of French,
showing neat diachronic progression is far from straightforward. We should there-
fore proceed with some degree of caution, keeping in mind that variation according

to the verse or prose status of a text, its regional provenance, and its genre[5] is closely entwined with any evidence for a change in progress.[6] Nevertheless, I will suggest that it is possible to propose a plausible pathway of diachronic change in the syntax of sɪ, which is in keeping with acknowledged principles of grammaticalisation and other recognised major syntactic changes affecting the language.

If we examine the distribution of sɪ in the *Chanson de Roland*, a text dating from approximately 1100, we see the clearest evidence to support Fleischman's (1991, 1992), Van Reenen & Schøsler (1992, 1993, 2000) and others' account of sɪ as a marker of Topic-continuity. Firstly, the most common context for sɪ in the text is when it occurs as the sole preverbal constituent (138/201 occurrences):

(15) *Si **recevrez**   la   lei de chrestïens*
      sɪ receive.2sg.fut the law of christians
      'You will receive the Christian rite'   (*Roland* 38)

This is not surprising under this analysis, where sɪ is predicted to show the same distribution as other constituents which can occupy the preverbal field. Indeed, witness numerous observations on Old French V2 and V2 languages in general which suggest that the preverbal field hosts a thematic constituent which is interpreted as a 'default Topic' in the unmarked case (Vikner, 1995, p. 41; Buridant, 2000, section 631; Rouveret, 2004, pp. 185–186; Steiner, 2014, pp. 187–245; Holmberg, 2015, p. 376). sɪ is therefore showing a distribution in keeping with a class of other thematic constituents, in particular nominal subjects and pronouns which give rise to a sense of discourse cohesion in early texts.

Perhaps the most compelling evidence for the Topic-continuity analysis, however, comes from the distribution of subject expressions in clauses featuring sɪ. This is an area where both Marchello-Nizia (1985, pp. 169–178) and Fleischman (1991, p. 265) have previously noted diachronic variation between earlier and later texts. Restricting ourselves for the moment to the *Roland*, we find that only 9.45% of clauses featuring sɪ in the corpus assembled (19/201) show an overt postverbal subject and, furthermore, only 1% of clauses (2/201) show a preverbal subject:

(16) *Jo si nen **ai**   filz ne fille   ne heir*
      I   sɪ neg have.1sg son neg daughter neg heir
      'I have neither a son, nor daughter, nor an heir [to succeed me]'
                                                               (*Roland* 2744)

---

5. For a case-study of the fundamental effect of text type on the progression of change in Old French syntax, see recent work by Balon & Larrivée (2016)

6. Indeed, the syntax of sɪ itself has been shown to be subject to regional variation (Reenen & Schøsler, 1992, 1993, 2000; Schøsler & Reenen, 1996).(P. van Reenen & Schøsler, 1992, 1993, 2000; Schøsler & Reenen, 1996) (Reenen & Schøsler, 1992, 1993, 2000; Schøsler & Reene, 1996)

This is again in keeping with the predictions of the Topic continuity account: if sɪ signals to the hearer or reader that the theme or Topic of the preceding discourse is the same in the present clause, we predict that the need for an overt subject to serve a disambiguating function should, in the idealised case, not be required at all and at the least represent a minority option, which it clearly does in the *Roland*.

A final point is of significance for any syntactic analysis: *Roland* also shows evidence that a range of verbal complements can precede sɪ. These include an informationally new direct object (17a), an indirect object, locative adverbials (17b), and manner adverbials (17c):

(17) a. *Reis Vivien si **succuras** en Imphe*
     King Vivien sɪ help.2sɢ.ꜰᴜᴛ in Imphe
     'You will help King Vivien there in Imphe'   (*Roland* 3995–3996)

   b. *Sur l'erbe verte si **est** caeit envers*
     on the-grass green sɪ be.3sɢ fall.ᴩᴛᴄᴩ on-back
     'He has fallen down on his back on the green grass'   (*Roland* 2269)

   c. *Cunquerrantment si **finereit** li bers*
     conquering.ᴀᴅᴠ sɪ end.3sɢ.ᴄᴏɴᴅ the nobleman
     'The nobleman would end his life as a conquering hero'   (*Roland* 2867)

As an interim summary, we see that the Topic-continuity analysis of sɪ gets into no major difficulties when held up against the data available from one of the first French texts of a substantial length. However, we will now see that many of the properties pointing towards this conclusion for the *Roland* do not hold true of later verse and prose.

Here I outline a number of differences between the distribution of sɪ in the *Roland* and *Charrette*, another verse text from approximately 80 years later (c. 1177–1181). The first revealing difference concerns the relative frequency of sɪ-initial clauses. As noted above, this is the overwhelmingly dominant order for sɪ in the *Roland* (68.66%), but accounts for only 45.92% of sɪ occurrences in *Charrette* (135/294). The main context which sees a notable concomitant rise in *Charrette* is that where sɪ occurs following an initial clause, which accounts for only 3.98% of the *Roland* corpus (8/201) but 19.05% of the corpus from *Charrette* (56/294):

(18) *Et, quant ele vint devant lui,// Si li **dit***
    and when she come.3sɢ.ᴩsᴛ before him sɪ him.ᴄʟ= say.3sɢ.ᴩsᴛ
    'And when she comes before him, she says….'   (*Charrette* 131–132)

There are also differences between these two texts concerning the relationship between sɪ and subject expressions. In particular, whereas overt preverbal subjects account for a mere 1% of sɪ occurrences in the *Roland*, these account for a moderately increased 4.76% of the data in *Charrette* (14/294). In Wolfe (2018) this rise in

the proportion of Subject + SI configurations is shown also to be typical of the text *Villehardouin* (1199–1213):

(19) a. *Et <u>cil</u> <u>autres</u> si **est**…*
        and this other SI be.3SG
        'And this other one is…'   (*Charrette* 5816)

      b. *Et <u>cil</u> <u>qui</u> <u>empereres</u> <u>seroit</u>   <u>par</u> <u>l'eslecion,</u>   si **aroit***
        and he who emperor be.3SG.FUT by the-election SI have.3SG.FUT
        lo   quart de tote la   conqueste
        the quarter of all the conquest
        'And he who was elected emperor by then, will have a quarter of [what was gained] from the whole conquest'   (*Villehardouin2*, 34, 234)

Furthermore there are revealing differences between *Roland* and *Charrette* in terms of other preverbal constituents that can co-occur with SI. Firstly, there are no direct objects co-occurring with SI in the corpus from *Charrette*, in contrast to *Roland* where there is a single instance. Secondly, however, an example on an indirect object prepositional phrase is attested with SI (20):

(20) *Au   roi, si li   comance a dire…*
      to-the king SI him.CL= start.3SG to say.INF
      'He starts to say to the king…'   (*Charrette* 27d, 85)

Thirdly, a seemingly important difference between *Roland* and *Charrette* concerns the types of adverbials licensed in ADVERBIAL + SI configurations. Whilst we saw above that manner adverbials are found with SI in the *Roland* (see also Marchello-Nizia, 1985, pp. 157–160), this is not the case in our corpus from *Charrette*, where all adverbials serve a frame-setting function, anchoring the clause in terms of its locative or temporal coordinates (Benincà, 2004; Benincà & Poletto, 2004, p. 66; Öhl, 2010, p. 62):

(21) a. *puis si s'an   va*
        then SI REFL.CL=LOC.CL= go.3SG
        'Then he goes…'   (*Charrette* 931)

      b. *Lors si **fust**   trop granz <u>la</u> <u>folie</u>*
        thus SI be.3SG.PST too great the foolishness
        'The foolishness was thus too great…'   (*Charrette* 9845)

To summarise, once more, our findings for this text, we see a number of differences when comparing it to the *Roland*. These concern the decrease of SI-initial clauses and concomitant rise of CLAUSE + SI configurations and the absence of VP adverbials and direct objects occurring before SI.

Evidence of further intertextual and, as I claim below, diachronic variation is observable if we turn to a 13th-century prose text, *La Queste del saint Graal* (1225–1230). This text warrants attention, as it frequently shows what might be termed 'innovative' developments in clausal syntax and has appeared extensively in syntactic studies of Old French (Roberts, 1993; Vance, 1997; Rouveret, 2004, pp. 212–216; Wolfe, 2016).

A major indication that by the time of *La Queste*, the syntax of sɪ has changed markedly comes from the relative frequencies of sɪ-initial and CLAUSE + sɪ configurations. In *La Queste* an initial clause precedes sɪ in 177/300 occurrences (59%), whilst sɪ-initial contexts only account for 32.33% of the corpus (97/300). This is markedly different from *Roland* where, recall, the sɪ-initial and CLAUSE + sɪ configurations account for 68.66% and 3.98% of contexts respectively. As suggested in Section 2, it is not at all obvious that the majority of clause-initial sɪ configurations are amenable to a Topic-continuity analysis, which calls the appropriateness of this route for 13th-century prose into question.

Further evidence of a difference between *La Queste* and the other texts considered here is found in the absence of any DIRECT OBJECT + sɪ attestations. Indeed, when Prepositional Phrases co-occur with sɪ in *La Queste*, they have the prototypical characteristics of frame-setters laid out above. Adverbials with sɪ in *La Queste* are not attested within my corpus of 300 sɪ-clauses.

Taken together, these data show, on the basis of three texts alone, that there is significant intertextual variation in the syntax of sɪ, a point which is engaged with in certain works on the particle (Marchello-Nizia, 1985; Reenen & Schøsler, 1992; Reene & Schøsler, 2000; Salvesen, Forthcoming) and suggests it is not the homogeneous entity many works on Old French syntax imply it is. In what follows, however, I suggest there is a stronger interpretation of the data that suggests the intertextual variation can be mapped onto a grammaticalisation cline for sɪ.

4.   Towards a grammaticalisation pathway for sɪ

A legitimate question with which to start our discussion of the grammaticalisation of sɪ is why we might treat this case of variation and change as grammaticalisation at all. There are several reasons to do so. First, as hinted at above and set out in detail below, sɪ follows the traditional characterisation of grammaticalisation phenomena in its shift from lexical to functional element (Heine & Kuteva, 2002, p. 2; Hopper & Traugott, 2003, pp. 4–6; Lehmann, 2015, p. 11), in changing from a (semi-)lexical manner adverb to a temporal deictic adverb and, later, V2-related expletive. Second, as we shall see, and as many scholars have pointed out (Foulet,

1928, section 300; Ménard, 1968, p. 81; Moignet, 1973, pp. 287–288), sɪ becomes increasingly 'bleached' of its previous semantic content.[7] Third, numerous scholars of grammaticalisation highlight a widening of distribution as a 'defining factor of grammaticalization' (Lehmann, 2015, p. 7), an observation that clearly fits with the ubiquity of sɪ in Old French. Witness, for example, comments by Foulet (1928, p. 300) that in Old French 'there is hardly a page where it doesn't appear several times'. Fourth, recent formally oriented work on grammaticalisation has emphasised the role of successive 'upward' reanalysis of the element in question throughout the clausal spine (Roberts & Roussou, 2002, p. 202; Van Gelderen, 2008, p. 247, 2011, Chapter 1; Roberts, 2010, pp. 49–50, 2012, p. 353); I will now suggest that sɪ provides a paradigm case study of such a change.

It is important to note from the outset that evidence of sɪc's upwards grammaticalisation pathway begins before the Old French period. Recall that in post-Classical Latin texts, sɪc can maintain its manner adverbial function in postverbal position, yet occurs preverbally when encoding temporal succession as a temporal deictic adverb (contrast (12) and (13) above), with the latter usage particularly associated with the *Peregrinatio* (see discussion in Löfstedt, 1959, p. 47). Given recent proposals that the finite verb in certain post-Classical Latin texts moves to a position within the high inflectional layer of the clause (Danckaert, 2017, pp. 207, 239), we can derive these facts if we hypothesise that sɪc can optionally move from a position in the extended Verb Phrase to a high position within the inflectional layer (IP/TP), argued by Ledgeway (In Press b) to host temporal deictic adverbs:[8]

i.   Reanalysis 1: $\mathbf{Adv_{Manner}} > \mathbf{Adv_{Temporal}}$
     …[$_{Fin}$ [$_{TDeixis}$ (**sɪc**) [$_{SubjP}$ … **V**…[$_{vP}$ **sɪc**…]]]]

The next stage along the grammaticalisation cline involves a small-scale change, where sɪ is no longer limited to encoding temporal continuity, but can, as Fleischman's (1991) account predicts, encode general thematic continuity. This is the system we see instantiated in the *Roland*, where sɪ like other adverbials can be fronted into the left periphery of the clause (CP) to satisfy the V2 constraint:

ii. Reanalysis 2: $\mathbf{Adv_{Temporal}} > \mathbf{Adv_{TopicContinuity}}$
     [$_{Frame}$ [$_{Force}$ [$_{Topic}$ [$_{Focus}$ [$_{Fin}$ (**sɪ**) [$_{Fin^o}$ **V**] [$_{TDeixis}$ (**sɪ**) [$_{SubjP}$ [$_{vP}$ …]]]]]]]]]

---

7.   On the role of semantic bleaching in grammaticalisation see Hopper & Traugott (2003, pp. 94–98), Roberts & Roussou (2002, pp. 218–219), Roberts (2010, Section 5) and Lehmann (2015, pp. 134–136) amongst many others.

8.   Under an alternative view, where late Latin is already a V2 language, sɪc could already be raising into the left periphery (Ledgeway, 2012, p. 153, 2017; Wolfe, 2015).

However, even in this very early text a Topic-continuity adverbial analysis may not capture a small minority of the data. If we consider (22), we find ꜱɪ following an initial wh-clause, where the subject of the clause *Rollant* 'Roland' is entirely accessible from context. One way of accounting for this data is to suggest in line with Donaldson (2012) that some initial clauses in Old French are base-generated (i.e. non-moved) high in the left periphery and as such cannot move via the V2 bottleneck to satisfy the V2 constraint. Put simply, clauses such as the *quant*-clause below do not count for V2. As such, the head responsible for V2 effects searches for the closest available constituent to satisfy V2, which is then attracted to its specifier (see Wolfe 2018). As such ꜱɪ performs a quasi-expletive function, satisfying V2 when a constituent that cannot do so is the sole preverbal constituent. In non-theoretical terms, ꜱɪ is acting as a kind of V2-related place-holder.

(22) *Quant l'ot   Rollant, si* **cumençat**   *a rire*
      when   it.ᴄʟ=hear.3ꜱɢ Roland ꜱɪ begin.3ꜱɢ.ᴘꜱᴛ to laugh.ɪɴꜰ
      'When Roland heard this, he began to laugh'   (*Roland* 303)

Given the articulated left periphery assumed by many scholars for 11th- and 12th-century French (Rouveret, 2004; Labelle, 2007; Donaldson, 2012; Salvesen, 2013), the question arises as to which position in the left periphery ꜱɪ moves to in the *Roland*. The fact that both topicalised (16) and focussed (17) constituents can precede it strongly suggests it stays low within the left periphery, in the specifier of Fin. This is an unsurprising finding, as much of the literature on Early Old French (i.e. Labelle, 2007) assumes that Fin and SpecFinP host the finite verb and constituent satisfying V2 respectively at this stage in the language's history. As an expletive-like element we predict its ultimate position to correlate with the locus of V2 effects more generally.

   The story of ꜱɪ in *Charrette* and 12th century texts in general is somewhat of a halfway-house between its distribution in the *Roland* and *La Queste*, showing that what we may conceptualise as distinct stages in a grammaticalisation process overlap to some extent. In particular, the substantial rise in ᴄʟᴀᴜꜱᴇ + ꜱɪ configurations and ꜱɪ's increasing incompatibility with focal elements, suggests strongly that the last-resort nature of the particle as a V2-related expletive has progressed. Further cross-textual evidence for this hypothesis comes from contexts where ꜱɪ co-occurs with adverbials in *Eneas*, another 12th century verse text. Crucially, all such adverbials are frame-setting elements, which, importantly in formal terms, are assumed to be base-generated (not moved) high in the left periphery (Ernst, 2009, Section 3). Crucially, these elements would not therefore count for V2, so ꜱɪ must be merged in SpecFinP in the absence of another appropriate element. Evidence that ꜱɪ in *Charrette* (and indeed other texts) should better be analysed as an expletive than a marker of Topic continuity, or in the terms of Fleischman (1992, p. 436),

a 'same-subject marker', comes from a discourse-pragmatic analysis with overt postverbal subjects co-occuring with SI. Strikingly, 77.78% of such cases involve subject expression which are discourse-NEW, a finding which would be entirely surprising if Fleischman and others' analysis held of this text. As such, I conclude that in at least some texts from the mid-12th century SI has been grammaticalised as a FinP expletive which is first merged there rather than reaching this position via movement:

iii. Reanalysis 3: **Adv$_{TopicContinuity}$ > Expletive$_{Fin}$**
$[_{Frame} [_{Force} [_{Topic} [_{Focus} [_{Fin}$ **SI** $[_{Fin^o}$ **V**$] [_{TDeixis} [_{SubjP} [_{vP} \cdots]]]]]]]]]]$

There is an emerging view in the literature that the V2 system of French undergoes substantial changes at the turn of the 13th century, many of which are instantiated in *La Queste*. Specifically, it has been claimed by both Rouveret (2004) and Wolfe (2016) that within *La Queste* we can observe the emergence of a Force-V2 grammar where both verb movement and phrasal merger target Force^o and SpecForceP respectively, as opposed to Fin^o and SpecFinP in earlier French.[9] Support for this hypothesis comes from the distribution of SI in *La Queste*, which itself appears to have been reanalysed as a ForceP expletive. Put briefly, there is clear evidence of less left-peripheral structure to the left of SI: Orders where SI is preceded by more than one constituent are hardly attested (1/300 attestations), in contrast to earlier texts, all PPs or adverbials co-occurring with SI are typical frame-setters and CLAUSE + SI configurations rise to account for 59% of attestations (177/300). This last finding is unsurprising if SI is merged in Spec-ForceP, as adverbial, wh- and circumstantial clauses are prototypical frame-setters so are predicted to be among the very small class of constituents that can precede SI:

iv. Reanalysis 4: **Expletive$_{Fin}$ > Expletive$_{Force}$**
$[_{Frame} [_{Force}$ **SI** $[_{Force^o}$ **V**$] [_{Topic} [_{Focus} [_{Fin} [_{TDeixis} [_{SubjP} [_{vP} \cdots]]]]]]]]]]$

Overall, we have seen that the data can plausibly be accounted for under an analysis where SI undergoes successive upwards reanalyses throughout the clausal spine. Whilst the textual database used for our study is representative rather than exhaustive, the proposals are in keeping with other findings concerning syntactic change during 9th to 13th-century French.

---

9. In simple terms this is formally a difference as to whether the verb and constituent which move to satisfy V2 must target a position low or high within the extended Complementiser Phrase. In descriptive terms, the height of V2 corresponds to its relative strictness in terms of V>3 orders.

5. Consequences

In this short chapter, it has been shown that the syntax of si shows considerable variation within the Old French period which, as has been suggested, can be mapped on to a grammaticalisation cline. The cline can be outlined from Classical Latin through to 13th century French as follows:

v. $\textbf{Adv}_{\textbf{Manner}} > \textbf{Adv}_{\textbf{Temporal}} > \textbf{Adv}_{\textbf{TopicContinuity}} > \textbf{Expletive}_{\textbf{Fin}} > \textbf{Expletive}_{\textbf{Force}}$

Further research is needed on wider textual basis to corroborate the findings, but as it stands there are several points to note for the cline proposed.

Firstly, si appears to be an unambiguous case of upwards reanalysis, examples of which are reported widely in the literature, but is perhaps unusual due to the fact that successive reanalyses see si moving upwards from a low position within an articulated clausal hierarchy to the very highest reaches of the clause. This stands in contrast to a number of familiar case-studies, where the reanalysis often takes place just within a particular clausal layer, that is to say the VP, TP or CP.

Secondly, the case of si clearly illustrates the interconnectedness of a number of morphosyntactic changes affecting the history of French. Of particular interest is the fact that si's grammaticalisation appears to 'leapfrog' off changes affecting the locus of verb movement and Extended Projection Principle (EPP) effects.

Overall, despite its appearance in a number of very early works on the history of Romance, the intriguing particle si still has much to offer both Romance and historical linguistics.

Abbreviations

ACC	accusative	NEG	negation	REFL	reflexive
ADV	adverb(ial)	NOM	nominative	SG	singular
CL	classifier	PASS	passive	SBJV	subjunctive
COND	conditional	PL	plural	1	first person
FUT	future	PRF	perfect	2	second person
INF	infinitive	PST	past	3	third person
LOC	locative	PTCP	participle		

Bibliography

Adams, M. (1987). From Old French to the Theory of Pro-Drop. *Natural Language & Linguistic Theory*, 5(1), 1–32. https://doi.org/10.1007/BF00161866

Ariel, M. (1988). Referring and accessibility. *Journal of Linguistics*, 24, 65–87. https://doi.org/10.1017/S0022226700011567

Balon, L. & Larrivée, P. (2016). L'ancien français n'est déjà plus une langue à sujet nul–nouveau témoignage des textes légaux. *Journal of French Language Studies*, 26(2), 221–237. https://doi.org/10.1017/S0959269514000222

Bauer, B. (1995). *The emergence and development of SVO patterning in Latin and French: diachronic and psycholinguistic perspectives*. New York: OUP.

Benincà, P. (1995). Complement clitics in medieval Romance: The Tobler-Mussafia law. In I. Roberts & A. Battye (Eds.), *Clause structure and language change*. (pp. 325–344). Oxford: OUP.

Benincà, P. (2004). The left periphery of Medieval Romance. *Studi Linguistici e Filologici Online*, 2(2), 243–297.

Benincà, P.(2006). A detailed map of the left periphery of medieval Romance. In: R. Zanuttini (Ed.). *Crosslinguistic research in syntax and semantics: Negation, tense and clausal architecture*. (pp. 53–86). Georgetown: Georgetown University Press.

Benincà, P. & Poletto, C. (2004). Topic, focus, and V2. In L. Rizzi (Ed.), *The Structure of CP and IP* (pp. 52–75). Oxford: OUP.

Buridant, C. (2000). *Grammaire nouvelle de l'ancien français*. Paris: Sedes.

Cruschina, S. (2006). Information focus in Sicilian and the left periphery. In M. Frascarelli (Ed.), *Phases of Interpretation* (pp. 363–385). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110197723.5.363

Cruschina, S. & Ledgeway, A. (2016). The Structure of the Clause. In A. Ledgeway & M. Maiden (Eds.), *The Oxford Guide to the Romance Languages*. Oxford: OUP. https://doi.org/10.1093/acprof:oso/9780199677108.003.0031

Danckaert, L. (2017). *The development of Latin clause structure: a study of the extended verb phrase* New York, NY: OUP. https://doi.org/10.1093/oso/9780198759522.001.0001

Den Besten, H. (1983). On the Interaction of Root Transfomrations and Lexical Deletive Rules. In W. Abraham (Ed.), *On the Formal Syntax of the Westgermania* (pp. 47–61). Amsterdam: John Benjamins. https://doi.org/10.1075/la.3.03bes

Diez, F. (1882). *Grammatik der romanischen Sprachen*. Bonn: Weber.

Donaldson, B. (2012). Initial subordinate clauses in Old French: Syntactic variation and the clausal left periphery. *Lingua*, 122(9), 1021–1046. https://doi.org/10.1016/j.lingua.2012.04.003

Ernst, T. (2009). Speaker-oriented adverbs. *Natural Language & Linguistic Theory*, 27(3), 497–544. https://doi.org/10.1007/s11049-009-9069-1

Ferraresi, G. & Goldbach, M. (2002). V2 Syntax and Topicalization in Old French. *Linguistiche Berichte*, 189, 2–25.

Ferraresi, G. & Goldbach, M. (2003). Particles and sentence structure: a historical perspective. In *Syntactic Structures and Morphological Information*. Berlin: De Gruyter. https://doi.org/10.1515/9783110904758.101

Fleischman, S. (1991). Discourse pragmatics and the grammar of Old French: A functional reinterpretation of 'si' and the personal pronouns. *Romance Philology*, 44, 251–283.

Fleischman, S. (1992). Discourse and diachrony: The rise and fall of Old French SI. In M. Gerritsen & D. Stein (Eds.), *Internal and External Factors in Syntactic Change* (pp. 433–474). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110886047.433

Fleischman, S. (2000). Methodologies and Ideologies in Historical Linguistics: On Working with Older Languages. In S. C. Herring, P. van Reenen, & L. Schøsler (Eds.), *Textual parameters in older languages* (pp. 33–58). Amsterdam: John Benjamins.

Foulet, L. (1928). *Petite syntaxe de l'ancien français* (3rd Edition). Paris: H. Champion.

Harris, M. (1976). A Typological Approach to Word-Order Change in French. In M. Harris (Ed.), *Romance Syntax: Synchronic and Diachronic Perspectives* (pp. 33–68). Salford: University of Salford.

Harris, M. (1978). *The evolution of French syntax: a comparative approach*. London; New York: Longman.

Harris, M. B. (1984). On the causes of word order change. *Lingua*, 63(2), 175–204. https://doi.org/10.1016/0024-3841(84)90053-6

Heine, B. & Kuteva, T. (2002). *World Lexicon of Grammaticalization*. CUP. https://doi.org/10.1017/CBO9780511613463

Holmberg, A. (2015). Verb Second. In T. Kiss & A. Alexiadou (Eds.), *Syntax – Theory and Analysis* (pp. 242–283). Berlin: Walter de Gruyter.

Hopper, P. J. & Traugott, E. C. (2003). *Grammaticalization*. CUP. https://doi.org/10.1017/CBO9781139165525

Jensen, F. (1990). *Old French and Comparative Gallo-Romance Syntax*. Tübingen: Walter de Gruyter. https://doi.org/10.1515/9783110938166

Klausenburger, J. (2000). *Grammaticalization: Studies in Latin and Romance Morphosyntax*. Amsterdam: Benjamins.

Labelle, M. (2007). Clausal architecture in Early Old French. *Lingua*, 117(1), 289–316. https://doi.org/10.1016/j.lingua.2006.01.004

Ledgeway, A. (In Press a). Second position elements and late Latin V2 syntax. In *Rethinking Verb Second*. Oxford: OUP.

Ledgeway, A. (In Press b). The Verb Phrase. In G. Longobardi (Ed.), *The Syntax of Italian*. Cambridge: CUP.

Ledgeway, A. (2008). Satisfying V2 in early Romance: Merge vs. Move. *Journal of Linguistics*, 44(02). https://doi.org/10.1017/S0022226708005173

Ledgeway, A. (2011a). Grammaticalization from Latin to Romance. In : B. Heine & H. Narrog (Eds.). *The Oxford Handbook of Grammaticalization*. (pp. 719–728). Oxford: Oxford University Press.

Ledgeway, A. (2011b). Syntactic and morphosyntactic typology and change in Latin and Romance. In M. Maiden, J. C. Smith, & A. Ledgeway (Eds.), *The Cambridge History of the Romance Languages* (pp. 382–471). Cambridge: CUP.

Ledgeway, A. (2012). *From Latin to Romance: morphosyntactic typology and change*. Oxford: OUP. https://doi.org/10.1093/acprof:oso/9780199584376.001.0001

Ledgeway, A. (2017). Late Latin Verb Second: The Sentential Word Order of the Itinerarium Egeriae. *Catalan Journal of Linguistics*, (Special Issue: Generative Approaches to Latin Syntax), 163–216. https://doi.org/10.5565/rev/catjl.212

Lehmann, C. (2015). *Thoughts on grammaticalization* (3rd edition). Berlin: Language Science Press. https://doi.org/10.26530/OAPEN_603353

Lemieux, M., & Dupuis, F. (1995). The Locus of Verb Movement in Non-Asymmetric Verb-Second Languages: The Case of Middle French. In I. Roberts & A. Battye (Eds.), *Clause Structure and Language Change* (pp. 80–110). Oxford: OUP.

Löfstedt, E. (1959). *Late Latin*. Oslo: H. Aschehoug & Co.

Malkiel, Y. (1988). *The Triple Janus Head of Romance Linguistics. Lecture delivered in October 1988 in the Senate House of the University of Liverpool*. Liverpool: Liverpool University Press.

Marchello-Nizia, C. (1985). *Dire le vrai : L'adverbe <<si>> en français médieval : Essai de linguistique historique*. Geneva: Droz.

Mathieu, É. (2012). The left periphery in Old French. In D. Arteaga (Ed.), *Research in Old French: The state of the art* (pp. 327–350). Dordrecht: Springer.

Ménard, P. (1968). *Manuel d'ancien français* (Vol. 3. Syntaxe). Bordeaux: Sobodi.

Moignet, G. (1973). *Grammaire de L'ancien français: morphologie – syntaxe*. Paris: Éditions Klincksieck.

Öhl, P. (2010). Formal and functional constraints on constituent order and their universality. In C. Breul & E. Göbbel (Eds.), *Linguistik Aktuell/Linguistics Today* (Vol. 165, pp. 231–276). Amsterdam: John Benjamins.

Pinkster, H. (2006). *On Latin Adverbs*. Amsterdam University Press.

Price, G. (1971). *The French Language: Past and Present*. London: Edward Arnold.

Prince, E. F. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics* (pp. 223–255). New York: Academic Press.

Reenen, P. van & Schøsler, L. (1992). Ancien et Moyen Français: SI 'thématique', analyse exhaustive d'une série de textes. *Vox Romanica*, 51, 101–127.

Reenen, P. van & Schøsler, L. (1993). SI 'thématique', étude de SI en ancien et moyen français, discours direct". In *Actes du XXe Congrès International de Linguistique et Philologie Romanes* (pp. 617–628). Zürich.

Reenen, P. van & Schøsler, L. (2000). The pragmatic functions of the Old French particles ainz, apres, donc, lors, or, pluis, and si. In S. C. Herring, P. van Reenen, & L. Schøsler (Eds.), *Textual Parameters in Older Languages* (pp. 59–105). Amsterdam: John Benjamins.

Roberts, I. (1993). *Verbs and diachronic syntax: a comparative history of English and French*. Dordrecht: Kluwer.

Roberts, I. (2010). Grammaticalization, the clausal hierarchy and semantic bleaching. In E. C. Traugott & G. Trousdale (Eds.), *Typological Studies in Language* (Vol. 90, pp. 45–73). Amsterdam: John Benjamins.

Roberts, I. (2012). Diachrony and Cartography: Paths of Grammaticalization and the Clausal Hierarchy. In L. Brugé, A. Cardinaletti, G. Giusti, N. Munaro, & C. Poletto (Eds.), *Functional Heads: The Cartography of Syntactic Structures, Volume 7* (pp. 351–367). OUP. https://doi.org/10.1093/acprof:oso/9780199746736.003.0027

Roberts, I. & Roussou, A. (2002). *Syntactic change a minimalist approach to grammaticalization*. Cambridge: CUP.

Rouveret, A. (2004). Les clitiques pronominaux et la périphérie gauche en ancien français. *Bulletin de La Société de Linguistique de Paris*, 99, 181–237. https://doi.org/10.2143/BSL.99.1.541914

Rowlett, P. (2007). *The syntax of French*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511618642

Salvesen, C. (Forthcoming). Resumptive Particles and Verb Second. R. Woods & S. Wolfe (Eds.), *Rethinking Verb Second*. Oxford: OUP.

Salvesen, C. (2013). Topics and the Left Periphery: A comparison of Old French and Modern Germanic. In T. Lohndal (Ed.), *In search of universal grammar: from Old Norse to Zoque* (pp. 131–172). Amsterdam: John Benjamins. https://doi.org/10.1075/la.202.10sal

Salvi, G. (2004). *La formazione della struttura di frase romanza: ordine delle parole e clitici dal latino alle lingue romanze antiche*. Tübingen: Max Niemeyer. https://doi.org/10.1515/9783110945508

Schøsler, L. & Reene, P. van. (1996). La fonction pragmatique des particules AINZ, APRES, DONC, LORS, OR, PUIS, SI en ancien français. In *Actes du XIIIe Congrès des Romanistes Scandinaves* (Vol. 2, pp. 635–658). Jyväskylä: Publications de l'Institut des Langues Romanes et Classiques.

Schwegler, A. (2013). *Analyticity and Syntheticity, A Diachronic Perspective with Special Reference to Romance Languages* (Reprint 2012). Berlin: De Gruyter Mouton.

Skårup, P. (1975). *Les premières zones de la proposition en ancien français*. København: Akademisk Forlag.

Skårup, P. (1975). *Les premières zones de la proposition en ancien français*. Copenhagen: Akademisk Forlag.

Sornicola, R. (2011). Romance Linguistics and Historical Linguistics: Reflections on synchrony and diachrony. In M. Maiden, J. C. Smith, & A. Ledgeway (Eds.), *The Cambridge History of the Romance Languages* (pp. 1–49). Cambridge: CUP.

Steiner, B. D. (2014). The Role of Information Structure in the Loss of Verb-Second in the History of French (PhD Thesis). University of Indiana, Bloomington, Boston, Massachusetts.

Thurneysen, R. (1892). Die Stellung des Verbums in Altfranzösischen. *Zeitschrift Für Romanische Philologie*, 16, 289–307. https://doi.org/10.1515/zrph.1892.16.1-4.289

Vance, B. (1995). On the decline of verb movement to Comp in Old and Middle French. In A. Battye & I. Roberts (Eds.), *Clause structure and language change*. (pp. 173–199). Oxford: OUP.

Vance, B. (1997). *Syntactic change in medieval French*. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-015-8843-0

Vance, B., Donaldson, B. & Steiner, B. D. (2009). V2 loss in Old French and Old Occitan: The role of fronted clauses. In S. Colina, A. Olarrea & A. M. Carvalho (Eds.), *Romance Linguistics 2009. Selected papers from the 39th Linguistic Symposium on Romance Languages (LSRL), Tuscon, Arizona* (pp. 301–320). Amsterdam: John Benjamins.

Vanelli, L., Renzi, L. & Benincà, P. (1986). Tipologia dei pronomi soggetto nelle lingue romanze medievali. *Quaderni Patavini Di Linguistica*, 5, 49–66.

Van Gelderen, E. (2008). Where did Late Merge go? Grammaticalization as feature economy. *Studia Linguistica*, 62(3), 287–300. https://doi.org/10.1111/j.1467-9582.2008.00150.x

Van Gelderen, E. (2011). *The Linguistic Cycle: Language Change and the Language Faculty: Language Change and the Language Faculty*. OUP, USA. https://doi.org/10.1093/acprof:oso/9780199756056.001.0001

Vikner, S. (1995). *Verb movement and expletive subjects in the Germanic languages*. New York: OUP.

Wolfe, S. (2015). Medieval Sardinian: New Evidence for Syntactic Change from Latin to Romance. In D. T. T. Haug (Ed.), *Historical Linguistics 2013: Selected papers from the 21st International Conference on Historical Linguistics, Oslo, 5–9 August 2013* (pp. 303–324). Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.334.16wol

Wolfe, S. (2016). A Comparative Perspective on the Evolution of Romance Clausal Structure. *Diachronica*, 33(4), 461–502. https://doi.org/10.1075/dia.33.4.02wol

Wolfe, S. (2018). Probing the syntax of a problematic particle: Old French 'si' revisited. *Transactions of the Philological Society* 116 (3), pp. 332–362. https://doi.org/10.1111/1467-968x.12123

The rise of the analytic Perfect aspect in the West Iranian languages

Vit Bubenik and Leila Ziamajidi

Memorial University of Newfoundland / Al-Zahra University, Tehran, Iran

This paper focuses on the long-term grammaticalization of tense/aspect systems in the West Iranian languages, beginning with Old Iranian (Section 1). In Middle Persian (Section 2) the Aorist and the reduplicative Perfect of Old Persian were replaced by a new system of analytic constructions. The fundamental mechanism in the rise of the innovative Preterit (perfective) and Perfect categories was the process of grammaticalization, reducing the auxiliary 'be' into suffixes of the innovative Preterit. In Early New Persian (Section 3) an unambiguous Perfect was recreated by attaching personal suffixes to the Perfect stem. In the second part of the paper (Section 4) we turn to the elaboration of the evidential ('non-witnessed') subsystem in New Persian through grammaticalization and possible Turkic influence. A typological parallel in the southernmost Slavic languages (Section 5) is provided.

Keywords: Middle and New Persian, Kurdish, Tajik, (analytic) Perfect, copula, grammaticalization, degrammaticalization, inferential and conjectural mode

1. Old Iranian Tense /Aspect/Mood system

Old Iranian is known to us in two varieties: earlier Gathic (documented in the earliest portions of the Avesta) and later Cuneiform Persian (used in the Achaemenid inscriptions of the 6th–4th c. BCE). The core tense /aspect/mood system of Gathic is presented in Table 1.

The Gathic system reflects the state of affairs during the 10th/9th c. BCE and is as archaic as that of Vedic. To mention the most obvious parallels: there is a relatively rich system of the Aorist and the Perfect; there are numerous modal forms (optative, subjunctive) found in both the Aorist and the Perfect; there are relics of the sigmatic Future (*vax-šy-ā* 'I will speak'); and the nasal Presents are relatively well preserved (*kr-ṇao-ti* 'he does'). As in Vedic, in Gathic the co-existence of primary

Table 1. Old Iranian (Gathic) Tense/Aspect/Mood system

PRESENT	AORIST	PERFECT
(Imperfective)	(Perfective)	(Retrospective)
kr-nao-ti	*vax-šy-ā*	*ta-taš-a*
'he does'	'I will speak'	'he has shaped'
	dār-š-t	
	'may he hold' (INJUNCTIVE)	
kr-nao-t		*yaug-t*
'he did'		'he yoked'
mrav-at(i)		*jam-at(i)*
'that he speak'	(SUBJUNCTIVE)	'that he went' (SUBJUNCTIVE)

and secondary suffixes in the subjunctive (*kr-ṇáv-āt(i)* 'that he do', *mrav-at(i)* 'that he speak') is also noticed. A remarkable archaism is the appearance of the passive Aorist formed by the suffix *-ī*: *vāčī* 'he was named', *srāvī* 'he was heard' (cf. Vedic *a-kār-i* 'it was done').

The aspectual and modal system of archaizing Old (Cuneiform) Persian is less archaic and more atrophied than that of Gathic but it has also preserved a number of original features: the augment in the Imperfect and the Aorist (appearing only sporadically in Gathic); consistent lengthening of the thematic vowel in the subjunctive (*ku-nav-ā-tiy* 'may he do'); and zero-grade in the plural forms of the root Aorist *a-ku-mā* (< *a-kr-ma) 'we made'. On the other hand, some other forms (such as the passive Aorist and non-modal forms of the reduplicative Perfect) are not documented at all. In spite of all the distinct morphology of the Aorist and the Perfect, these two forms came to be used interchangeably to express actions in the past "whether in progress or definitely terminated or habitual and repeated" (Kent, 1953, p. 90). For instance the same completed past event of 'having made battle' can be expressed by either the Aorist or the Imperfect:

(1)   *avaθā hamaranam a-   ku-   mā*   (Kent, 1953 [DB 1.90])
      then   battle   AUG- make.AOR- 1PL
      'Then we made battle'
      avadā hamaranam a-   ku-   nau- š   (Kent, 1953 [DB 2.23])
      then   battle   AUG- make- IMPF- 3SG
      'Then HE made battle'

2. The rise of analytic aspectual formations during the Middle Persian period

The inflectional categories of Old Iranian (Imperfect and Aorist) were gradually lost in Middle Persian but were recreated analytically, combining the past participle with the substantive verb (see Estaji & Bubenik 2007: 32).

Table 2. Middle Persian Tense/Aspect system

IMPERFECTIVE	PERFECTIVE	PERFECT
		(ergative construction)
kun-am/-ēm	*(u)=m kard*	*(u)=m kard h-am*
'I do'	'(and) I did'	'(and) I have done'
		(u)=m kard būd h-am
		'(and) I had done'
(IMPERFECT lost)	(AORIST lost)	

In the case of transitive verbs it can be argued that the ergative construction had already appeared during the Old Iranian period as a 'pseudo-possessive' construction *manā kart-am* lit. 'of-me done' 'I did' in (2). However, as long as the finite perfective and perfect active forms (i.e. Aorist and Perfect) were part of the aspectual system, the possessive construction *manā kart-am* could be interpreted passively:

(2) *ima tya   manā kart-am pasāva yaθā xšāyaθiya abavam*
      this that I-GEN do-PP   after when king   become.IPFV-1SG
      'This (is) that (which) was done by me after (I) became king'
                                                    (Kent, 1953 [DB I, 28–29])

To judge by the Achaemenid inscriptions, the construction *manā kart-am* (lit. 'of me done') became an alternative way of expressing the past completed events instead of the active Aorist and Perfect; i.e. the speakers of Old Persian had a choice between the active Aorist *avaθā hamaranam a-ku-mā* 'Then we made battle' (in (1)) and the passive construction with the passive participle: **avaΘā ahmākəm hamaranam kart-am* (then we-GEN battle do.PPP-NEUT) 'then battle was made by us'. During the Middle Persian period the neuter suffix marking on the PP was lost, leaving only the bare form *kard*. The eventual reinterpretation of the passive construction as an ergative construction is to be noted in the personal clitics. The Old Persian passive construction *utā=maiy kartam* (and=I-OBL do.PPP-NEUT) 'and it was done by me' was reduced to *u=m kard*, and the conjunction *uta* 'and' hosting the pronominal agent (by Wackernagel's Law) was reduced to *u*. Through attrition and an ultimate complete loss of the finite forms, the construction *u=m kard* became the only means of expressing the active perfective category, moving into the domain of the former Aorist (*u=m kard*, 'and I did' *u=t kard* 'and you did', *u=š kard* 'and he did').

In Middle Persian, the Aorist and synthetic reduplicative Perfect of Old Iranian were replaced by a new system of analytic constructions based on the copula *h-* 'be' (*h-am* 'I am', *h-ē* 'you are' *ast* 'he is') in combination with the PP form in *-ta* (< adjectival form *-t-ag* < * -t-ak). With intransitive verbs such as *šaw-am* 'I go' the past (Preterit) was *šud h-am* 'I went' and its past counterpart (Pluperfect), *šud būd h-am* 'I had gone', was formed by adding the PP of the copula *būd* 'been'. With transitive verbs in late Middle Persian texts, the analytic 'be'-Perfect *u=m kard h-am* 'and I have done' is derivationally linked with the passive in exploiting the copula auxiliary in conjunction with the PP form of the main verb. We observe that in the Perfect *u=m kard h-am* the copula is in the 1st sg. (lit. of-me done I-am, i.e. 'I am the (one) who has done') and guarantees thus an unambiguous active interpretation of the former passive construction *manā kart-am* 'done by me' (> *u=m kard*) as shown in (3):

(3)   *ud   pas   man   namāz   burd   ham   ud   goft*
      and   then   I.OBL   homage   bring.PP   be.1SG   and   say.PP
      'and I bowed and said'   [Ardā Wirāz Nāmag 14.6]

In Early New Persian the Perfect/Passive auxiliary, copula *h-* 'be', was replaced by another auxiliary, namely the stative verb *īstādan* 'stand' (< * st(h)ā 'to stand'), exemplified in (4):

(4)   Passive Perfect of transitive verbs (based on *kard būd* do.PP be.PP 'it was done'):
      (Perfect)   *kard būd* 'it was done'
      >   *kard est-ēd* (do.PP stand.3SG) 'it has been done'
      (Pluperfect) *kard būd ast* 'it had been done'
      >   *kard est-ād* (do.PP stand.PP is) 'it had been done'

It is of fundamental importance to realize that the same auxiliary *īstādan* 'to stand' was used for the formation of both the Perfect of intransitive verbs and the passive of transitive verbs, as shown in (5):

(5)   Active Perfect of intransitive verbs (based on *šud est-* go.PP stand.PP. lit. 'gone stood')):
      (Perfect)   *šud est-am*   (go.PP stand.1SG)   'I have gone'
      (Pluperfect) *šud est-ād h-am*   (go.PP stand.PP be.1SG)   'I had gone'
      (*je suis allé*)
      (*j'étais allé*)
      Passive Perfect of transitive verbs (based on *nibišt est-* (write.PP stand.PP. lit. 'written stood')):
      (Perfect)   *nibišt est-ēd* (write.PP stand.3SG) 'it stands/has been written'
      (Pluperfect) *nibišt est-ād* (write.PP stand.PP)   'it stood/had been written'

This state of affairs is a result of the loss of mediopassive (middle) voice morphology of Old Persian. As a parallel for the polysemy of the stative verb *īstād-an* 'to stand' and the copula *(h)ast-an* 'to be' (i.e. as both the active Perfect marker and the passive marker), we can mention the so-called 'deponent' verbs in Latin, where a number of verbs of 'middle' semantics (such as *speak*, *admonish*, etc.) use the passive morphology to form their active Perfect; contrast: *vir locū-tu-s est* 'the man has spoken' (Active Perfect) with *vir laudā-tu-s est* 'the man is (has been) praised' (Passive Perfect).

In Examples (4)–(5), note the co-existence of two formations of the Pluperfect exploiting two different past participles for marking the [+Past], namely the PP of 'to be' *būd* (be.PP) and the PP of 'to stand' *estād* (stand.PP). The 'stand'-Perfect (in the present) was realized by the present tense forms of *īstādan* 'stand' and the Pluperfect (past Perfect) by the attachment of the present tense forms of the copula to the PP *estād*. The Pluperfect of the 'be'-Perfect used the same strategy of attaching the present tense forms of the copula to the PP *būd*. All these Middle Persian formations in the singular are surveyed in Table 3. (based on Skjærvø 2009, p. 219)

Table 3. Middle Persian Perfect and Pluperfect

Perfective	'Stand'-Perfect	'Stand'-Pluperfect	'Be'Pluperfect
1SG *šud h-am*	*šud est-am*	*šud estād h-am*	*šud būd h-am*
2SG *šud h-ē*	*šud est-ē*	*šud estād h-ē*	*šud būd h-ē*
3sg *šud-Ø*	*šud est-ēd*	*šud estād-Ø*	*šud būd-Ø*
'he went/has gone'	'he has gone'	'he had gone'	'he had gone'

The past counterpart of the ambiguous perfective *šud h-am* 'I have gone' (Perfect) ~ 'I went' (Preterit) was formed by adding the PP of the copula *būd* 'been' (corresponding to the Slavic *l*-participle *byl*). The 'be'-Pluperfect 'he had gone' is called "past preterit" by Skjærvø (2009, p. 219). Observe that the same strategy of using two past participles for the formation of the Pluperfect, *šud būd h-am* 'I had gone', is found in West Slavic languages. For instance, the Czech Pluperfect *byl js-em šel* 'I had gone' features two past participles, the auxiliary *byl* and main verb *šel* corresponding to Middle Persian *būd* and *šud*, respectively.

In later development of the simple past (Preterit), the copula came to be cliticized to the PP of the main verb: New Persian *šud=(h)am* 'I went' > lexicalized as *šod-am* 'I became'. In typological terms the same process is observed in the history of Polish *šeł=jeś-m* > *šł-em* 'I went' (see Andersen, 1987, 2006).

At the end of the Middle Persian period the 'stand'-Perfect was discontinued (Khanlari, 1986, p. 249) but note that Tajik continues to exploit the auxiliary 'stand' for the formation of the progressive aspect (*raft-a istod-a ast* (go.PP stand.PP is) 'he is going'), to be discussed in connection with (21). The 'be'-Perfect was rebuilt

on the basis of an innovative perfect stem in *-a*, *kard-a*, and the PP *būd* was grammaticalized as the past tense marker [+P] (to be dealt with in Section 3). During the New Iranian period, the passive construction was rebuilt again, this time by means of the verb *šod-an* as an auxiliary (grammaticalized lexical verb 'to go'). In contemporary New Persian, it functions as a full-fledged passive auxiliary (*kardé šod* 'it was done'). As such the form of the innovative perfective participle *kard-é* is polysemous between the active perfective participle (*kard-é=am* in the 'be'-Perfect, lit. I am (the one) having done) and the passive perfect participle (*kard-é* become. PP) as in *kard-é šod-é bud* (do.PP become.PP be.PP) 'it had been done'.

3.   New Iranian

There is considerable variation in the shape of the innovative perfective participle in the Iranian languages. In New Persian the form *kard-é* goes ultimately back to Proto-Indo-Iranian *kr-tá (> *kar-tá > *kard-á > kard-é). The final stress in New Persian *kard-é* continues the PII state of affairs (preserved in Vedic *kr-tá*). According to Khanlari (1986, p. 250) the analytic Perfect is based on the adjectival form *kard-ag* in an attributive structure; we understand this innovation as an instance of the grammaticalization of a derivational morpheme *-ag*. This attributive adjectival form was recategorized as the new perfect stem to which the forms of the copula were added: *kard-ag ast*. This new PP form is documented in Balochi *kurt-a(g)* 'done, *šut-a(g)* 'gone'; the PP in *-ig* is documented in Sorani (Sinayi) *wut-ig-* 'spoken' (< * *wuk-t-ig*) and there are several other g-less PP forms such as *kird-ū-* 'do' (< * kir-d-ug) in Mokri and Suleimani, *kird-ī-* in Kirmanshahi (Mokri, 2003, p. 328–9). A remarkable form is found in the Gurani perfective participle *karda-n-* (*karda-n-im* 'I have done'), which appears to have been enlarged by *-n*.[1]

3.1   Early New Persian

3.1.1   *Canonical 'be-Perfect'*
As was shown above (in Table 3) the Middle Persian 'be'-Perfect was ambiguous between the perfective (Preterit) and perfect reading (*šud h-am* 'I went' ~ 'have gone'). With the demise of the 'stand'-Perfect (*šud est-ēd* 'he has gone'), it became essential to renew the 'be-Perfect in its perfect reading. The auxiliary *h-* 'be' continued

---

1.   However, it is more likely that the Perfect in Gurani was built on the infinitive *kard-an* 'to do' > perfect *kard-an-im* 'I have done' as in Pashto (East Iranian) where the Perfect is clearly built on the infinitive in *-ə́l* as in *dared-ə́l-ay yəm* stop.INF-ADJ be.1SG 'I have stopped' (*-ay* is the adjectival suffix comparable with West Iranian *-ag*).

being used and attached to a P(ast) P(article) or to a PP enlarged by *-a* (< *-ag*). The picture emerging from the large-scale analysis of the Early New Persian texts of the 10th–11th c. by Lenepveu-Hotz (2012, p. 123–135, 212–235) is summarized in Table 4.

Table 4. Compound forms in Early New Persian Texts (10th–11th c.)

		Imperfective	Perfect
Present	*kunad*	*(ha)mē kunad*	*kard(a) ast*
Past	*kard*	*(ha)mē kard*	*kard(a) būd*
Evidential/Distant past	*kard(a) ast*	*(ha)mē kard(a) ast* ?	*karda būda ast* ?

(The formation *(ha)mē kunad* > *mēkunad*, ancestral of New Persian imperfective *mi-konad* 'he does', arose by the grammaticalization of adverb *hamēw* 'always).

   During the following centuries (Lenepveu-Hotz 2012, p. 235–6) the use of the plain PP (*kard*) was discontinued and in the texts composed during the 13th–14th centuries the Perfect is based on the PP enlarged by the suffix *-a*: *kard-a ast* 'he has done' (Pluperfect *kard-a būd* 'he had done'). While Lenepveu-Hotz (2012, p. 128–130) in her analysis of Early New Persian texts has not observed any semantic difference between the Perfects based on the plain PP (*kard ast*) and the PP enlarged by *-a* (*kard-a ast*), both 'he has done', a significant difference can be observed in Early Judaeo-Persian texts (of 8th–12th c.) analyzed by Paul (2013). Their importance for the study of the transition from the Middle Persian state of affairs to the Early New Persian stage lies in their independent development of Early New Persian for several centuries and in their preservation of a number of dialectal features not found in ENP. On the whole, the EJP is closer to MP than is ENP, and it may be considered "an important dialectal link between MP and ENP" (Paul, 2013, p. 11; Lazard, 1987). There are two analytic Perfects (Paul's "past tenses") in Early Judaeo-Persian formed either (i) with the plain PP (*rasīd* arrive.PP 'arrived') or (ii) with the PP in -a (*rasīd-a*). The copula in the 3SG is *hest*. The forms in (6) are "more regular past tenses" while those in (7) "still show an adjectival static character" (Paul, op. cit. 131–135).

   (6)   *nibišt hest* 'he has written' (Perf), *nbyšt bwd* [nibišt būd] 'he had written' (Plpf)

   (7)   *nibišt-a hest* 'it is written' (Perf), *nbyšt' bwd* [nibišt-a būd] '(the words) were written' (Plpf)

3.1.2   *Dialectal / 'Nishapuri' Perfect* kard-ast-am *'I have done'*
In Early New Persian poetic and historical texts (Attār, Bal'ami, Beyhaqi, Tabari) there appears another analytic Perfect with the 3SG form of the verb 'to be' /æst/ inserted between the personal suffix and the past base. Some Iranian scholars

(Khanlari, 1986; Ahmadi Givi, 2005) maintain that these forms arose by the full conjugation of the verb 'to be' *astan/ astidan* plus the PP *kard*. Some scholars consider them to be typical of the dialect of Nishapur (Bahar, 1942/1994) or simply dialectal (Khanlari, 1986). According to Lazard (1963) and others (Utas, 2000; Paul, 2008), they are typical of northeastern regions and represent a relic of the Middle Persian 'stand'-Perfect (with the auxiliary *īstādan* 'stand') and not the 'be'- Perfect. Paul (2013, p. 134) reports two examples from Early Judaeo-Persian as typical of northeastern New Persian, one of them displaying the form *-ist-* (kt'b' ḥtmy *krdy-stym* [kard-ist-ēm] 'we have sealed the book'). It would seem to us that in other dialectal regions, these forms could have arisen on the basis of the 3sG form of the 'be'-Perfect *kard-ast* 'he has done' by adding personal suffixes to it: *kard-ast-am*, *kard-ast-ī*, *kard-ast-īm*, *kard-ast-īd*, *kard-ast-and*. Several salient examples of these formations are presented in (8).

(8)   *to   midāni   ke goft-ast-am to   rā   rāz*   (Attār, 11th c.)
       you IPFV-know-2sG that say-COP-1sG you DAT secret
       'You know that I have told you the secret'
       to   eydar mānd-ast-i   (Bal'ami, 9th c.)
       you   here   stay- COP-2sG
       'You have stayed here'
       barkat   kard-ast-im   (Tabari, 9th c.)
       blessing do- COP-1PL
       'We have blessed [a bounty]'   (examples from Ahmadi Givi, 2005)

The strategy where the unmarked form of the 3rd sG functions at the pivot for the rebuilding of the whole paradigm is covered by Kuryłowicz's second 'law' of analogical extension (1947, modified by Harris and Campbell, 1995, p. 97). Fairly well-known examples are available from Persian (*ast* 'is' > *(h)ast-am*, *(h)asti*, …); Polish (*jest* 'is' > *jest-em*, *jest-eś*, …) and other languages.

3.2   The analytic Perfect in *-Vg* in Kurdish, Balochi and 'conjectural' mode in Tajik

As mentioned above the suffix of the innovative perfect stem appears in several allomorphs (*-ig* in Kurdish Sorani (Sinayi) as in *wut-ig-m-a* 'I have said' (Mokri, 2003, p. 334–5); there are several *g*-less PP forms such as *kird-ū-* 'do' (< * kird-ug) in Mokri and Suleimani, and *kird-ī-* in Kirmanshahi (Mokri, 2003, p. 328–9). In some Iranian languages and dialects, such as Zāboli, Bakhtiyāri (the dialect of Ardal) and Varāmini (the dialect of New Persian), this suffix appears as *-ak* (Kalbasi, 2004). In Balochi (Eastern Iran) there is a suffix *-ag* in the same function:

kapt-a(g) 'fallen', *kurt-a(g)* 'done'. The *-g-* in Balochi is optional but it appears consistently in the formation of the Pluperfect *kapt-ag-at-un* 'I had fallen' (Jahani & Korn, 2009, p. 665–6). According to Windfuhr & Perry (2009, p. 453) Standard Tajik features *g*-less forms through the whole paradigm but Kalbasi (1995, 2004) also provides examples of the Perfect built on the PP in *-ag* from Tajik dialects. In Standard Tajik, the *-ag* appears in the formation of the 'conjectural' mode based on the PP in *-ag-ī* and the forms of 'to be' (to be dealt with in Section 4). There are two sets of forms: one formed with the reduced form of the copula (*ast > -st*) as in *[raft-agī]-st* 'I suppose he went' and the other with the copula *-yam, -yī, y-em, -yed, -yand*. All these forms are surveyed in Table 5.

Table 5. Formation of the Perfect in Balochi and the 'conjectural' mode in standard and dialectal Tajik

	Balochi	**Tajik**	**'Conjectural' mode**	**Dialectal Tajik**
1SG	*kapt-a(g)-un*	*raft-a am*	*raftagī-st-am*	*raftagī-yam*
2SG	*kapt-a(g)-ay*	*raft-a ī*	*raftagī-st-ī*	*raftagī-yī*
3SG	*kapt-a(g)-Ø*	*raft-a ast*	*raftagī-st (zero)*	*raftagī-st*
1PL	*kapt-a(g)-an*	*raft-a em*	*raftagī-st-em*	*raftagī-yem*
2PL	*kapt-a(g)-it*	*raft-a ed*	*raftagī-st-ed*	*raftagē-yed*
3PL	*kapt-a(g)-an*	*raft-a and*	*raftagī-st-and*	*raftagī-yand*

3.3   New Persian

The development of the New Persian analytic Perfect *kard-é am* 'I have done' is summarized in (9):

(9)	Middle Persian	Early New Persian	New Persian	Colloquial
1SG	*kard h-am*	*kard(a) am*	*kard-é am*	*kard-ä́m*
	do.PP be.1SG	do.PP 1SG	do.PP 1SG	do.PP.1SG
2SG	*kard h-ē*	*kard(a) ī*	*kard-é ī*	*kard-í*
3SG	*kard-Ø*	*kard(a) ast*	*kard-é ast*	*kard-é́*

In Colloquial New Persian, the final *-e* of the Perfect stem *karde* contracts with the initial vowel of the suffix of the copula yielding distinct forms of the present Perfect *kard-ä́m* 'I have done', *kard-í* 'you have done', *kardé* 'he/she has done', while the forms of the Preterit (past perfective) are accented on the root: *kárd-am* 'I did', *kárd-ī* 'you did' *kárd* 'he/she did' (see Ziamajidi 2019). The minimal pair of sentences is displayed in (10):

(10)   *man ghazā-m o   xórd-am*
     I   food=my ACC eat.PP-1SG
     'I ate my food [at 4 p.m.]'   (past perfective)
     man ghazā-m o   xord-ā́m
     I   food=my ACC eat.PERF-1SG
     'I have eaten my food'   (present Perfect)

In Sorani Kurdish, there is an identical construction with the auxiliary 'to be' reduced to the particle *-æ* which is not conjugated. But unlike in New Persian, this reduced form of the copula is attached to the finitized form of the perfective participle throughout the entire paradigm (examples from Rokhzadi, 2000):

(11)   *xʷend-um-æ* 'I have read'   [Sorani]
     xʷend-ut-æ 'you have read'

Similarly, the formation of the Perfect in Luri parallels that of Kurdish in that the reduced form *-ε* of the copula is attached as the particle to the finitized form of the perfective participle (examples from Soleymani & Haghbin, 2015):

(12)   *srəv-εn   hard-ij-ε*   [Luri, Bālāgerive dialect]
     apple-ACC eat.PP-2SG-COP
     'You've eaten the apple'

3.4   Sequencing of morphemes expressing tense and person/number

In some of the West Iranian languages, the present Perfect is formed by the auxiliary 'to be' conjugated in the present or its reduced form *-æ* attached as the particle (degrammaticalized copula) to the finitized form of the perfective participle. The difference between the formation of the Perfect in New Persian and 'weakly ergative' Kurdish dialects (Haig, 2008) lies in the sequencing of morphemes expressing the tense and person/number in the Perfect and Imperfect. In New Persian the person/number markers appear as suffixes in both aspectual categories while in Kurdish various imperfectivizing particles host the person/number markers before the past stem as shown in (13) in examples from Sorani: *a-m-wut* IPFV-1SG-say 'I was saying' (Imperfect) vs. *wut-ig-m-a* say.PP-1SG-COP 'I have said'. This difference is, of course, a consequence of the fact that Kurdish dialects finitize their PPs by means of the possessive suffixes, which attach directly to their possessor, while New Persian possessive clitics may only attach to nouns and prepositions.

(13) Present Perfect in New Persian and Kurdish
         New Persian   Kurdish (Sorani)
      Perfect   Imperfect   Perfect   Imperfect
      goft-é am   *mi-goft-am*   *wut-ig-m-a*   *a-m wut*
      goft-é i   *mi-goft-i*   *wut-ig-t-a*   *a-t wut*
      goft-é ast *mi-goft*   *wut-ig-y-a*   *a-y wut*

The past Perfect (Pluperfect) is formed in a parallel fashion by attaching the past tense forms of the copula with a differential treatment of intransitive and transitive verbs in Sorani Mokri and Suleimani (Mokri, 2003, p. 336; McCarus, 2009, p. 610):

(14) Intransitive verbs   Transitive verbs
                                                 [Sorani Mokri and Suleimani]
      1sg *hāt-i bū-m* 'I had come''   *xwārd-i bū-m* 'I had eaten' (Plpf)
      2sg *hāt-i bū-y(t)* 'you had come' *xwārd-i bū-t* 'you had eaten' (*-t*)
      3sg *hāt-i bū* 'he had come''   *xwārd-i bū-y* 'he had eaten' (*-y*)
      1pl *hāt-i bū-yn* 'we had come' *xwārd-i bū-mān* 'we had eaten' (*-mān*)

An interesting example of 'reversing' the order of the sequence of person/number and tense markers for non-3sg Pluperfect and present Perfect forms has been reported by Paul (2013, p. 134) in his study of Early Judaeo-Persian texts. In some instances (15) the person ending is directly attached to the verbal stem and the degrammaticalized pp *būd* follows, with no parallels to this order in Middle or (Early) New Persian:

(15) *s'lt'n kyrdwm bwd* [kerd-om būd] 'I had asked questions'
      mwlk gryptynd bwd [giript-and būd] 'they had seized kingship'

But there are also less common instances (in (16)) of the ending attached to *būd* conjugated as an auxiliary:

(16) *hmgyn nbyšt bwdwm* [nibišt būd-om] 'I had written it all'

The reduction of the person-number inflection and the generalization of the 3sg form as an 'impersonal' (subjectless) predicate is referred to as 'degrammation' by Andersen (2010, p. 144–5) with a number of parallels from Russian, South Slavic, and Greek. As a parallel to (15) we may quote the Russian Pluperfect of the type *ona zasnul-a* (f) *byl-o* (n) 'she had just fallen asleep' (Turgenev) with the degrammaticalized pp *byl-o* 'was'.

4.   Evidential ('non-witnessed') subsystem

Given the different distribution of evidential values in language-specific grammatical systems, it is difficult to characterize the evidential mode in universal terms. Its implementation varies all the way from lexical renderings such as 'apparently, allegedly, it is said' to grammatical (sub)systems using (paradigmatic) morphology. Evidentials in the Balkan languages (Bulgarian, Macedonian, Albanian) and in the languages of Western Asia (Turkish, Persian, Kurdish, Luri, Tajik, etc.) possess special forms with an evidential register absent in 'neutral' discourse. While the neutral register does not say anything about the source of information, the evidential register expresses an action or state as assumed/inferred or reported/dubious by the narrator. In terms of morphology all these evidential forms are based on the perfect stem. (We may remind ourselves that the evidential interpretation of the Perfect was recognized at the dawn of linguistic science in Ancient India by Pāṇini who defined the inflectional Perfect as *parokṣe* 'unnoticed by the eyes'). Contrast Osmanli Perfect *gel-miş-im* come.perf-1sg 'I have come' with its evidential counterpart *gel-miş i-miş-im* come.perf be.ev-1sg 'I am said to have come/they say that I have come'. In the grammars of New Persian, it is stated that the simple past is used when the speaker is or was on the scene and is the witness of what happens, but when the event is reported through someone else, the present Perfect is used (see Sadeghi & Arjang, 1979; Ahmadi Givi, 2005, p. 134–35). Ahmadi Givi (*ibid* p. 135) refers to the evidential/inferential Perfect by the 'Pāṇinian' term *gozašteye nadide* 'unseen past'. It is maintained that there are two major differences between the simple past and the present Perfect: (i) in the simple past the speaker usually witnesses the event, but in the present Perfect the speaker knows it by hearsay or inference; (ii) in the simple past the time of the event is specified but in the Present Perfect it is not (Ahmadi Givi, 2005, p. 147); the latter example is in (17):

(17)   *Širāz raft-id ?*   vs.   *Širāz rafté-id ?*
        Shiraz go.p-2pl (past)   Shiraz go.pp-2pl (present Perfect)
        'Did you go to Shiraz?'   'Have you ever gone to Shiraz?'
        (last week, last year)

In New Persian the Perfect form *rafté ast* 'he is gone' (*il est parti*) is actually ambiguous (or rather 'polysemous') between the retrospective Perfect 'he has gone' and evidential perfective 'he is/was said to be gone' or 'I infer that he is/was gone' (or to use Windfuhr's terminology between the Perfect and inferential Aorist). As in all the Balkan and Asian languages the evidential developed from the various usages of the Perfect: hear-say, inference, remoteness, retrospectivity (Lazard, 1989, p. 272). Its earliest attestations are seen in certain Early New Persian historical texts of the 13th c. displaying 'surcomposé' forms of the type *gurēxta būda ast* flee.pp be.pp

is '(as reported) he had fled' (Lenepveu-Hotz, 2012, p. 230). In New Persian, the evidential mode is available in the three basic aspectual categories (imperfective, perfective and retrospective) and also in the progressive aspect. As shown in (18) it is formed by adding the copula to the neutral forms (in spoken Persian, the copula in the 3rd person can be omitted):

(18)		Neutral Categories	Evidential
	Imperfect	*mi-raft*	*mi-rafte (ast)*
	Progressive (Past)	*dāšt mi-raft*	*dāšte mi-rafte (ast)*
	Aorist	*raft*	*rafte (ast)*
		'he went'	'(apparently)he went'
	Perfect	*rafte ast*	*rafte bude (ast)*
		'he is gone'	'he is said to have gone'

It is observed that the formation of the innovative evidential Imperfect (*mi-rafté=ast*) parallels that of its non-evidential counterpart (i.e. by adding the indicative/imperfective prefix *mi-* to the perfective form *raft*).

An example of the evidential mode in the imperfective aspect in (19) is quoted from Lazard (1989, p. 272):

(19) *ba'd … Nane dide bud-aš bâz-ham jelow-e xâne*
afterwards Nane see.PP be.PP=him again in front-EZ house
râh mi-rafte
road PROG-go.PP
'Afterwards Nane had seen him, who again walked in front of the house'

The narrator knows about the presence of the person of the house through Nane; if he himself witnessed him he would have used the past imperfective form *mi-raft* instead of the evidential *mi-rafté*.

Example (20) provides an illustration of the evidential Perfect progressive, where the speaker reports that he has heard that Ali had an accident:

(20) *Ali dâšte mi-raft-e ke tasâdof kard-e*
Ali have.PP IPFV-go.PP that accident make.PP
'While Ali was reportedly going, he had an accident'

If the speaker himself had witnessed Ali's accident, he would have used the past progressive: *Ali dâšt mi-raft ke tasâdof kard* 'While Ali was going, he had an accident'.

It should be mentioned that New Persian does not form an evidential counterpart to the verb in the progressive aspect in the present tense (*dâr-am mi-kon-am* 'I am doing'). However, both options are available in Tajik, which forms its progressive aspect by a different auxiliary *istod-an*, grammaticalized lexical verb *istād-an* 'stand' (Windfuhr & Perry 2009, p. 464).

(21) Progressive Evidential 'durative'   [Tajik]
       Present   *[raft-a istod-a]-ast*   *me-[raft-a]-ast*
                   'he is going'   'he is (evidently) going, goes, will go, used
                                              to go'
                   Progressive past   Evidential
       Past   *[raft-a istod-a]-bud*   *[raft-a istod-a]-bud-a ast*
                   'he was going'   'he is/was (evidently) going'

Notice the ambiguity of the formation *me-[raft-a]-ast* (evidential Present progressive ~ evidential Imperfect) corresponding to New Persian *mi-rafte ast* in (18). Observe also that in the Tajik progressive aspect the auxiliary *istod-an* follows the main verb in *[raft-a istod-a]-ast*, while in the double finite construction of New Persian, the auxiliary in *dārad mi-ravad* 'he is going' precedes the main verb.

In addition, Tajik forms its related category of 'conjectural' mode (originally a peculiarity of Northern dialects "assimilated" into modern literary Tajik) by attaching the degrammaticalized form of the copula *-st-* (< *hast*) to the perfect participle in *-ag-ī*. In Colloquial Tajik these forms are contracted with the copula *-st-* elided (Windfuhr & Perry 2009, p. 466).

(22) Conjectural mood in Tajik
       1SG *raft-agī-st-am* (contracted forms *raft-ag-em/-im*)
       2SG *raft-agī-st-ī*   (*raft-ag-ī/-i*)
       3SG *raft-agī-st*   (*raft-ag-e/-i*)

As is shown in (23) these formations (available in the past, present-future, and the present progressive) express "various degrees of supposition or conjecture" (Windfuhr & Perry 2009, p. 467):

(23) *ū   az   šahr [omad-agī]-st*   [Tajik]
       he from city [come.PP]-COP
       'he may have come from the city'
       ū   pagoh   [me-omad-agī]-st
       he tomorrow [IPFV-come.PP-PP]-COP
       'he'll probably come tomorrow'
       ū   [raft-a istod-a]-gī-st
       he [go.PP stand.PP]-PP-COP
       'he might be going'

The same process of adding the copula to the neutral forms to derive their evidential counterparts is observable in other West Iranian languages, Kurdish, Luri (Soleymani & Haghbin (2015)) and Bahtiāri (Taheri (2007)). Consider the following examples:

(24) *ženaft-em-a Ali čū-a bo širāz*
     'I heard that Ali had gone to Shiraz' vs.
     Ali čū bo širāz   [Kurdish, p.c.]
     'Ali went to Shiraz' (last week)'
     ešnaft-em-a Ali rat-a wa širāz
     'I heard that Ali had gone to Shiraz' vs.
     Ali rat wa širāz   [Luri (Bakhtiāri) p.c.]
     'Ali went to Shiraz'

5.   Typological parallels

The strategy of deriving evidential counterparts to basic aspectual categories by means of the copula is plausibly linked with the "probable influence of Turkic" (Windfuhr, 1982), i.e. language contact with Osmanli and Uzbek (Kononov, 1960) in the case of Tajik. The Osmanli 'model' is provided in (25):

(25) Basic categories   Evidential counterparts   [Osmanli]

Present	*gel-iyor-um*	*gel-iyor-muş-um*	'I am/was said to be coming'
Aorist	*gel-ir-im*	*gel-ir-miş-im*	'I am/was said to come'
Perfect	*gel-miş-im*	*gel-miş i-miş-im*	'I am said to have come'
	come.perf-1sg	come.perf be.ev-1sg	

In terms of morphology, Turkish forms unambiguous evidential counterparts for all aspects on their stems (Present *gel-iyor*, Aorist *gel-ir*, Perfect *gel-miş*) by adding the perfect marker *-miş* (called *miş*-past); the analytic evidential Perfect is formed by adding the evidential form of the copula (neutral *gel-miş-iz* 'we have come' vs. evidential *gel-miş i-miş-iz* 'we are/were said to have come'). New Persian in its elaboration of evidential counterparts to neutral aspectual categories relies on the perfect stem which is enlarged by the imperfectivizing prefix *mi-* (for the evidential Imperfect) and the auxiliary *bude* be.PP 'been' (for the evidential Perfect). Unlike Turkish with its more elaborate system of evidentiality, New Persian relies on the polysemy of the Perfect. In New Persian one form *rafte-im* – polysemous between neutral Perfect and evidential Aorist – may translate both Turkish examples of the neutral and evidential Perfect. The salient difference between these two systems – Osmanli in (25) and New Persian in (26) – lies in the appearance of the innovative Imperfect and the evidential progressive aspect in New Persian.

(26)

	[−Evidential]	[+Evidential]	[+Ev without cop]
Imperfect	*mi-raft*	*mi-rafte (ast)*	*mi-rafte*
Progressive	*dāšt mi-raft*	*dāšte mi-rafte (ast)*	*dāšte mi-rafte*
Aorist	*raft*	*rafte ast*	*rafte*
Perfect	rafte ast	rafte bude ast	rafte bude

While the evidential interpretation of the Aorist is based on the polysemy of the basic Perfect (*rafte ast* 'he is gone' (Perfect) or 'he is/was said to be gone' (evidential Aorist)), the evidential Imperfect apparently arose by analogy (defined as 'extension' by Harris & Campbell, 1995, p. 97) with the evidential Aorist without the copula. In other words, unlike the Perfect, the neutral Imperfect *mi-raft* 'il allait' (he was going) cannot be interpreted evidentially. This is shown in (27):

(27) *raft*   : *rafte*   [−EV] AOR : [+EV] AOR
      mi-raft : *mi-rafte* [−EV] IMPF : [+EV] IMPF

To conclude this section, we wish to provide an interesting typological parallel to this state of affairs from South Slavic languages (Bulgarian, Macedonian) where the "structural influence of Turkish seems to be indubitable" (Gołąb, 1960, p. 37).

(28)   Basic categories   Evidential counterparts [Bulgarian]
      Present   *píš-e* 'he writes'   –
      Imperfect   *píš-eše*   *píš-e-l*
      Aorist   *pis-á*   *pis-á-l e*
      Perfect   *pis-á-l e*   *pis-á-l*

In Bulgarian (in (28)), as in New Persian (in (26)), the evidential interpretation of the Aorist is based on the polysemy of the basic Perfect *pis-á-l e* write.PP be.3SG 'he has written' (Perfect) or 'people say that he wrote' (evidential Aorist). Also, as is shown in (29), the stem of the innovative evidential Imperfect apparently arose by analogy with that of the evidential Aorist:

(29) *pis-á* : *pis-á-l* [−EV] AOR : [+EV] AOR
      píš- : *píš-e-l* [−EV] IMPF : [+EV] IMPF

In Iranian at the end of the continuum of Turkish influence is Tāliši (Northwest Iranian) which modeled its evidential mood on the pattern of the Turkic evidential subsystem by featuring a single marker *ban* (adopted from the 3rd Pl form of 'to be' < *baon* 'they were'). Here, an interesting typological parallel can be provided from Cypriot Greek, which – in some registers – forms its evidential statements by adding the marker *mis(i) mou* adopted from the Osmanli evidential present tense form of the verb 'to be' *i-miş* (as in *ev-de i-miş-im* 'I am said to be at home'). In conclusion, the matters of the grammaticalization of evidentiality in Persian should be studied in terms of Persian/Turkish language contact (Johanson (2010)) and a typological context of the Balkan and Asian languages.

6.   Conclusions

In Middle Persian the Aorist and the reduplicative synthetic Perfect of Old Iranian were replaced by a new system of analytic constructions based on the copula *h-* 'be' (*h-am* 'I am', *h-ē* 'you are' *ast* 'he is') in combination with the PP form in *-t/d* (*nibiš-t* 'written', *kar-d* 'done'). The morphology of the innovative Preterit and Perfect categories (with intransitive and transitive verbs) was provided in Section 2. We explored the fundamental grammaticalization process whereby the forms of the auxiliary 'to be' were reduced to the suffixes of the innovative Preterit: *šud=(h)-am* 'I went' > *šod-am* 'I became', and that its past counterpart (Pluperfect) was formed by adding the PP of the copula: *šud būd=(h)am*.

Significant changes are observable in the texts composed in the following centuries discussed in Section 3. The use of the plain PP (*kard*) was discontinued and in the Early New Persian texts composed during the 13th–14th c. the Perfect is based on the PP enlarged by the suffix *-a*: *kard-a ast* 'he has done' (and Pluperfect *kard-a būd* 'he had done'). While no semantic difference between the Perfects based on the plain PP (*kard ast*) and the PP enlarged by *-a* (*kard-a ast*), both 'he has done', could be observed in Early New Persian texts, a significant difference between the two analytic Perfects was detected in Early Judaeo-Persian texts formed either (i) with the plain PP *rasīd* arrive.PP 'arrived' or (ii) with the PP in *-a*, *nibišt-a* write. PP-*a* 'written' The forms in (i) are regular active 'be'-Perfects while those in (ii) are used in passive constructions *nibišt-a hest* 'it is/has been written' (replacing earlier *nibišt-Ø estēd*) with their double-marked PP (write.PP-PP) displaying adjectival characteristics (< * nibiš.t-ag).

In Section 3.2 we observed that the 'degrammaticalized' forms of the copula (defined by Andersen (2010) in terms of the generalization of the 3SG form as an 'impersonal' (subjectless) predicate) were exploited in the formation of more complex aspectual and modal categories. In Balochi there is a suffix *-ag* forming the perfect stem (*kapt-a(g)-* 'fallen', *kurt-a(g)-* 'done'); while the *-g-* in the Perfect in Balochi is optional, it appears consistently in the formation of the Pluperfect *kapt-ag-at-un* 'I had fallen' when another element – the degrammaticalized form of the copula (*hast > at*) – is added: *kapt-ag-at-un* 'I had fallen', *kapt-ag-at-ay* 'you had fallen' etc. Another form of the degrammaticalized copula (*hast > st*) appears in the formation of the 'conjectural' mood in Tajik as in *[raft-agī]-st* 'I suppose he went'

In Section 4 we ascertained that the evidential subsystem developed from the various usages of the Perfect: hear-say, inference, remoteness, and retrospectivity. In its elaboration of evidential counterparts to neutral aspectual categories, New Persian relies on the perfect stem (*raft-e* 'gone') which is enlarged by the

imperfectivizing prefix *mi-* (for the evidential Imperfect) and the auxiliary *bude* be.PP 'been' (for the evidential Perfect of the type *rafte ast* 'he is gone'). Evidential counterparts to neutral forms are available in all the three aspectual categories (imperfective, perfective and perfect) and both voices, but there is no evidential counterpart to the verb in the progressive aspect in the present tense (*dâr-am mi-kon-am* 'I am doing'). Both options (evidential present and evidential past progressive) are available in Tajik, which forms its progressive aspect by a different auxiliary *istod-an*, grammaticalized lexical verb *istād-an* 'stand'. Tajik (in contact with Turkic Uzbeki) expanded its evidential subsystem by developing the 'conjectural' mood to express unsubstantiated conjectures/assumptions. It is formed by attaching the degrammaticalized form of the copula *-st* (< *ast*) to the past participle in *-ag* (*raft-agī-st* 'he may have gone'/' I suppose he went'). In Section 5 we concurred with the current assumptions regarding probable Turkish influence on the formation of the evidential subsystem in other West Iranian languages (Kurdish, Luri, Tajik). From a more expansive perspective of probable Turkic influence on inflectional languages, we provided the interesting typological parallel of the development of the inferential mood in the southernmost Slavic languages (Bulgarian, Macedonian).

To conclude, this paper has provided a longitudinal analysis of the rich data of the West Iranian languages in order to contribute to a deeper understanding of the mechanisms through which temporal and aspectual systems are maintained and restructured over long periods of time. A propos the matter of the grammaticalization of evidentiality in Persian (and other West Iranian languages), much more work remains to be done in terms of Persian/Turkish language contact and cross-linguistic comparisons with Balkan and West Asian languages.

Abbreviations

ACC	accusative	GEN	genitive	PLUPF	pluperfect
ADJ	adjective	IPFV	imperfective	PP	past participle
AOR	aorist	IND	indicative	SG	singular
AUG	augment	INF	infinitive	1	first person
COP	copula	PERF	perfect	2	second person
DAT	dative	PL	plural	3	third person
EV	evidential				

References

Ahmadi Givi, H. (2005). *Dastur e Zabān e Fārsi: Fe'l*. [The grammar of the Persian language] Tehrān: Našr e Qatreh.

Andersen, H. (1987). From auxiliary to desinence. In M. Harris & P. Ramat (Eds.), *Historical development of auxiliaries*, (pp. 21–52). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110856910.21

Andersen, H. (2006). Grammation, regrammation, and degrammation: tense loss in Russian. *Diachronica* 23, 231–258. https://doi.org/10.1075/dia.23.2.02and

Andersen, H. (2010). From morphologization to demorphologization. In: S. Luraghi & V. Bubenik (Eds.) *The Continuum Companion to Historical Linguistics*, (pp. 117–146). London/New York: Continuum.

Bahar, M. (1942/1994) (6th ed.) *Sabkšenāsi. Tārix e Tatavvor e Nasr e Fārsi* (3 vols) [Stylistics: The history of the evolution of Persian prose texts] Tehrān: Amir Kabir.

Estaji, A. & Bubenik, V. (2007). On the development of the tense/aspect system in Early and New Persian. *Diachronica* 24, 31–55. https://doi.org/10.1075/dia.24.1.03est

Gołąb, Z. (1960). The influence of Turkish upon the Macedonian Slavonic dialects. *Folia Orentalia* 1, 26–45.

Haig, G. (2008). *Alignment change in Iranian languages: A Construction Grammar approach*. Berlin/New York: Mouton de Gruyter. https://doi.org/10.1515/9783110198614

Harris, A. & Campbell, L. (1995). *Historical syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511620553

Jahani, C. & Korn, A. (2009). Balochi. In: G. Windfuhr (Ed.) *The Iranian languages*, (pp. 634–692). London/New York: Routledge.

Johanson, L. (2010). Turkic language contacts. In: R. Hickey (Ed.), *The handbook of language contact*, (pp. 652–672). Chichester, West Sussex: Wiley-Blackwell. https://doi.org/10.1002/9781444318159.ch32

Kalbasi, I. (1995). Tafāvothā ye vojuh e fe'li dar Fārsi ye Tājiki va Fārsi ye Irān [The differences in aspects of verbs in Iranian Persian and Tajiki Persian] *Majaleye Dāneškaye adabiyāt o ʔolum e ensāni dānešgāh e Tabriz*. No. 153 & 154, 155–170.

Kalbasi, I. (2004). Gozašte ye naqli dar lahjehā va guyešhā ye Irān [The Perfect in dialects and languages of Iran] *Guyeš šenāsi*, No. 2, 66–89.

Kent, R. (1953). *Old Persian*. New Haven, CT: American Oriental Society.

Khanlari, P. (1986). *Tārix e Zabān e Fārsi*. [History of the Persian Language] Tehrān: Našr e No.

Kononov, A. N. (1960). *Grammatika sovremennogo uzbekskogo literaturnogo jazyka*. [The grammar of Contemporary Literary Uzbek]. Moscow/Leningrad: Izdatel'stvo Akademii Nauk SSSR.

Kuryłowicz, J. (1947). La nature des procès dits analogiques. *Acta Linguistica* 5, 17–24.

Lazard, G. (1963). *La langue des plus anciens monuments de la prose persane*. Paris: Klincksieck [Etudes linguistiques, 2].

Lazard, G. (1987). Le judeo-persan ancien entre le pehlevi et le persan. In *Transition periods in Iranian history*. Paris: Association pour l'Avancement des Etudes Iraniennes Leuven: Peeters, XI, 263 [Studia Iranica, Cahier 6].

Lazard, G. (1989). Le persan. In R. Schmitt (Ed.) *Compendium linguarum orientalium*. (pp. 263–293). Wiesbaden: Reichert.

Lenepveu-Hotz, A. (2012). *Etude diachronique du système verbal person (Xe-XVIe siècles): d'un équilibre à l'autre?* Paris: Ecole pratique des hautes études.

McCarus, E. (2009). Kurdish. In G. Windfuhr (Ed.) *The Iranian languages*. (pp. 587–633). London & New York: Routledge.

Mokri, M. (2003). *Grammaire et lexique comparés des dialects kurdes*. Paris: Karthala.

Paul, L. 2008. Remarks on the evolution of the New Persian system of past tenses. In: Sh. Shaked & A. Netzer (Eds.) *Irano-Judaica VI, studies relating to Jewish contacts with Persian culture throughout the ages*, (pp. 185–195). Jérusalem: Ben-Zvi Institute.

Paul, L. (2013). *A grammar of Early Judaeo-Persian*. Wiesbaden: Reichert.

Rokhzadi, A. (2000). *The phonetics and the grammar of Kurdish*. Tehran: Tarfand.

Sadeghi, A. & Arjang, Gh. (1979). *Dastur e Zabān e Fārsi* [The grammar of the Persian language]. Tehran: Sāzmān e āmuzeš va parvareš.

Skjärvø, P. (2009). Middle West Iranian. In: G. Windfuhr (Ed.), *The Iranian languages*. (pp. 196–278). London / New York: Routledge.

Soleymani, A. & Haghbin, F. (2015). Tense and aspect in the Bālāgerive dialect of Lori. *Journal of Lingual Inquests*, 20, 245–262. Tehran: Tarbiat Modarres University.

Taheri, E. (2006). The verb in Bakhtiari, *Nāme ye Farhangestān*, 2, 86–108. Tehrān: Farhangestān e Zabān.

Utas, B. (2000). Traces of evidentiality in Classical New Persian. In L. Johanson & B. Utas (Eds.) *Evidentials: Turkic, Iranian and neighbouring languages*. New York: de Gruyter. https://doi.org/10.1515/9783110805284.259

Windfuhr, G. (1982). The verbal category of inference in Persian. *Acta Iranica* 22, 263–287.

Windfuhr, G. & Perry, J. (2009). *Persian*. In: G. Windfuhr (Ed.) *The Iranian languages*, (pp. 416–544). London/New York: Routledge.

Ziamajidi, L. (2019) A diachronic study of grammatical aspect in Persian. Tehran: Alzahra University (unpublished PhD dissertation).

On the grammaticalization of the *-(v)ši-* resultative in North Slavic

Andrii Danylenko

Pace University, New York

It is commonly accepted that the *-(v)ši-* resultatives in North Slavic are a product of the areal diffusion of a similar resultative formation in *-vęs-* (< *-u̯es-*) from Baltic. The author argues that the anterior grammaticalization of the *-(v)ši-* resultative in North Slavic has been largely an internally motivated process. The analysis is premised on the distinction of spoken vs. written discourse which preconditions the parallel use of finite and non-finite predication, including the resultative structures, in the history and dialects of North Slavic. The author argues that language contact was but one of several factors influencing the convergent tendencies in the use of resultatives in the Circum-Baltic zone and well beyond.

Keywords: resultative, spoken vs. written discourse, anterior grammaticalization, North Slavic, Circum-Baltic zone, sociolinguistic typology

1. Introduction

This article is devoted to the areal-typological and historical profiling of the *-(v)ši-* (< *-(vъ)ši-*) resultatives in North Slavic in comparison with similar formations in *-vęs-* in Baltic. I intend to prove that the development of Slavic resultatives, particularly in the northwestern Russian, some Belarusian, Polish, and Ukrainian dialects have been internally motivated and, although genetically related to their Lithuanian and Latvian counterparts, they are not an innovation spread from the Baltic area (cf. Drinka, 2017, p. 349–353; Wiemer, 2012, p. 76).

The article is structured as follows. After a brief overview, in Section 2, of the problems germane to the descent and the grammaticalization of the *-(v)ši /-vęs-* (< *-u̯es-*) resultatives in Slavic and Baltic, I focus on their historical and areal-typological sampling as reflected in both "written" discourse (Sections 3–3.2) and "spoken" discourse (Sections 4–4.3). In Section 5, I discuss sampling results with an emphasis on the sociolinguistic aspects of the grammaticalization of Slavic resultatives. General conclusions are offered in Section 6.

The reconstruction of the grammaticalization of the *-(v)ši-* resultatives in North Slavic is premised on the distinction of two types of discourse pressure or information flow in discourse which are, over long periods, characterized by different degrees of special background information of the narrated event that might be shared by members of a particular speech community. Thus, speaking about the interrelation of the ergative and accusative grammar, Du Bois (1987) singled out a type-independent Preferred Argument Structure which is determined by a certain pattern of information and grammar. Following Du Bois, I argue that language-internal phenomena as fundamental as the structuring of grammatical relations and type of predication in North Slavic, can be driven by forces arising out of a particular discourse, viewed as the aggregate of instances of language (Danylenko, 2002, p. 119–121).

I distinguish between "spoken discourse" versus "written discourse", whose prevalence in either non-standard or standard varieties with the respective text genres might have preconditioned the emergence of two parallel pathways of the anterior grammaticalization in the aspect-tense system in North Slavic (see Danylenko, 2018a, p. 380, 382). Thus, on the one hand, one can speak about the prototypical written discourse based on the use of finite verbal predicates (hereafter, FP) which historically lose the imperfect first, followed by the aorist, replacing them with the *l*-perfect (Andersen, 2009). The latter historically began to take on preterital status, thus becoming a host of aspect; as a result of this process, the *l*-perfect turned into the only marker of the past tense and led to the loss of the BE-auxiliary, both for copulas and for auxiliaries, as first attested in the vernacular ("spoken") usage of Old Novgorodian (Zaliznjak, 2004: 173, 350). On the other hand, the prototypical spoken discourse, which, according to Du Bois (1987, p. 839–840; cf. Orr, 1989) best coincides with an ergative-like alignment, facilitates the anterior use of non-finite verbal predicates (hereafter, NFP), including past active participles in **-u̯es-* (> Slavic *-(v)ši-* and Baltic *-vęs-*).[1] Historically, in the process of their anterior grammaticalization, the non-finite (participial) predicates tended to acquire auxiliaries and adapt to the new tense-aspect system of North Slavic (Danylenko, 2018a).

---

1.   Our understanding of spoken and written discourses resembles the interpretation of Tannen (1983, p. 3), according to whom, spoken discourse is highly context-bound, while writing is decontextualized; also cohesion is established in spoken discourse through para-linguistic and non-verbal channels, while cohesion is established in writing through lexicalization and complex syntactic structures. This distinction is reminiscent of Živov's (2017, p. 325–333) discussion of the written language versus the spoken language in the history of East Slavic and its written tradition. He believes that the syntax of "non-bookish" texts is different from that used in various registers of the bookish texts. In other words, changes within each type of the texts is not only evolutionary but primarily structural; this is why one can speak about rearrangement and redistribution of the linguistic means within either bookish or non-bookish texts as defined by written and spoken discourse, respectively (Živov, 2017, p. 326–327).

I argue that the aforementioned anterior grammaticalization pathways are parallel trends in the evolution of the primary (non-finite) predicate which historically branched off into the periphrastic construction BE + active *l*-participle and the predicative **-ṵes* participle (cf. Delbrück, 1897, p. 478–493); the latter, as has been mentioned, acquired auxiliation later, in the process of its grammaticalization. It is not surprising, then, that the *-(v)ši-* resultatives are not used with auxiliaries, for instance, in the Ostromir Gospel (1056–1057) written in the East Slavic redaction of Old Church Slavonic, while the canonical Church Slavonic manuscripts and older East Slavic chronicles have just a handful of examples of the *-(v)ši-* resultative with auxiliaries (Potebnja, 1888, p. 132).

The postulated anterior grammaticalization pathways for FP and NFP, respectively, can be represented tentatively in the following way:
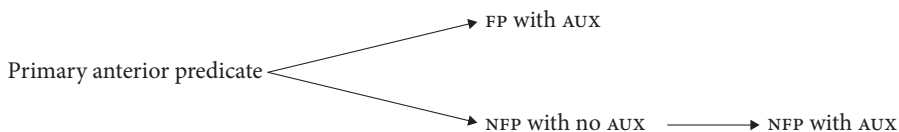
Scheme 1. Two pathways of the anterior grammaticalization in North Slavic

The NFP anterior grammaticalization pathway is typical of spoken discourse encompassing both the colloquial (vernacular) and non-bookish registers, while the FP anterior grammaticalization pathway is determined by written discourse which defines the bookish registers, although the registers can overlap in some narratives (Živov, 2017, p. 375). What is left to determine here is under what circumstances the NFP branching in the process of anterior grammaticalization occurred in Northwest Russian and some other northern Slavic languages and dialects. I home in on these circumstances in Section 5.

In Section 2 I offer a critical survey of the theory of the diffusion of the **-ṵes*-resultatives from Baltic to Slavic. This will allow me to outline an alternative explanation based on historical and areal-typological evidence from North Slavic in Sections 3–3.2 and 4–4.3.

2. Issues of the descent and grammaticalization of the *-(v)ši-* resultatives

In their comprehensive study of Baltic and Slavic resultatives, Wiemer and Giger (2005) explored the present-day distribution of the periphrastic resultative structures in Baltic and North Slavic, especially those built with the help of participles in *-n/-t* or *-(v)ši-/-vęs-*, with an eye to areal spread and the effects of grammaticalization. Drinka (2017) has recently used their findings in order to sort out the historical foundations

for the current, complex distribution of these structures, in particular the *-u̯es-* resultatives attested in the Circum-Baltic area (contact zone). Among those resultatives one can name the corresponding structures found in Lithuanian like (1a) and (1b) (Geniušienė & Nedjalkov, 1988, p. 370), in some Belarusian dialects like (2a) and (2b) (Wiemer & Giger, 2005, p. 54–55), and in North Russian, see (3) (Kuz'mina, 1993, p. 169–170); similar structures are found in Latvian (Endzelin, 1923, p. 758) and Latgalian based on the high Latvian dialect, see Example (2) (Nau, 2011, p. 50).[2]

(1) Lithuanian
   a. *Jis [yra] šiltai   ap-si-rengęs*
       he [is]   warmly PREF-REFL-dress.PAP.M.SG.NOM
       'He has dressed himself warmly'
   b. *Ji [yra] ap-si-vilk-usi   paltą*
       she [is] PREF-REFL-put.on-PAP.F.SG.NOM coat.F.SG.ACC
       'She has put on a coat'

(2) Latvian (Latgalian)
    Asmu   redzējs
    be-PRS.1SG see.PAP.M.SG
    'I have seen'

(3) Belarusian
   a. *adzeža   namokšy*
       clothing.F.SG.NOM drench.PAP.INDECL
       'The clothes are drenched'
   b. *my byli   dastau̯šy   hrošy*
       we be.AUX.PL.PST receive.PAP.INDECL money.PL.NOM/ACC
       'We have received money'

(4) Northwest Russian
   a. *Ja [byl]   vymyvšy*
       I   [be.AUX.SG.PST] wash.PAP.INDECL
       'I have washed'
   b. *pol   pomyvši*
       floor.M.SG.NOM wash.PAP.INDECL
       'The floor has been washed'

---

2.   I use basic diathetic concepts as developed by Russian scholars (Nedjalkov & Jaxontov, 1983; Trubinskij, 1983; Knjazev, 1983; Sobolev, 1998). Thus, Examples (1a), (2), (3a), (4a) can be called subjective or, subject-oriented resultatives, while Examples (1b) and (3b) are the so-called primary possessive resultatives, and Example (4b) is a typical object-oriented structure. Questions of the diathetic classification of the *-(v)ši-/-vęs-* resultatives are not discussed in detail in this article. The primary objective of this study is to delve into the diachrony and the areal distribution of independent (internally-motivated) grammaticalization of the *-u̯es-* resultatives in North Slavic.

One should add here another type of resultative structure, consisting of a HAVE auxiliary + PAP in combination with the accusative or genitive (under negation or for indefinite quantification) direct object. This type is found in Lithuanian, which has a transitive verb *turėti* 'to have', and some adjacent northern Slavic dialects which also use the HAVE verb, see (5) (Geniušienė & Nedjalkov, 1988, p. 385) and (6) (Adomavičute & Čekmonas, 1991b, p. 98 , also 1991a). Absent from Latvian, this HAVE construction in Lithuanian as exemplified in (5) is, according to Wiemer (2012, p. 72–73), a "natural consequence" of two conditions, which are the presence of a HAVE verb and the complementary voice distribution of participles in Lithuanian.

(5) Lithuanian
Jis turi   nu-si-pirkęs   žemės
he have.AUX.PRS.3SG PREF-REFL-buy.PAP.SG.NOM land.F.SG.GEN
'He has bought some land'

(6) Northeastern Borderland Polish
Čy vy juž   mac'e   kupifšy   plita?
PART you already have.AUX.PRS.2PL buy.PAP.INDECL gas.stove.F.SG.NOM
'Have you already bought a gas stove?'

Drinka's study of the periphrastic perfect in Europe, including the distribution of the *-u̯es-* resultatives in the Circum-Baltic area, is fully based on contact as a key element in the diffusion of the perfects and resultatives across the map of Europe.[3] With regard to the interrelations between Baltic and Slavic resultatives, she advanced, in particular, the following arguments:

First, only in Latvian and Lithuanian is the *-vęs-* participle used to form a regular, productive perfect which freely combines with other tenses and moods (Drinka, 2017, p. 352; cf. Wiemer & Giger, 2005, p. 111).

Second, in terms of areal distribution, the Belarusian dialects show productive use of the *-vši-* resultative in the north, but much more sporadic use in the south (Wiemer & Gieger, 2005, p. 40, 54). The productivity of the construction in dialects of Northwest Russian and North Belarusian lying closest to the Baltic languages supports, according to Drinka (2017, p. 353), the view of the Baltic area and the Baltic-Slavic Contact Zone as the epicenter of the *-u̯es-* resultative innovation.

Third, from the diachronic point of view, Drinka (2017, p. 353) insists that the *-vši* participle, though attested in Church Slavonic, usually appears there as a gerund or adverbial participle. She also agrees with the relative chronology of the formation of the *-vši-*resultative in Slavic as summarized by Wiemer & Giger (2005, p. 40, 57). According to this chronology, the forerunners if these resultatives appear

---

3.   For a critical survey of Drinka's book, see Danylenko (2018d).

in Novgorod as early as the 13th century; the Belarusian resultatives are first attested somewhat later, in the 15th century (Drinka, 2017, p. 434).

Fourth, Drinka (2017, p. 353) also mentions the use of the resultative construction by the "Old Believers" who migrated from the Novgorod and Pskov region into Lithuania, Latvia, Belarus', and northeastern Poland in the 17th–18th century (cf. Čekmonas, 2001).

To begin with, one can safely concur with Drinka's first argument since it is largely corroborated by Baltic dialect and historical data (Ambrazas, 1990, p. 186–191, 2006, p. 351–373). As far as the fourth argument is concerned, it appears circular. This argument does not support the role of Baltic as the epicenter of the *-ṷes-* resultative innovation. As always, Ukrainian might provide alternative argumentation. In Sections 4–4.3, the pertinent Ukrainian material proves crucial in refuting the contact-induced interpretation of the *-(v)ši-* resultative in Slavic. Suffice it to mention here the use of the *-(v)ši-* resultative in the resettled Russian dialects spoken in the northwestern part of Ukraine (Žytomyr region), see (7) (Samoxvalova, 1982, p. 133).

(7)   Russian (Žytomyr) dialect
      Ljudej   stol'ko   pъnaexatčy
      people.GEN so.many arrive.PAP.INDECL
      'So many people have arrived here'

Like the Old Believers in Lithuania, the speakers of these dialects might have first migrated from Pskov and Novgorod to Lithuania and later to Ukraine. There, these resettled dialects could have been influenced by local dialects of the northern, southwestern and transitional Ukrainian dialect types (Samoxvalova, 1982, p. 130, 151–152, 161). However, the alleged migration of the Russian speakers to the Ukrainian lands in the period of the Polish-Lithuanian Commonwealth does not necessarily prove that the Russian resultatives were originally replicated on the Baltic structure construed with participles in *-(v)ši-*. In general, the occurrence of *-(v)ši-* resultatives in the Russian resettled dialects spoken in both Lithuania and Ukraine brings us back to the source Russian dialects and needs interpretation as part of the Slavic dialect sampling.

The validity of the third, diachronic argument is probed in Sections 3–3.2 where I trace the continuity in the predicative use of the *-(v)ši-* participle in Slavic as reflected in some texts of the Old Church Slavonic canon and, likewise, in some religious and secular texts of the northern Slavic provenance. I argue that the anterior grammaticalization of the resultatives in North Slavic as hypothesized by Drinka and Wiemer & Giger (2015, p. 40) who based their historical insight on the works of Obnorskij (1953, p. 213) and Dmitrieva (1962, p. 158–160) looks patchy. For instance, the postulated 15th century emergence of the Belarusian resultatives

does not reflect the actual processes which took place, on the one hand, in the written discourse of (Old) Church Slavonic and, especially, its eastern Slavic recension and, on the other hand, in the spoken discourse of North Slavic. In fact, the 15th century as the cutting point in the development of Belarusian resultatives can hardly be taken for granted. Similar examples are found in texts written in Ruthenian, a vernacular standard used by both Belarusians and Ukrainians in the Grand Duchy of Lithuania and the Kingdom of Poland. It would be extremely difficult, if at all possible, to pinpoint the Belarusian origin of such constructions in the Ruthenian texts (cf. Danylenko, 2017). Consequently, the earliest attestations of such constructions in Ruthenian can potentially prove the existence of resultatives in both Belarusian and Ukrainian dialects as early as the 15th century. As I show in Section 4.3, this possibility is supported by the use of -(v)ši- resultatives in some modern Ukrainian dialects.

However, the most vulnerable is Drinka's thesis about the alleged spread of the *-ṷes- resultative from the Baltic area to Russian and Belarusian dialects. Thus, Drinka (2017, p. 352–353, 392) maintains that the Baltic languages form the nexus of the *-ṷes- innovations since it supports her principle thesis about the western origins of the HAVE resultative; she also claims that early Prussian German influence brought about the emergence of resultatives like the one exemplified in (5) in Lithuanian. Wiemer's (2012, p. 76) reasoning is more cautious in this case. He finds it difficult to ascertain how widespread a diffusion of this resultative from Lithuanian to Belarusian might have been. However, the occurrence of comparable constructions in other Slavic varieties is better explained, according to him, on independent grounds, mainly on the basis of direct German influence (as far as Kashubian and Pomeranian are concerned) or as a spontaneous development via analogical expansion as is the case of the Hanakian dialects of Moravia (Wiemer & Giger, 2005, p. 87–93).

Overall, the possibility of multiple influences as posited by Drinka and, though with some reservations, by Wiemer is open to challenges. At first blush, one can resort to the postulated existence of the Baltic substrate in the Belarusian-Lithuanian-Polish contact zone (Ambrazas, 1990, p. 191; Proxorova, 1991, p. 36–37; Mackevič & Hrynaveckene, 1993, p. 10). However, this scenario does not explain the mechanisms of the anterior grammaticalization of the *-ṷes- participle in either Baltic or Slavic, especially in those Slavic varieties which are found outside this zone. Nor did one offer a valid explanation of the diffusion of such constructions into the neighboring dialects of Belarusian, Lithuanian, and Polish (cf. Erker, 2014, p. 141).

3.   Evidence from written discourse

Sections 3–3.2 are concerned with both the predicative and nominative absolute usage of the *-ᵤes-* participle in written discourse.

3.1   Old Church Slavonic evidence

Old Church Slavonic examples of the independent (absolute) predicative use of the *-ᵤes-* participle are quite rare (Večerka, 1961, p. 93; Růžička, 1963, p. 226). Such examples have been treated either as (1) mere mistakes in translations from the Greek (cf. Marguliés, 1927, p. 140) or (2) indigenous structures (Večerka, 1961, p. 93).

   In total, around 10 examples are usually cited, depending on their syntactic, phonological, and orthographic features. Moreover, the explanation of predicative participles in *-ᵤes-* is commonly influenced by parallel Greek resultative structures or some other syntactic patterns (Večerka, 1961, p. 96, 1993, p. 92–92). Nevertheless, one finds several reliable Old Church Slavonic examples, which are repeated in different reference books. Remarkably, in various codices, absolute predicative participles in *-ᵤes-* with resultative meaning can be substituted by finite predicates in the same matrix clause. Here is one such example from the Codex Zographensis (Večerka, 1961, p. 94; Růžička, 1963, p. 227):[4]

(8)   Old Church Slavonic   (Luke 21: 38)
   *I   vsi ljudьe   iz   utra   **prixodęšte**   kъ nemu   vъ*
   and all people.NOM from morning.N.GEN come.PRAP3PL to him.GEN in
   crъkovъ   poslušatъ ego   (Lk 21:37–38)
   church.F.ACC listen.SUP him-ACC
   Greek
   kaì   pâs ho   laòs   ốrthrizen   pròs autòn ev tōi
   and   all the people come_at_dawn.IMPERF.3PL to   him   in the
   hierôi   akoúein   autoû
   temple listen.INF him
   'And all the people came early in the morning to him in the temple, for to hear him'

An additional example is provided in (9). Due to some orthographic ambiguities, this example looks dubious since the *-ᵤes-* participle used as an independent predicate can likewise be treated as an aorist form: *viděvъ* (PAP.M.SG) '(he) has seen' = *vidě* (AOR.3SG) 'he saw'. The latter reading is corroborated in the Codex

---

4.   For translation, I am using here and hereafter the Authorized King James Version of the Holy Bible.

Zographensis by the occurrence of the preposition *vъ* at the beginning of the next line following the use of the aorist *vide* in place of the participle *vidĕvъ* (Jagić, 1883, p. 167; Růžička, 1963, p. 228):

(9) Old Church Slavonic   (Mark 12:41)
I   **sĕdъ**   *i(isu)sъ*   *prĕmo gazofilakiovi.* **vidĕvъ**   *kako*
and sit.PAP.SG.M Jesus.M.NOM straight treasury.F.DAT see.PAP.M.SG how
narodъ.   *mĕštetъ   mĕdъ   vъ gazafilakijǫ. i   mnozi*
people.M.NOM cast.PRS.3SG copper.F.ACC in treasury.F.ACC and many
bogatii.   *vъmĕtaaxǫ   mnogaĕ*   (Mk 12:41)
rich.PL cast.IMPERF.3PL many.PL
Greek
kai **kathísas**   *katénanti   toû gazophulakíou*
and sit/M.SG.NOM.PART.AOR1 over.against the treasury
etheṓrei   *pỗs   ho óchlos bállei chalkòn   eis   tò*
behold.IMPERF.ACT.3SG anyway the people cast copper.money into the
gazophulákion kaì   polloì ploúsioi éballon pollá
treasury   and many rich cast much
'And Jesus, sat over against the treasury, and beheld how the people cast money into the treasury: and many that were rich cast in much'

Like (8), Example (9) can also be treated as a case of the nominative absolute construction with the subject distinct from the matrix subject (Potebnja, 1888, p. 194–197). Discussing examples from the East Slavic written discourse, Živov (2017, p. 424) argued that the nominative participle can be viewed as a variety of the predicative resultative participles since both cases represent independent predication in clauses, linked with the help of a coordinating conjunction like *i* 'and' as found in the following citation from the Codex Suprasliensis (Večerka, 1961, p. 107).

(10) Old Church Slavonic   (Supr 37: 7–8, *Vita Cononis martyris in Isauria*)
I   povelĕvъ   imъ priti   kъ sebĕ.   i   prišedъše
and order.PAP.M.SG them come.INF to him(self) and come.PAP.M.PL
stašę   prĕdъ njimъ
stay.AOR.3PL before him
Greek
kaì **kaleúsantos**   *autoū*   **ēlthon**   *kaì*
and call.PART.ACT.AOR1.M.SG.GEN himself.GEN come.3.PL.AOR.2 and
paréstēsan
come_to_the_side.3PL.AOR1
'And he ordered them to come to him, and they came and stood in front of him'

Večerka (1961, p. 98, 107) also surmised that, in terms of prosodic cues like pauses and changes in intonation which are likely to reduce or remove ambiguity, it is practically impossible to delimit cases of the predicate resultatives from the use of the nominative absolute.

Particularly noteworthy are parallel absolute constructions in the written discourse of Old Church Slavonic and its later recensions, for instance, the nominative absolute in Example (11a), excerpted from the Book of Savva, and the dative absolute in Example (11b) found in the Codices Zographensis, Marianus, and Assemanianus (Večerka, 1961, p. 109).

(11) Old Church Slavonic (Matthew 8:5)

    a. ***Prišedъ***   *i(isu)sъ*   *vъ kapernaoumъ.*   ***pripade***
       come.PAP.M.SG Jesus.NOM in Capernaum.M.ACC come_unto.AOR.3SG
       emou   *sъtъnikъ*   (Mt 8:5)
       him.DAT centurion.M.SG.NOM

    b. *vъšedъšou*   *že*   *emou*   *vъ kafernaoumъ.*
       come.PAP.M.SG.DAT PART him.DAT in Capernaum.M.ACC
       pripade   *emou*   *sъtъnikъ*   (Mt 8:5)
       come_unto.AOR.3SG (unto) him.DAT centurion.M.SG.NOM
       Greek
       eiselthóntos   *dè*   *autoû*   *eis Kapharnaoùm*
       enter.M.SG.GEN.PART.AOR2 PART himself.GEN into Capernaum
       prosēlthev   *autōi*   *hekatóntarchos*
       approach.3SG.AOR.2 to.him centurion
       '[And when] Jesus was entered into Capernaum, there came unto him a centurion'

The parallel absolute structures in (11a) and (11b) can be traced to the prehistoric locative absolute (morphologically close to the dative case) as reconstructed for Indo-European (cf. Haudry, 1977, p. 136–137; Hristova, 2004). One can therefore argue that the use of the predicative resultative participles in *-ųes-* in the texts of the Old Church Slavonic canon was largely an independent development, even if paralleled in the Greek. It comes as no surprise that a similar distribution of the nominative absolute and the dative absolute is attested in East Slavic written discourse (Živov, 2017, p. 417).

The growing frequency of the nominative absolute at the cost of the dative absolute in Old Church Slavonic was accompanied by the anterior (resultant) grammaticalization of the *-ųes-* participle. This participle went historically through decategorialization to become an indeclinable form used with an auxiliary, particularly in East Slavic spoken discourse (Sections 4–4.3) as exemplified in (12) (Miklosich, 1926, p. 834):

(12)   Ruthenian
    Umyslyvšy   *jesmo* (16th)
    think.PAP.INDECL be.AUX.PRS.1PL
    'We have considered [this]'/'consideravimus'

3.2   Later medieval evidence

Cases of the predicative resultative participles and the nominative absolute are commonly attested in the written discourse of medieval North Slavic (Klemensiewicz, 1974, p. 310; Večerka, 1961, p. 96, 107; Stanislav, 1958: 405–406; Gebauer, 1929, p. 613–615).

(13)   Middle Polish
    Tedy przykazawszy   *im*   *anioł,*   *aby*   *sie*
    then order.PAP.INDECL them angel.M.SG.NOM so.that REFL
    nie   *wrocili*
    NEG return.PST.3PL
    '[And] then the angel ordered them not to return'

Both nominative and predicative **-(v)ši-* participles are attested in the East Slavic recension of Church Slavonic. One should mention, for instance, the Anthology of the Dormition Cathedral in Moscow written in the late 12th century from the copy of the 11th century which contains several dozen predicative **-(v)ši-* participles (Meľničuk, 1958, p. 108–112). At this juncture, it is worth citing a typical sentence from the Lives of the Saints Boris and Gleb (Meľničuk, 1958, p. 115; cf. Živov, 2017, p. 330).

(14)   East Slavic Church Slavonic
    I   *se*   *uvěděvъ*   *Svjatoplъkъ, poslav*   *dva*
    and this see.PAP.M.SG Svjatoplъkъ send.PAP.M.SG two
    varjaga,   *i*   *probodosta*   *i*   *mečьmь*
    Varangians.M.DU.ACC and pierce.AOR3PL him sword.M.SG.INST
    vъ s(ь)rdce
    in heart.N.SG.ACC
    'And Svjatoplъkъ saw this, [and] sent two Varangians, and they pierced him with a sword in the heart'

In the light of ample evidence from early East Slavic Church Slavonic texts, Meľničuk (1958, p. 111–112) concluded that the predicative use of the **-ųes-* participle in the East Slavic written discourse of Church Slavonic reflected the actual state of this type of predication in 11th century Old Rusian (Old East Slavic).

The usage of the predicative *-(v)ši-* resultatives became ever more pronounced in Church Slavonic texts written in the early middle period. Meľničuk (1958, p. 130–133), in particular, found 42 reliable examples of this kind in the 1462 redaction of the Kyivan Cave *Patericon*, a collection of tales about the monks of the Kyivan Cave Monastery, whose original version arose in the early 13th century.

(15) East Slavic (Ruthenian) Church Slavonic
I   otveščavъ   blažennyj: Ašče ko knjazju   idu,
and answer.PAP.M.SG blessed.M.SG.NOM if   to prince.DAT go.PRS.1SG
to   i   ko vsěmъ idu
then and to all.DAT go.PRS.1SG
'And the blessed answered: if I go to the Prince, then I go to all [of them]'

Remarkably, out of the 42 examples coming from the *Patericon* one finds 4 cases of non-agreement of the predicative resultative participles with the matrix subject. The latter fact tentatively testifies to the early stage of the decategorialization of the *-(v)ši-* resultatives in written discourse under pressure from the narrative structures typical of spoken discourse; consider the following Example (16) which contains diagnostic vernacular forms like *nikoli* 'never' and *boľnyj* 'sick'.

(16) East Slavic (Ruthenian) Church Slavonic
Boľnyj   že   skoro   vsъtavъ,   jako nikoli
sick.M.SG.NOM PART immediately get_up.PAP.M.SG as   never
že   bolěvъ
part sick.PAP.M.SG
'And immediately the sick man was made whole as if he never was ill'

4. Evidence from spoken discourse

Cases of both nominative and predicative participles *-(v)ši-* are well attested in East Slavic spoken discourse. In Section 4.1, I discuss the resultative participles in the annalistic texts, while Section 4.2 deals with other genres. The dialect data are introduced in Section 4.3.

4.1 Annalistic texts

Of paramount importance, in this respect, is the language of the Rusian chronicles, particularly, the Tale of Past Days or, the Primary Chronicle, originally compiled in the early 12th century in Kyiv. The paradoxical fact that the content of the Primary Chronicle is oral but its form, written, raises questions about the mechanism

whereby these tales were written down and about the language used in the oral tales and in their written reflexes. Timberlake (2008, p. 211, 212, 217) rightly maintains that the narrative of the chronicle, from legends to reportage, was oral in origin; this explains why resultative participles are used more in narrative than in homiletic sections of the text (cf. Živov, 2017, p. 329–330).

Found in the Primary Chronicle, the number of predicative participles and nominative absolute constructions becomes somewhat larger by the end of the Kyivan Chronicle (12th century). The use of *-(v)ši-* participles is commonly attested in later East Slavic annalistic texts, including the Middle Russian chronicle of 1619–1691, in which Živov (2017, p. 420) noted a gradual loss of the morphosyntactic distinction between FPs and NFPs. This very tendency was first discussed by Meľničuk (1958) for Old Rusian and Middle Ukrainian. According to this scholar, the use of *-(v)ši-* participles in the Hypatian codex (1425) of the Primary Chronicle reflects the oldest layer of the resultative structures in Old Rusian. It is not surprising that the participle in Example (17) from the Hypatian codex is substituted by the plural aorist form *poxoroniša/poxraniša* 'buried' in all of the other eight redactions of the Primary Chronicle (Ostrowski, 2003, p. 1537).

(17)   Old Rusian   (Hypatian codex)
   i   braťja   poxoronivše   telo   ego   (1074)
   and brothers.NOM bury.PAP.M.PL body.N.SG.ACC his
   'And the brothers [have] buried his body'

Since such constructions are attested in various codices and redactions of the Primary Chronicle copied in different parts of Old Rus' (cf. Bjørnflaten, 2012), one can concur with Meľničuk (1958, p. 112) that the predicative usage of *-(v)ši-* resultatives is traceable in spoken discourse across vast territories, from Galicia to Kyiv to Novgorod.

As mentioned, syntactic parallels are easy to find in other northern Slavic languages of the middle period, particularly in spoken discourse. The predicative resultative participles are, for instance, found in the secular genre of Middle Czech *Alexandreida*, particularly its 1429 Svatovítský witness, which is based on oral narrative structures (Gebauer, 1929, p. 616).

(18)   Middle Czech
   král   uzřěv   to,   otstúpi   ho   vš
   king.M.SG.NOM see.PAP.M.SG that leave.PST.3PL him.ACC all
   nádějě   (1429)
   hope.F.PL.NOM
   'The king saw that, [and] all hopes left him'

4.2   Charters and beyond

The number of aorists and imperfects in the Old Novgorod birchbark charters is scanty, which can be explained by their oral narrative. As has been mentioned in Section 1, spoken discourse, characteristic of prototypically small speech communities, facilitates the anterior use of NFPs, including participles in *-ṷes-*. Timberlake (2008, p. 220–221) states that the Old Novgorodian charters have a two-part structure: at the outset, the speaker announces that a problem exists in the present; subsequently, the speaker proposes some possible action – redress, payment, threat, avoidance of responsibility. Events that occurred in the past are seen from the perspective of the speaker in his present moment and are likely to be understood as "perfect" – that is, as results of prior actions relevant in the speaker's current world (Zaliznjak, 1995, p. 616, 682; Timberlake, 2008, p. 219).

(19)   Old Novgorodian   (Charter 235)
     žǫdke poslavъ   ębetnika   dova i   pograbila   mę
     Žǫdke send.PAP.M.SG bailiff.M.DU.ACC two   and rob.PST.DU me.ACC
     '[And so] Žadko sent two bailiffs from the court and they [have] robbed me'

The same kind of oral narrative is observed in the written discourse of the Ruthenian administrative language as exemplified by the following citation (20) from a charter written in 1445 (Rozov, 1928, p. 150).

(20)   Ruthenian
     A   Pan   Ivanъ Mušata rekъ   peredъ name   (1445)
     and master Ivanъ Mušata say.PAP.M.SG before us.INSTR
     'And Master Ivan Mušata said before us'

In the 14th–15th century the use of auxiliaries with the predicative participle in *-(v)ši-* was minimal (Kolomijec', 1958). As early as the 16th century one comes across, however, more auxiliaries in the Ruthenian charters of that period, see Example (21) excerpted from Miklosich (1926, p. 834):

(21)   Ruthenian
     stojavšy   jesy   peredъ namy   i   opysalъ
     stand..PAP.INDECL be.AUX.2SG.PRS before us.INSTR and describe.PAP.M.SG
     jesy   jixъ
     be.AUX.2SG.PRS them
     'You have stood before us, and described them'

Some parallels can be found in other secular genres defined by spoken discourse in the medieval varieties of Czech, Polish, and Slovak (Miklosich, 1926, p. 834; Komárek, 2012, p. 206; Klemensiewicz et al., 1964, p. 384–385; cf. Taszycki, 1924,

p. 58). The diary written by Jan Władysław Poczobut (1640–1703) is particularly interesting from this point of view (Smolińska, 1983, p. 124); consider Example (22) below.

(22) Middle Polish
Dnia   8 maia   przybywszy   do Kupiszek
day.M.SG.GEN 8 may.M.SG.GEN arrive.PAP.INDECL at Kupiszki
'I arrived at Kupiszki [Lithuanian *Kupiškis*] on the 8th of May'

4.3   Dialects

In seeking the origin of the *-(v)ši-* resultative and its anterior grammaticalization in North Slavic, particularly in medieval Czech and Slovak (Gebauer, 1929, p. 613–618; Stanislav, 1958, p. 405–406), one finds dialect evidence especially useful.

Of particular interest is the Ukrainian dialect material found at the periphery of the postulated Circum-Baltic diffusion zone. Miklosich (1926, p. 834) was the first to cite Example (23) recorded in the 19th century in Transcarpathian Ukraine (cf. Golovackij, 1878, p. 104; Meľničuk, 1958, p. 133–134):

(23) Transcarpathian Ukrainian
Ja na ližku   ležala,   try lysta   pysala…
I on bed.N.SG.LOC lie.PST.F.SG three letters.ACC write.PST.F.SG
Mylyj   vzjavšy,   ta j   perečytavšy
sweetheart.M.SG.NOM take.PAP.INDECL and PART read.PAP.INDECL
'I lay on a bed, wrote three letters […] My sweetheart took [them] and read them'

Used without auxiliaries, the predicative participles in the second clause can nevertheless be juxtaposed with the preterites in the first clause, thus signaling an advancement in the anterior grammaticalization of the *-(v)ši-* resultative.

Similar predicative participles were sporadically attested in other parts of Ukraine in the 19th century (Meľničuk, 1958, p. 134). However, the bulk of their examples come from Southwest Ukrainian in the 18th–19th century. Examples from this part of Ukraine reflect the use of both predicative participles as in (24a) and nominative participles as in (24b) which were typical of older texts contingent on oral narratives introduced into written texts (cf. Franko, 1894, p. 62, 136).

(24) Southwest Ukrainian
a. *Ax uzrivšy   Honta* […] *xlopci*
   oh see.PAP.INDECL Honta   fellows.NOM/ACC
   'Oh, Honta has seen/saw his fellows'

b. *Mudryj mladenec pročytavšy lyst,*   *A*   *v*
   wise   baby   read.PAP.INDECL letter.M.SG.ACC and in
   tom lysti   *stoit* […]
   that letter.M.SG.LOC stay.PRS.3SG
   'The wise baby read the letter, and this letter states […]'

The -*(v)ši*- resultative in both standard and non-standard varieties of Southwest Ukrainian does not take an auxiliary and, if used as an anterior, resembles similar structures found in northeastern borderland Polish in the 18th century. Consider the following example dating back to 1726 from Kurzowa (2006, p. 296):

(25) Northeastern Borderland Polish
   mało co piwszy,   *tylko łyszkę*   *iedną*
   a little what drink.PAP.INDECL only spoon.F.SG.ACC one.F.SG.ACC
   barszczu   *wziowszy*   (1726)
   borsch.M.SG.GEN take.PAP.INDECL
   '[He] has drunk a little, he has eaten only one spoon of borsch'

The -*(v)ši*- resultatives derived from both transitives and intransitives were attested in Polish as early as the 14th century, although their frequency has ostensibly dropped by the 19th century (Kurzowa, 2006, p. 295).[5] As Meľničuk (1958, p. 112) argued, the usage of resultatives in East Slavic spoken discourse is traced back to the 11th century; they are attested in modern Ukrainian dialects.

(26) Transcarpathian Ukrainian[6]
   Kozak rano vstavšy,   *pys'mo*   *pročytavšy,*   *tjažen'ko*
   cossack early get.up.PAP.INDECL letter.N.SG.ACC read.PAP.INDECL heavily
   zitxnuvšy,   *koxannja*   *spiznavšy*
   sigh.PAP.INDECL love.N.SG.ACC get.to.know.PAP.INDECL
   'The Cossack got up early, read a letter, heavily sighed, [and] got to know love'

Tentatively, both the southwestern Ukrainian and northeastern borderland Polish resultatives could have reached, during the late medieval period, a similar level

---

5. Bednarczuk (2010, p. 9) viewed the use of the resultative construction in northeastern borderland Polish as one of the syntactic features of a linguistic league which emerged in the Polish-Lithuanian Commonwealth. It is not, however, clear how this hypothesis can be reconciled with the theory of the diffusion of the *-ṷes-* resultatives from Baltic to Slavic (see Section 1).

6. Example (26) was kindly provided by Vasyľ Nimčuk († 27 November, 2017) from the Institute of the Ukrainian Language, National Academy of Sciences of Ukraine, who had recorded this song in the Village of Miškarovycja, Volovec'kyj district, Transcarpathian region, in the 1950s. Nimčuk assumed that, due to the use of the appellative *Cossack* and some other features, this song could have been brought to this area from Dnieper Ukraine.

of the grammaticalization as reflected in their initial semantic shift of anterior to preterite (Kurzowa, 2006, p. 296; Karaś, 2002, p. 252–253). This assumption can be corroborated extralinguistically. The two language varieties were employed by the corresponding speech communities within the same state borders; consequently, similar societal factors tended to influence Polish and Ukrainian, as well as other speech communities in contact. Under pressure of spoken discourse, conducive to the analyticization of the language system, the shift of anterior to preterite seems to have run to its completion in modern northeastern borderland Polish, Belarusian, and Lithuanian.

5. Discussion

The foregoing diachronic and areal-typological sampling of the resultatives in East Slavic and other northern Slavic varieties casts serious doubt on the theory of the diffusion of the resultatives from Baltic into Slavic. Instead, one legitimately asks about what could bring about their full grammaticalization in Baltic, and some Belarusian and Polish dialects, including the development of subjective, objective, and bi-diathetic patterns and the introduction of HAVE in the resultative structure with a *-ṷes-* participle, see Examples (5) and (6).[7]

One of the alternative theories was offered by Jung (2014) who rightly stated that it would be too strong a claim to suggest that the *-(v)ši-* resultative in Northwest Russian appeared as a direct result of the Balto-Finnic substratum effect, i.e., as a syntactic borrowing; according to her, this construction developed too late to be regarded as a direct result of the language contact in the northwestern region neighboring Lithuania, Latvia, Estonia, and Finland. Instead, this author claimed that the geographical distribution of the *-(v)ši-* resultative conicides with the area where *v* and *l* were neutralized after the jer shift in Old Russian (Novgorodian), the establishment of what she called the "*-(v)ši-* perfect" as opposed to the preterite,

---

7. The advanced level of the anterior grammaticalization of resultatives in Northwest Russian allowed Razmusen (1891, p. 403–404) to posit for the non-standard variety of Russian three "absolute" tenses, i.e., past – *ja napisal/pisal* 'I wrote', present – *ja pišu* 'I write', future – *ja napišu/ja budu pisat'* 'I will write', and three "relative" tenses, i.e., "anterior past" – *ja byl prišedši* (come-PAP. INDECL) 'I had come', "anterior present" – *ja prišedši* (come- PAP.INDECL) 'I have come', "anterior future" – *ja budu prišedši* (come- PAP.INDECL) 'I will come'. Razmusen compared the Russian "anterior" tenses with the analytic perfect tenses in Western European languages like German *ich hatte geschrieben* 'I had written' (anterior past), *ich habe geschrieben* 'I have written' (anterior present)), *ich werde geschrieben haben* 'I will have written' (anterior future) and other forms more variegated than in Russian. Kuz'mina (1993, p. 157–158) argued along the same lines that the framework of anteriors in Russian is too small for accommodating the participial constructions, including the resultative structures in Northwest Russian.

must have been based on this phonological condition (Jung, 2014, p. 211–212). Strong in its critique of the Balto-Finnic substratum, Jung's explanation, nevertheless, is open to doubts.

First, the phonological basis of the grammaticalization in question is untenable. To begin with, one should bear in mind similar constructions in northeastern borderland Polish characterized by a phonological voicing system with neutralization before all obstruents, e.g., *pšyšetšy* from *pšyše[d]šy* 'having come'; this system differs from phonemic protensity developed in the phonological system of Northwest Russian, (Southeast and North) Ukrainian, and Belarusian where *v* and *l* are neutralized in *v* [ŭ] (Danylenko, 2006, p. 189–191), see Examples (3a), (3b), (4a) and (4b).

Moreover, in the same variety of borderland Polish, one finds cases without reflexes of the voicing system, e.g., *dowiodłszy* '(having) led' (1648) and *ukradłszy* '(having) stolen' (1762) (Kurzowa, 2006, p. 295; Klemensiewicz et al., 1964, p. 385). In some western Russian dialects the resultative construction utilizes forms in *-lši* and not *-vši*, thus resembling the aforementioned borderland Polish forms.

Second, Jung (2014, p. 212) argues that the Old Russian predicative *-(v)ši-* participle developed into a fully independent predicate denoting the perfect in some Russian dialects, while the preterite-denoting function was assigned to the suffix *l* [ŭ]. According to Kuz'mina (1993, p. 155; Markova, 2014, p. 139), however, the predicative participles are practically not opposed to the *l*-preterites in modern Russian dialects, including Northwest Russian. To draw a parallel, in northeastern borderland Polish, participles in *-(f)šy-* also tend to be used as preterites (Karaś, 2002: 251–253).

In Section 1, I sketched out another alternative explanation premised on the distinction of two parallel types of predication brought about by two different – spoken vs. written – discourse styles operative in the history of North Slavic. Quite in the same vein, Živov (2017) posited the idea of functional parallelism of spoken and written discourse in the complexification of East Slavic syntax, arguing that "no evolutionary cline" from parataxis to hypotaxis can be reconstructed for such a development. Elsewhere (Danylenko, 2018a) I also challenge the framework of unidirectionality and linearity used for the explanation of the rise of syntactic complexity in the relativization strategies of East Slavic. I posit, instead, two separate (parallel) pathways of relative clause chaining in East Slavic, one leading from parataxis to "paratactic subordination", and a second leading from parataxis to hypotaxis and ultimately to "hypotactic subordination".

The aforementioned pathways of relative clause chaining can be matched with the two separate (parallel) pathways of the anterior grammaticalization for FP and NFP, in particular the *-(v)ši-/-vęs-* (< *-ųes-*) participles in Baltic and (North) Slavic. Unlike the FP pathway typical of written discourse, the NFP anterior grammaticalization

is posited for spoken discourse. As has been shown, the two types of discourse (as well as the two types of predication) can intermingle, a process which is reflected in the use of narrative elements alongside homiletic components in the Church Slavonic canon and other northern Slavic bookish texts extant from the medieval period. However, the application of spoken discourse has demonstrated consistency and continuity in choosing the respective NFP forms in *-(v)ši-* which became fully grammatized in Lithuanian and most of the northern Slavic varieties, primarily Northwest Russian, some Belarusian dialects, and northeastern borderland Polish.

What is left to determine is what preconditions the anterior grammaticalization of the *-(v)ši-* resultative would have in the aforementioned varieties since, as is evidenced from Ukrainian and some Polish dialects, a contact hypothesis would be only a partial explanation. In fact, to cast around for exclusively contact explanations of this grammaticalization which presuppose either borrowing or replication is heuristically unproductive. As was hypothesized elsewhere (Danylenko, 2018a, also 2018b, 2018c) in terms of sociolinguistic typology, the formation of paratactic and hypotactic subordination is dependent on a historically prevalent type of communication (discourse) practiced within a particular speech community. The prevalent type of communication is determined ultimately by a particular constellation of societal (extralinguistic) factors like type of social stability, size of a speech community, type of social networks, and degrees of communally shared information, and finally the amount of adult language contact (Trudgill, 2011, p. 13–14, 62, 146; Danylenko, 2018b, 2018c). One needs to repeat here that language contact is only one of the confluent societal factors that are likely to trigger changes in linguistic patterning. Among such changes, one can mention the strengthening of analyticity leading to the reduction of grammatical patterns such as conjugations, declensions, and inflected forms, the rise of articles, an increase in the use of adpositional and periphrastic constructions, the development of fixed word order, and the like (Dahl, 2004, p. 281; Haspelmath & Michaelis, 2017).

When resorting to the sociolinguistic typology in the spirit of Trudgill, one can posit a series of societal variables explaining similarities in the anterior grammaticalization of the *-(v)ši-/-vę̇s-* (< **-u̯es-*) resultatives in North Slavic and Baltic. I argue, in particular, that the prototypical small speech communities that had found themselves in the Circum-Baltic area and beyond were characterized by convergent societal variables. This may explain the strengthening of NFP accompanied by the use of the nominative object in some Russian dialects (Kuz'mina, 1993, p. 145), see also Examples (3b) and (6). The introduction of the HAVE auxiliary (typical of the analytic western European languages) in Lithuanian, northeastern borderland Polish and some transitional Belarusian dialects is a result of what I called analytic (secondary) simplification (analyticization) in the inflecting Slavic and Baltic languages (Danylenko, 2002, p. 119–121, 2018c: 119–121)

6.   Conclusions

I venture to surmise that the driving force behind the anterior grammaticalization of the *-(v)ši-* resultatives in North Slavic can hardly be found in the alleged diffusion of this construction from Baltic to Slavic (cf. Večerka, 1958, p. 233). The evidence challenging this scenario is enough to argue that the aforementioned grammaticalization was internally motivated and largely driven by spoken discourse characterized by a tendency to non-finite predication. Convergence of the resultative structures in North Slavic and Baltic, primarily in their non-standard varieties, was brought about by the historical convergence of societal variables determining communication within speech communities which had not been necessarily in close contact with one another.

In sum, the theory of areal diffusion of the *-(v)ši-* resultative in the Circum-Baltic area can hardly be taken at face value. In fact, it should be further tested against sociolinguistic and typological argumentation.

Abbreviations

ACC	accusative	INF	infinitive
ACT	active	LOC	locative
AOR	aorist	M	masculine
AUX	auxiliary	NFP	non-finite verbal predicates
DAT	dative	NOM	nominative
DU	dual	PART	participial
F	feminine	PL	plural
FP	finite verbal predicates	PRS	present
GEN	genitive	PST	past
IMPERF	imperfect	REFL	reflexive
SG	singular	1	first person
INDECL	indeclinable	2	second person
INSTR	instrumental	3	third person

References

Adomavičute, I. & Čekmonas, V. (1991a). Obrazovanie i morfologičeskij status form na *-(f)šy* v "poľščizne vilenskoj" [The development and the morphological status of forms in *-(f)šy* in Vilnius Polish]. *Studia nad polszczyzną kresową*, 6, 85–94.

Adomavičute, I. & Čekmonas, V. (1991b). Paradigma kvaziperfektnyx form na *-(f)šy* v "poľščizne vilenskoj" [The paradigm of pseudo-perfect forms in *-(f)šy* in Vilnius Polish]. *Studia nad polszczyzną kresową*, 6, 95–105.

Ambrazas, V. (1990). *Sravnitel'nyj sintaksis pričastij baltijskix jazykov* [A comparative syntax of the participles in Baltic]. Vilnius: Mokslas.

Ambrazas, V. (2006). *Lietuvių kalbos istorinė sintaksė* [A historical syntax of the Lithuanian language]. Vilnius: Lietuvių kalbos institutas.

Andersen, H. (2009). On the origin of the Slavic aspects. In V. Bubenik, J. Hewson, & S. Rose (Eds.), *Grammatical change in Indo-European languages* (pp. 123–140). Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/cilt.305.14and

Bednarczuk, L. (2010). Zróżnicowanie terytorialne północno-wschodniej polszczyzny kresowej [The territorial differentiation of northeastern borderland Polish]. *Rozprawy Komisji Językowej*, 55, 5–39.

Bjørnflaten, J. I. (2012). The Slavonic gerund and its development in Russian. In A. Danylenko & S. Vakulenko (Eds.), *Studien zu Sprache, Literatur and Kultur bei den Slaven. Gedenkschrift für George Y. Shevelov* (pp. 62–78). Munich: Otto Sagner.

Čekmonas, V. (2001). Russian varities in the southeastern Baltic area. In Ö. Dahl and M. Koptjevskaja-Tamm (Eds.). *The Circum-Baltic Languages: Typology and Contact*, vol. 1: *Past and Present* (pp. 101–136). Amsterdam/Philadelphia: John Benjamins.

Dahl, Ö. (2004). *The Growth and Maintenance of Linguistic Complexity*. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/slcs.71

Danylenko, A. (2002). The East Slavic 'HAVE': Revising a developmental scenario. In K. Jones-Bley, M. E. Huld, A. Della Volpe, & M. Robbins Dexter (Eds.), *Proceedings of the Thirteenth Annual UCLA Indo-European Conference. Los Angeles*, 2001 (pp. 1045–127). Washington: Institute for the Study of Man.

Danylenko, A. (2006). *Slavica et Islamica. Ukrainian in Context*. Munich: Otto Sagner.

Danylenko, A. (2017). A Missing Chain? On the sociolinguistics of the Grand Duchy of Lithuania. *Acta Baltico-Slavica*, 41, 31–57. https://doi.org/10.11649/abs.2017.002

Danylenko, A. (2018a). A tale of two pathways: On the development of relative clause chaining in East Slavonic. In J. Grković-Major, B. Hansen, & Sonnenhauser, B. (Eds.). *Diachronic Slavonic syntax* (p. 361–386). Berlin/Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110531435-013

Danylenko, A. (2018b). The correlation of linguistic patterning and societal structures in systemic typology. *Studia Linguistica Universitatis Iagellonicae Cracoviensis*, 135, 81–96.

Danylenko, A. (2018c). Complexity development and simplification in systemic typology: Types of changes in the internal determinant. *Studia Linguistica Universitatis Iagellonicae Cracoviensis*, 135, 113–129.

Danylenko, A. (2018d). The Arabs, the Slavs, and United Europe or, the vagaries of the development of Indo-European perfect. *Lingua*, 203, 16–26. https://doi.org/10.1016/j.lingua.2017.10.001

Delbrück, B. (1897). *Vergleichende Syntax der indogermanischen Sprachen*, part 2. Strassburg: Karl J. Trübner.

Dmitrieva, L. (1962). K voprosu o značenii i genezise otpričastnyx form (na materiale sovremennyx pskovskix govorov i pskovskix letopisej) [On the question of the meaning and origin of participial forms (on the material of modern Pskov dialects and the Pskov chronicles)]. *Pskovskie govory*, 1, 143–161.

Drinka, B. (2017). *Language Contact in Europe. The Periphrastic Perfect through History*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781139027694

Du Bois, J. (1987). The discourse basis of ergativity. *Language*, 63, 805–855. https://doi.org/10.2307/415719

Endzelin, J. (1923). *Lettische Grammatik*. Heidelberg: Carl Winder.

Erker, A. (2014). Sposoby vyraženija prošedšego vremeni v belorusskom smešannom govore na balto-slavjanskom pograničʹe [Means of the expression of past time in a Belarusian mixed dialect at the Baltic-Slavic borderland]. In I. A. Seržant & B. Wiemer (Eds.), *Contemporary approaches to dialectology. The area of North, North-West Russian and Belarusian dialects* (Slavica Bergensia 12) (pp. 130–149). Bergen: University of Bergen.

Franko, O. (Ed.). (1894). *Žytje i slovo. Vistnyk literatury, istoriji i folʹkloru* [Life and word. The messenger of literature, history, and folklore], vol. 2. Lviv: Instytut Stavropihijsʹkyj.

Gebauer, J. (1929). *Historická mluvnice jazyka českého* [A historical grammar of the Czech language], vol. 4: *Skladba* [Syntax]. Prague: F. Tempsky.

Geniušienė, E. & Nedjalkov, V. (1988). Resultative, passive, and perfect in Lithuanian. In V. P. Nedjalkov (Ed.), *Typology of resultative constructions* (pp. 369–386). Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/tsl.12.27gen

Golovackij, J. F. (1878). *Narodnye pesni Galickoj i Ugorskoj Rusi* [Folk songs of Galician and Ugric Rusʼ], part 1. Moscow: Imperatorskoe Obščestvo Istorii i Drevnostej Rossijskix pri Moskovskom Universitete.

Haspelmath, M. & Michaelis, S. M. (2017). Analytic and synthetic: typological change in varieties of European languages. In I. Buchstaller & B. Siebenhaar (Eds.). *Language Variation – European Perspectives VI: Selected Papers from the 8th International Conference on Language Variation in Europe (ICLaVE 8), Leipzig 2015* (Studies in Language Variation 19) (pp. 1–17). Amsterdam: John Benjamins. https://doi.org/10.1075/silv.19.01has

Haudry, J. (1977). *Lʼemploi des cas en vedique. Inroduction á lʼétude des cas en indo-européen*. Lyon: LʼHermès.

Hristova, D. S. (2004). Absolute constructions in Slavic: case diversityand originality. *Journal of Indo-European Studies*, 32. 297–318.

Jagić, V. (Ed.). (1883). *Quattuor evangeliorum versiones palaeoslovenicae Codex Marianus glagoliticus*. Berlin: Weidmann.

Jung, H. (2014). Conditioning factors in the development of the *-vši* perfect in West Russian. In I. A. Seržant & B. Wiemer (Eds.), *Contemporary Approaches to Dialectology. The Area of North, North-West Russian and Belarusian Dialects* (Slavica Bergensia 12) (pp. 195–215). Bergen: University of Bergen.

Karaś, H. (2002). *Gwary Polskie na Kowieńszczyźnie* [The Polish dialects in the Kowno region]. Warsaw/Puńsk: Aušra.

Klemensiewicz, Z. (1974). *Historia języka polskiego* [A history of the Polish language]. Warsaw: Państwowe Wydawnictwo Naukowe.

Klemensiewicz, Z., Lehr-Spławiński, T. & Urbańczyk, S. (1964). *Gramatyka historyczna języka polskiego* [A historical grammar of the Polish language]. Warsaw: Państwowe Wydawnictwo Naukowe.

Knjazev, J. P. (1983). Rezulʹtativ, passiv i perfekt v russkom jazyke [Resultative, passive, and perfect in the Russian language]. In V. P. Nedjalkov (Ed.), *Tipologija rezulʹtativnyx konstrukcij: rezulʹtativ, stativ, passiv, perfekt* (pp. 149–160). Leningrad: Nauka.

Kolomijecʹ, L. I. (1958). Do pytannja pro formuvannja katehoriji diepryslivnyka v ukrajinsʹkij movi [On the question of the formation of the category of adverb in the Ukrainian language]. *Trudy filolohičnoho fakulʹtetu Xarkivsʹkoho deržavnoho universytetu im. O. M. Horʼkoho*, 6, 201–213.

Komárek, M. (2012). *Dějiny českého jazyka* [History of the Czech language]. Brno: Host.

Kurzowa, Z. (2006). *Język polski Wileńszczyzny i kresów północno-wschodnich XVI-XX w.* [The language of the Vilnius region and the northeastern borders of the 16th – 20th century]. Cracow: Universitas.

Kuz′mina, I. B. (1993). *Sintaksis russkix govorov v lingvogeografičeskom aspekte* [Syntax of the Russian dialects from the linguistic-geographical point of view]. Moscow: Nauka.

Mackevič, J. F. & Hrynaveckene, E. I. (1993). Dzeeprysloŭi na *-(ŭ)šy* ŭ belaruskix narodnyx havorkax [Adverbs in *-(ŭ)šy* in the Belarusian dialects]. *Lietuvių kalbotyros klausimai*, 30, 105–109.

Marguliés, A. (1927). *Der altkirchenslavische Codex Suprasliensis.* Heidelberg: Carl Winter.

Markova, N. V. (2014). Dialektnye perfektnye konstrukcii v odnom zaonežskom govore i idiostile [Dialectal participial constructions in one Onega dialect and in idiostyle]. *Severnorusskie govory*, 13, 137–148.

Meľničuk, A. S. (1958). Razvitie predikativnogo upotreblenija pričastij na *-(v)ъ, -(v)ъš-* v vostočnoslavjanskix jazykax [The development of the predicative use of participles in *-(v)ъ, -(v)ъš* in East Slavic]. In *Slov″jans′ke movoznavstvo*, vol. 1 (pp. 91–159). Kyiv: Akademija Nauk URSSR.

Miklosich, F. (1926). *Vergleichende Grammatik der slavischen Sprachen*, vol. 4: *Syntax.* Heidelberg: Carl Winder.

Nau, N. (2011). *A Short Grammar of Latgalian* (Languages of the World. Materials 482). Munich: LINCOM Europa.

Nedjalkov, V. & Jaxontov, S. (1983). Tipologija rezuľtativnyx konstrukcij [Typology of resultative constructions]. In V. Nedjalkov (Ed.), *Tipologija rezuľtativnyx konstrukcij: rezuľtativ, stativ, passiv, perfekt* (pp. 5–41). Leningrad: Nauka.

Obnorskij, S. (1953). *Očerki po morfologii russkogo jazyka* [Essays on the morphology of the Russian language]. Moscow: Izdateľstvo AN SSSR.

Orr, R. A. (1989). Russo-Goidelic syntactic parallel: *u nego svoja izba postavlena / tá sé déanta agam. General Linguistics*, 29 (1), 1–21.

Ostrowski, D. (Ed.). (2003). *The Pověst′ vremennykh lět. An interlinear Collation and Paradosis* (Harvard Library of Early Ukrainian Literature. Texts 10/3). Cambridge, MA: Harvard Ukrainian Research Institute.

Potebnja, A. (1888). *Iz zapisok po russkoj grammatike* [From the notes on Russian grammar], vol. 1–2. Kharkiv: D. N. Poluextov.

Proxorova, S. (1991). *Sintaksis perexodnoj russko-belorusskoj zony: areaľno-tipologičeskoe issledovanie* [Syntax of the transitional Russian-Belarusian zone: an areal-typological study]. Minsk: Universitetskoe.

Razmusen, L. P. (1891). O glagoľnyx vremenax i ob otnošenii ix k vidam v russkom, nemeckom i francuzskom jazykax [About verb tenses and their relationship with the aspects in Russian, German, and French]. *Žurnal Ministerstva narodnogo prosveščenija*, 275, 376–417; 276, 1–56; 277, 1–39; 278, 185–190.

Rozov, V. (1928). *Ukrajins′ki hramoty* [Ukrainian charters], vol. 1 (Pam″jatky i studiji z istoriji ukrajins′koji movy 1). Kyiv: Ukrajins′ka Akademija Nauk.

Růžička, R. (1963). *Das syntaktische System der altslavischen Partizipien und sein Verhältnis zum griechischen* (Deutsche Akademie der Wissenschaften zu Berlin. Veröffentlichungen des Instituts für Slawistik 27). Berlin: Akademie Verlag.

Samoxvalova, E. (1982). Osobennosti meždialektnogo vzaimodejstvija russkix i ukrainskix go-vorov na sintaksičeskom urovne [Particularities of the interdialectal interrelations between the Russian and Ukrainian dialects at the syntactic level]. *Russkiew govory na Ukraine* (pp. 129–180). Kyiv: Naukova dumka.

Smolińska, B. (1983). *Polszczyzna północnokresowa z przełomu XVII i XVIII w.* [Northeastern borderland Polish at the turn from the 17th to the 18th century] (Prace Instytutu języka polskiego 51). Wrocław, etc.: Zakład narodowy imienia Ossolińskich, Wydawnictwo Pols-kiej Akademii Nauk.

Sobolev, A. N. (1998). O predikativnom upotreblenii pričastij v russkix dialektax [About the predicative use of participles in Russian dialects]. *Voprosy jazykoznanija*, 5, 74–89.

Stanislav, J. (1958). *Dejiny slovenského jazyka* [History of the Slovak language], vol. 2. Bratislava: Slovenská Akadémia Vied.

Tannen, D. (1983). Oral and literate strategies in spoken and written discourse. In R. W. Bailey & R. M. Fosheim (Eds.), *Literacy for life: The demand for reading and writing* (pp. 79–96). New York: Modern Language Association.

Taszycki, W. (1924). *Imiesłowy czynne, teraźniejszy i przeszły 1. w języku polskim* (Active parti-ciples, the present and past participle 1 in the Polish language) (Rozprawy Wydziału Filo-logicznego. Polska Akademia Umiejętności, Kraków Wydział Filologiczny 61/5). Cracow: Polska Akademia Umiejętności.

Timberlake, A. (2008). The grammar of oral narrative in the *Povest' vremmenykh let*. In C. Bethin (Ed.), *American contributions to the 11th International Congress of Slavists. Ohrid 2008, vol. 1: Linguistics* (pp. 211–226). Bloomington, IN: Slavica.

Trubinskij, V. I. (1983). Rezul'tativ, passiv i perfekt v nekotoryx russkix govorax [Resultative, pas-sive, and perfect in some Russian dialects]. In V. P. Nedjalkov (Ed.), *Tipologija rezul'tativnyx konstrukcij: rezul'tativ, stativ, passiv, perfekt* (pp. 216–226). Leningrad: Nauka.

Trudgill, P. (2011). *Sociolinguistic Typology*. Oxford: Oxford University Press.

Večerka, R. (1958). Poznamky ke genesi a vývoji ruských nářečních obratů s přísudkovým *s*-ovým participem [Comments on the origin and development of the Russian dialectal construction with the predicative participle in -*s*]. *Československá rusistika*, 3, 227–233.

Večerka, R. (1961). *Syntax aktivních participií v staroslověnštině* [Syntax of the active participles in Old Slavic]. Prague: Státní pedagogické nakladatelství.

Večerka, R. (1993). *Altkirchenslavische (altbulgarische) Syntax, vol. 2: Die innere Satzstruktur* (Monumenta linguae slavicae. Dialecti veteris. Fontes et dissertationes 34/37 (2)). Freiburg i. Br.: U. W. Weiher.

Wiemer, B. (2012). The Lithuanian *have*-resultative – a typological curiosum? *Lingua Posna-niensis*, 54, 69–81. https://doi.org/10.2478/v10122-012-0015-z

Wiemer, B. & Giger, M. (2005). *Resultativa in den nordslavischen und baltischen Sprachen*. Munich: LINCOM Europa.

Zaliznjak, A. (1995). *Drevnenovgorodskij dialect* [Old Novgorodian dialect]. Moscow: Jazyki rus-skoj kuľtury.

Zaliznjak, A. (2004). *Drevnenovgorodskij dialect* [Old Novgorodian dialect], 2nd ed. Moscow: Jazyki russkoj kuľtury.

Živov, V. (2017). *Istorija jazyka russkoj pis'mennosti* [The history of the language of Russian writing culture], Vols. 1–2. Moscow: Universitet Dmitrija Požarskogo.

Atomizing linguistic change

A radical view

Dieter Stein

Heinrich-Heine-Universität Düsseldorf

This paper discusses the pragmatic processes of semiosis involved in real speaker-hearer interaction in the process of an initial innovative form-function construal. This atomistic view of the locus of linguistic change suggests that more prominence needs to be given to motivations which are present at the point of origin. Discussing two unrelated instances from the history of English, the paper argues that the initial motivation can become non-transparent in later stages of the process. Hypotheses based on structural comparisons may therefore be quite different from hypotheses of motivations effective at the point of origin.

Keywords: atomistic, micro-history of change, inter-subjectivity, morphosyntactic change, onset context, coercion, real-time pragmatics, construction grammar, emergent, invisible hand, comprehensible input, semiosis, Do-support, extravagant progressive

1.   Levels of explanation

Building on a view of variation as structured and rule-governed, Weinreich, Labov and Herzog (1968) re-directed the focus of historical linguistics to how a linguistic innovation moves through social space towards categoriality and thus to a change in the linguistic system. Their work represented a significant alternative to views that discussed linguistic change only in terms of types of changes in "the grammar", and motivations, teleologies and mechanisms derived from these views of language. On the whole, the field of historical linguistics has produced, not a unitary view of linguistic change, but a large number of theories building on specific synchronic views of language. The logical priority of a view of language as the backdrop of a theory of linguistic change has been highlighted by Luraghi (2010): "Ultimately, one's views on the causes of change are inextricably connected with one's general assumptions on language and on the real object of linguistic research" and, consequently, "…

current theories of causation in language change are quite disparate, and, depending on the perspective from which they are seen, may also look rather unlikely" (Luraghi, 2010, p.358).

A common denominator of the bulk of work on morphosyntactic change, on which this paper will be focused, is that it either implicitly or explicitly presupposes a notion of synchronic structure as a default assumption, with the task of linguistic theory being to account for how we – or languages – get from one state to another state. The discussion of linguistic change has therefore been largely holistic compared to a view of language as individual acts of using language.

There is, then, a considerable range of levels on which different kinds of theories of change have been focused. At one end of the whole process of change is the actual production of individual utterances by the individual, in the middle ground is sociolinguistic diffusion, and at the other is "the language," "the system," "langue," or "the grammar". No matter how the latter group of concepts differ from each other, they have one thing in common: they are Platonic hypotheses about abstractions from the individual speech act, postulating how discrete units of the code, hypostasized as isolated, decontextualized, lonely units are replaced by others at a later system state. Essentially, theories of linguistic change as system comparisons are built on "figments of thought" (Enfield 2014, p.5). Very often, system-based theories, extrapolating backwards, arrive at conclusions that are very different from hypotheses that are closer to the individual speech act. System-based theories stress commonalities, abstractions, and generalizations (on the part of the analyst, not necessarily on the part of the speaker), not differences, between the individual acts and their completely different contextualizations in each individual act. This fact is only partially represented by the concept of "variation," which is still predicated on groups, and less on individuals. The very notion of "linguistic change" as an explanandum has as a logical presupposition the idea that language should be seen primarily as a repository of a static, more or less fixed set of rules and norms, specifying how to perform all kinds of functions for which language is used (Bailey, 1980, esp. pp.139–140 for an analysis of these views).

Naturally, then, any change in these norms – or call it "structure" or "grammar" – constitutes a "problem" that needs to be explained. So linguistic change as an explanandum really arises as a result of a view of language as something basically static and stable, and not what it really is, a permanently changing tool: as Bailey (1980, p.149) puts it, paraphrasing Coseriu (1958 / 1974), "changing belongs to the nature (*Seinsweise*) of language". Synchronic linguistics, with staticity and immutability of the notion of "structure", may be seen as a necessary heuristic and an operative assumption for linguistics, but this assumption obscures the very nature of communication by language and, more importantly in the present context of discussion, it obscures the nature and our view on the processes of "change"

that are intrinsic to language. The consequence of this perspective is having to declare language change "a problem" and approaching these intrinsic properties with the wrong perspective, a perspective which must necessarily abstract away from the ground level of linguistic communication, or, at best, must assume some unspecified process of "fixed-code literalist telementation" (Toolan, 2002, p.1964) to account for the embarrassing presence of the pragmatics of communication, of hearers constructing inferred speaker meanings, and of innovation and change as inherent in each and every act of speakers' use of language to communicate.

In recent decades, the view that a theory of language change should derive from a theory of grammar viewed as "an object apart from the speaker and separated from the uses which the speaker may make of it" (Hopper 1987, p. 3) with the corollary of a "fixed code", is increasingly, though not in all quarters, being substituted by a view of language as "emergent": "its structure is always deferred, always in a process but never arriving, and therefore emergent" (Hopper 1987, p.3). A view of language and linguistic change in permanent flux is a concept developed in the framework of construction grammar, an approach hospitable to accommodating persistent changes in the form and meaning of subsets of language, the so-called "constructions." This approach is exemplified in Hilpert's (2013) study of parenthetical concessives. He shows that the persistent changes and gradient nature of fixation requires, not concepts for fixed states, but concepts that reflect gradient processes and degrees of construction schema formation, also reflected in a differentiation of concepts such as "meso-" and "macro-constructions" (Hilpert 2013, p.202). For the theory of linguistic change, one of the essential forward steps that construction grammar has made is the perspective of breaking down one structural change into several smaller changes:

> "Constructional changes then can be hypothesized by the analyst to precede and enable or 'feed' constructionalization" typically involving "expansion of pragmatics, semanticization of that pragmatics, mismatch between form and meaning [i.e. false steps, D.S.], and some small distributional changes. We call these 'preconstructionalization constructional changes' […]. In turn, constructionalization may feed further constructional changes.   (Traugott and Trousdale 2013, p.27)

If explanations derived from system correspondences operate on the macro-level of the theory of linguistic change, construction grammar has taken a significant step towards the description of the micro-level and individual steps in the micro-history of change. Arguably, however, the most important step in identifying the sources of language change continues to be largely out of sight: it is that change arises in communication between real speakers in physical space and real time. While there has lately been some change in perspective, a recent study of syntactic change is correct in characterizing the new mainstream approach of syntactic change as follows:

> Studies in syntactic change typically treat change as happening to some abstract object 'language'. Yet it is individual language users who do the talking and innovate. Innovation, when conventionalized, leads to syntactic change […] It is therefore essential that historical syntacticians fully embrace this individual dimension.
>
> (Petré 2017, p. 227)

A satisfying explanation of innovation processes can only be developed if this pragmatic aspect of language use as a fully contextualized real-time event is acknowledged. All norms – and the "rule" for forming questions in English with "do" is such a norm of a speaker community – are reasonably hypothesized to have an abstract existence, be that as schemas or as remembered patterns, but they are manifested, changed, or entrenched only in the actual speech act in a fully specified context with speaker intentions and hearer-side performative acts.

It is this more pragmatic side of innovation that is the focus of this article. While in modern approaches to linguistic change, and especially in the framework of construction grammar, the individual speaker has moved into the picture as the agent of change, both in their cognitive (Bybee 2010) and in their psycholinguistic (Hundt, Mollin, & Pfenninger 2017) dimensions, what is true for construction grammar is also true for the wider picture: "speakers and hearers as embodied agents […] have not been sufficiently integrated into Construction Grammar" (Hilpert 2013, pp. 209–10). Construction grammar has been locally, but not yet globally, instrumental in shaping a more modern view of linguistic change mainly for two reasons. The scope of its inquiry has been restricted mainly to individual expressions, whose ancestry can more easily be traced in more minute detail. The other reason is that the context and steps of constructionalization that allow gradient developments and series of small-scale changes (neoanalyses) to become transparent have not been included. Its much more constrained scope has provided more differentiated views of local developmental stages that can be extended to the study of language change in general.

2.   A gap in explanation

Given the dominant system-based approach to linguistic change, an explanation of persistent change at the source level which includes speaker innovation with ensuing diffusion – Traugott and Trousdale's (2013) "onset" context or Petré's (2017, p. 227) "switch context" – is therefore incompatible with or at least problematic for system-based explanations. This source-level explanation is difficult to link to the more abstract processes of change. But besides the practical difficulty of accessing data at this level, there has also been an absence of interest in linking

individual utterance production by a speaker to what is traditionally called "a linguistic change". The interest in the source of linguistic change has received only cursory and unsystematic treatment in work on linguistic change. Notable exceptions include Hockett (1965) and Ohala (1981), who regard the hearer as essential in motivating sound change, Gal (1979, 1989), who points to the role of individual speakers as agents of change as they respond to macrohistorical processes, Milroy (1992, 1999) and Croft (2000), who highlight the role of the individual in the instigation of change, and Drinka (2017, p.12-23), who focuses extensively on the issue of morphosyntactic innovation at the source level. If we accept as an a priori assumption that an explanation of linguistic change must include an explanation at its source and must be able to relate it to the final stages observed at the system level, then there continues to be a gaping hole in our attempts to explain linguistic change.

To the extent that individual speakers want to be polite or choose a more "extravagant" (cf. §5, below) form, or accommodate to and identify with other speakers, they do not do so with the intent to cause the language to ultimately end up being more analytic, more transparent, or more grammaticalized. These are, as a rule, unintended consequences.[1] So it is this juncture of explanatory levels – between the micro, the source level, and the macro, the system level – where attention needs to be directed. For a theory of one of the most salient aspects of language, its persistent change as a mode of existence, a theory is required that ultimately accounts for the whole extension of the process, from micro level to system change under one single perspective, including the false starts, the changes of direction.

Cicourel (1981) has pointed to the deficiencies arising from an approach that does not link macro- and micro-approaches, with "macro" referring to "*typified, normalized, context-free summary descriptions*" (Cicourel, 1981, p.51, italicization in original) of "single and local utterances as if they were self-contained meaningful units" (Cicourel, 1981, p.60). Cicourel's postulate, elaborated on a more general level in explaining evolutional changes by Enfield 2014), applies in a general way to an explanation of linguistic change at large, in particular to the very first inception of any change, at the onset context. An analysis of the very first stages of morphosyntactic variation and innovation would ideally have to include all of this information in the initial stages of an innovation process. Cicourel's discussion highlights the tension between macro- and micro-approaches in an illuminating and radical way.

---

1.   A theory that has in fact addressed the logical relationship between the motivation underlying the individual utterance and resulting system consequences is Keller's theory of unintended consequences of individual actions (1994).

3. Origin of variants

Ever since the early work of Coseriu (1958 / 1974), a distinction has been common between "innovation" and "diffusion." It would appear that most work on the early stages of linguistic change has focused on diffusion, with the most important concept of initial social biases arising through identification and accommodational processes through weak social ties between speakers. These social mechanisms have – rightly – shaped our idea of the early stages of linguistic change to an extent that they actually tend to be identified with notions of the source of linguistic change, as the discussion in Luraghi (2010) and Drinka (2017, pp.16–19) suggest. Work by Mufwene (2001, pp.14–15, 150–151), Croft (2000, p.178), Bergs (2006) and Hernández-Campoy & García-Vidal (2018) is also in line with the argumentation in the present article that it is the choice of the individual speaker that is at the source of linguistic change. However, it is a fact that selecting one variant over others in a social process has the existence of such variants as a logical precondition. They must come from somewhere. It is also clear that the source of variants is different in cases of language contact, where new forms are taken over from the other language in acts of identification, not in a blanket way, but in selective strategies involving processes of "replication" (Heine and Kuteva 2005; and further types of selectivity such as those discussed in Drinka, 2017, p.18).

The notion of innovation discussed here is thus one of "internal" causes of language change, meaning "internal" sources of innovation. This assumes that there is indeed language change that is not contact-related. Even if one assumes that "all" change involves contact, such as, on a very fine-grained level, between small groups of speakers, there is still the issue, at one last remove, that the variants must at some point come from *somewhere*, i.e. must come from one or several innovating speakers. We cannot simply presuppose their existence, but must include them, even if we acknowledge the dominance and ubiquity of contact-induced change in our theorizing as a pristine source event of speaker innovation.

The great difficulty in empirically studying internal innovation is even more severe than in studying the initial stages of diffusion. Recent work in corpus linguistics has offered valuable data for both establishing and further differentiating hypotheses about the emergence of biases (Gries 2012). While corpus linguistics is now indispensable as a methodological tool in analyzing linguistic change and has clearly moved us closer to the early stages of innovations, its contribution is above all to our knowledge of internal diffusion of innovations through linguistic and social space and, although it may suggest hypotheses about the "motivation" for an initial breach of a norm, it does not acknowledge the source innovation by the individual speaker in the individual speech act as responsible. Even if a process

may be difficult to document, this does not mean it does not exist or should be excluded from consideration.

With these restrictions and the focus on internal innovation in mind, the initial stage of a potential beginning of a process of change – the point of speaker innovation – may be conceptualized in the shape of two essential steps:[2]

1. The rise of speaker innovations in the individual's speech production in real time. This essentially amounts to a minimal departure from a default, from some sort of a norm, i.e. a false step.
2. The point when several speakers choose the same "deviant" alternative, resulting in an initial process of diffusion, producing utterances that are reflected in data corpora as a first "bias," starting to turn the initially false step into – ultimately – a normatively right step.

In most accounts, the origin of any change is pre-existing variation in the packaging of ideas or utterance intentions, on the assumption that we keep this "deep" level of – ultimately – intended perlocutionary acts stable, and exclude effects of the utterance level back to this intention level.

The existence of realization variation has been a basic assumption for all discussions of linguistic change (as in evolutionary biology). Variants may exist a priori, either as a kind of variation as in evolution, fortuitous or random, or they may be due to other factors not yet discovered but potentially recognizable as initial biases which may emerge through, for example, corpus analysis. The other alternative seems to be the existence of variants caused by some sort of higher-order systemic tendency through which a language might become more analytical or more structurally transparent. We must then assume that these abstract system tendencies are physically represented in the production mechanism, as factors influencing the choice at an individual utterance level. However, in a radical atomization of linguistic change, it is necessary to ask how these variants come into being in the first place: innovation in an onset context creates variation, and the further spread of innovation, in this first approach to the problem, creates an incipient bias, as set out in Milroy and Milroy (1985). If an initial, minimal departure from a norm is at the source of an eventual system change, we know that one person alone cannot make a rule. For a step to become a right one, or to create a first initial bias, that is, for its social evaluation to change, several people must take the same false step. There are many such false steps, and many such initial biases, few of which are eventually actualized. As Petré (2018) has documented, it is common for different individuals

2. See Petré 2018:3; 28 for another version of the logical steps in language change.

to choose different options, and to do so according to "different degrees of grammaticalizations" (Petré 2018, p. 29). For example, one English speaker in 1580 might say "Come you?", while for others, the more "grammaticalized" "Do you come?" might already be obligatory, illustrating variable levels of grammaticalization.

Some modern studies of linguistic change, especially in the framework of construction theory, pinpoint a moment of speaker innovation when a new form-function semiosis is offered by the speaker for the hearer to construct. What we have termed "speaker innovation" is variously referred to as "switch context" or "onset context" (See Traugott and Trousdale 2013, p. 199 for a broader discussion of this term). The formulations often imply that this "happens", but it is important to remember that this initial act is at some point first performed, under the monogenesis assumption, by one speaker at one time for the first time, even if we may never be able to empirically locate it. As with gravitational waves, the process must be theoretically and logically postulated, even if it is unlikely ever to be observed.

For diffusion to arise, there are two possibilities:

– one person's initial innovation must be taken over by others, for whatever social reason (monogenesis)
   or
– many people must depart in the same direction (polygenesis). What one must ask is, what is it that would make many people make the same initial "mistake", without meta-agreement to do so? Candidates for such a meta-agreement that have been hypothesized to reside in the individual speaker are "natural tendencies," invisible hand motivations, or universal tendencies. Each individual speaker must utilize the same "knowledge" about the next step. This process is completely local, operative only in the individual utterance and has no knowledge of any long-distance teleology.

On the whole, theories of linguistic change have been notoriously silent on the possibility of using the polygenesis criterion as a test of plausibility for views on linguistic changes. For syntax and referential expressions we might assume a specific meaning intention in each individual case, a social motivation, like wanting to achieve politeness, for example, or the opposite. So at the base is some purposefulness at source – wanting to be efficient, wanting to accommodate or distance, wanting to express an innovative aspect of meaning, aptly summarized by Enfield (2014, p.8): "we select the items that best enable us to manipulate other people's attentional and interpretive resources.". All of these intentions are plausible at the level of the individual utterance by the individual speaker. Are these also plausible for a polygenetically-originated bias, as the early source of a norm change and a candidate for diffusion? It is a well-known fact that biases or initial preferences may

well be differently interpreted structurally if they are sufficiently frequent, leading to a new directionality of their own, divorced from the original speaker motivation, which may by then be obscured. Although speakers tend to have an awareness that a certain expression is "old-fashioned," the original motivation for an new meaning for an old form or a new form for an old meaning is mostly lost for the then "synchronic" speaker, blind to the linguistic past, and only identifiable, if at all, for the later linguistic analyst in the study of individual speaker data – historical idiolectal texts – or in the form of statistical biases in computerized historical corpora. Two such cases from English will be cited here, the development of "Do-support" (below) and of the Progressive (in § 5).

A first case in point is the analysis of the historical development of "do" in questions in English. Traditional historical analysis pointed to the preverbal position of modal auxiliaries in a unilinear development. Stein (1990) has documented a very different origin. In a first phase, late Middle English shows a surge in uses of *do* in all syntactic contexts – a process similar in timing and properties to the second case of morphosyntactic change – the development of the English progressive (§ 5). In the second phase, starting in the later 16th century, this semantically suitable, propositionally empty form is co-opted to provide a syntactic bypass for phonotactically dispreferred forms in the second singular, the most frequent form in the spoken language ("thou dost expect" for "thou expect'st"), and quantitatively by far most strongly so in the case of questions ("expect'st thou," bypassed by "dost thou expect"). Essentially, an initial speaker motivation to bypass phonotactically dispreferred structures has created a massive initial quantitative bias in spoken second person dialogic contexts, as a precondition for early entrenchment that was re-interpreted – as a neoanalysis – as a grammaticalization of "do" in questions. Traditional hypotheses on the development of *do*, preferring level-internal explanations, tend to view the innovation as syntax-internal. Syntax-internal hypotheses go back to an older study by Ellegard (1953), which, for all its merits in shaping our knowledge of the historical development of *do*, does not subcategorize according to types of phonological environments (like the inflectional endings plus personal pronouns that cause consonantal clusters in questions) that makes the strong effect of the resulting consonantal cluster on syntactic evasion strategy transparent. That this effect in its strength is overlooked is also a natural consequence of the fact that Ellegard (1953) excludes data and corpora that contain large numbers of 2nd singulars such as plays that contain represented speech (as in the Shakespeare corpus), not to mention the fact that the potential effect of phonology on syntax was hardly considered at the time of Ellegard's writing. The various versions of syntax-internal arguments are summarized and critiqued by Hudson (1997). They show a predilection, typical of structure-based explanations, for unilinearism, as any presupposed

structural motivation should then also be present at the very inception, the source level of the individual utterance, as some "structural pressure," which would then have to be assumed to be present at each individual speech act in determining form choice. So the inclusion of substantially more differentiated variational data for the inception of the process in the 16th century renders an all too straightforward syntax-internal hypothesis questionable, despite its understandable popularity and its apparent simplicity.

The aspect that matters in this context is that some motivations are more plausible at the level of motivations of selection – that is, at the level of the choice in the individual utterance – than others. Amongst these are those identified by Keller (1994), basically a version of economy of behavior in reaching a goal such as articulatory ease. Others are semiotics-based tendencies that are not based on underlying assumptions of a "lonely organism," but assume an implied hearer, such as to assure a sufficiently rich signal for the hearer to trigger in them the construction of the intended hearer meaning, to counterbalance economy-driven attrition, such as conceptualized by Lüdtke (1980b). Traugott (2010, p.3), in her discussion of subjectivization and intersubjectivization, finds the systematic inclusion of the hearer and their interaction with the speaker a perspective that is indispensable in the development of the semantics of expressions in this domain.

In the context of a semiotics-based perspective, tendencies identified in work on so-called natural tendencies (Haiman, 1985; Mayerthaler, 1981; Wurzel, 1984) can also be viewed as candidates for motivating speaker innovations at the source level. It would appear that source-level motivations are the more plausible the more they reflect general cognitive or behavioral tendencies and the less they reflect structural or even universal linguistic tendencies. It may be speculated that these system-inherent tendencies may come into play at later stages after initial diffusion processes, at points where biases are or are not reinterpreted, and restructuring options taken or not. But these secondary processes are then initiated as new processes with speaker innovations and source problems of the very same kind discussed earlier, with the same open questions.

4.   The pragmatics of innovation

Why do we have, in the first place, a sea of variants, one of which is selected to spread through the "weak ties" mechanism? While the issue is a priori clearer for phonology, matters related to morphosyntax are more complex. As regards the rise of morphosyntactic speaker innovations, this is arguably the area where the least work has been done, both empirical and theoretical. Drinka (2017, pp.12–23), a notable exception, sums up the specific situation for the case of language contact. For

the specific case of language contact, she emphasizes the "choosiness" of speakers in adopting forms of the other language depending on what social advantage a form is likely to afford them as an important element in the microstructure of the process (Drinka, 2017, p.18). While this appears to be a key process in contact-related change, for the non-contact source of variants as speaker innovations these forms must be "produced" in the first place. So the formulation employed in the last section of "a variant" being "taken over" by a speaker who is socially linked to the innovator by weak ties requires further qualification, in order to get closer to a pragmatically more satisfying perspective on the process.

Essentially, a speaker innovation is the ad hoc creation of a new sign offer, the offer and acceptance of a slightly and more or less new content (referential, grammatical, discourse orienting). In order to get closer to a window on the microstructure of speaker innovations, we must invoke categories of analysis from interactional pragmatics. Before it "is" a sign, i.e. decoded, comprehended, constructed, and ratified as a new sign, and before neoanalysis happens, a number of other processes and conditions must be at work. If we assume, with Tomasello (1999) that cognition comes before language, a plausible assumption is that, given an utterance intention, the speaker first makes an estimate of what the current knowledge state of the hearer (genre, mutual knowledge, focused content) will be at a given moment. This appears to be the crucial point that enables speaker innovation. The speaker will calculate that the knowledge state of the hearer will afford enough redundancy to choose a term that is slightly "off" the previous remembered use history (described as "linguistic meaning" or "language meaning", or the compositional share). The speaker believes that they can take this "risk" given the calculated knowledge state of the hearer: the speaker trusts that the hearer can still construe their (the speaker's) intended meaning, even though the hearer might have to work a little harder, depending on how far "off" this use is. A corollary of this view is the one current in interactive pragmatics: that the share of "linguistic meaning" in real-time language is by far overrated, and that a substantial amount of extralinguistic and cognitive information is used by hearers to construct that meaning. It is this very contextual redundancy that frees up the "space" and creates the flexibility for innovative use.

Another area of language development shows a parallel standard process involving changes of meaning and solving the problem of discontinuity of discreteness. Pienemann (1998) has elaborated a concept of "comprehensible input" as a key concept in the continuous development of language learning: new forms are always slightly in advance of any given present state of mastery by the learner, but the availability of contextual knowledge guarantees that the learner can construct the intended meaning. Traugott (2010, p.25) points to the role of a similar limit of comprehensibility for innovative meanings:

> [T]he examples discussed here do suggest that some combination of maxims must be at work, such as "Talk like the others talk" (with pragmatic invited inferences of quantification) and "Try something very similar that will not be incomprehensible" (and therefore does not violate the maxim "Talk in such a way that you are understood").

So it is again the redundancy provided by context and its much greater share in guiding the construction of the intended meaning that enables both types of innovations to occur: on the one hand, the endowment of a pre-existing form with a slightly changed meaning, and, on the other, an innovation in the language learner's grammatical repertoire.

Many studies of morpho-syntactic change identify a stage of syntactic ambiguity (involving mostly non-propositional meanings, cf. Stein, 1990), where potentially innovative forms can have several, and several types of, meanings in a transitional phase, with the additional burden of context-based meaning construction placed upon the hearer, before firmer associations between one meaning and one form are observed. This transitional phase of "ambiguity" could then be understood at this level of discussion (but see § 5) as a phase when different speakers offer a variety of more individual meanings, that may be picked up by weak-speaker ties to form a bias that would end up in a newly grammaticalized structure, as did in fact happen in the case of "do" in English.

This mechanism of meaning change exploiting contextually given redundancy, then, appears to be the very kernel process of speaker innovation and, in the final analysis, eventually of linguistic change. The point to emphasize, above all, is that this fundamental process of language use is not something deviant, aberrant, or unusual. For Coseriu (1958 / 1974, pp.95–96) this "Sprachfreiheit der Sprecher" is the essential creativity in "meaning": in each act of meaning, what is referred to is different, as the contextual discourse situation is always different, no matter how slightly. This is also the point emphasized by Cicourel (cf. § 2). In each and every act of communication the expressive needs are different – this is why "change" is inherent in language use. Coseriu's dictum, though by now "old," is still valid. Inquiry into language change deals with "… auf welche Weise die Sprache sich den Ausdrucksnotwendigkeiten der Sprecher angepaßt, beziehungsweise wie und unter welchen Bedingungen das durch die Ausdrucksfreiheit Geschaffene angenommen und verbreitet, das heißt in die Sprachtradition eingereiht und verbreitet … wird" (Coseriu, 1958 / 1974, p.95).[3]

---

3.   'In what way language adapts itself to the expressional needs of speakers, or how and under what conditions the constructs established through expressional freedom are adopted and spread, that is, fit into pre-existing classes and disseminated across the language community'

But how does a sign come to be a sign in a situation of a true speaker innovation? What would make the speakers assume they will succeed in using an old form in a new, deviating way to make the hearer construct a new, deviating meaning? How can they hope that the hearer will construct the intended speaker meaning even though the meaning content and the form – the form of the sign – will depart from an expected default or a range history? How can they be sure hearers will be ready and able to cooperate and do their additional construction work?

Signs do not simply "arise" in the speaker- and hearer-less abstract. Signs are co-created by speakers in a complicated interaction with hearers. It would appear that some of the difficulties and mysteries in explaining language change stem from what Roy Harris (2002) has termed the two major language myths, "telementation" and the "fixed code" fallacy, the tacit presupposition that meaning is – at least ideally – entirely compositional and that the source of meaning extraction is only linguistic meaning with more or less fixed and locked-in boundaries. This all-compositional way of looking at expressions is overvalued and stands in the way of our seeing a potential for change in each individual usage act.

5. Semantic consequences

The most difficult issue is to question what the benefit would be for deviating from a pre-existing norm. The meaning to be expressed may itself be novel, or speakers may have non-propositional reasons to use a novel form, availing themselves of a structural resource that happened to be available at that point. Petré (2017) presents a seminal study of the microstructure of an innovation, the historical rise of the [BE V*ing*] ("I am singing") construction. As it offers crucial empirical findings relevant to the present context of discussion, it will be reported at greater length. Petré (2016, 2017) describes the development of the modern meaning of "ongoingness" or "in progress" starting from a Middle English situation where the meaning was stative and adjective-like as in Mod. Eng. "Wolves are living on the outskirts of Paris, wildlife groups claim. " (Petré 2017, p. 230), where the meaning can be paraphrased as static "exist", or "are present". A further development supplied a structure that was the hotbed for the development described here in the form of a preponderance of past progressives in subordinate clauses: "Middle English saw the rise of subordinate clauses that provided a background frame to some foregrounded event expressed in the main clause." (Petré 2017, p. 231), as the increasing occurrence of [BE Ving] was pragmatically reinterpreted as "not [...] predicating a temporary state or quality of a non-agent, but of giving background" (Petré 2017, p. 231) to the information supplied in the main clause. A typical example in a subordinate clause in past tense is represented in ex. 4 (Petré 2017, p. 230):

(4)   "So the meanwhile that thys knight **was makynge** hym redy to departe, there com into the courte the Lady of the Laake. (a1470)"   [emphasis Petré's] ["So while this knight was getting himself ready to depart, there comes into the court the Lady of the Lake" D.S.]

With an increase in subordination in Middle English, "the interpretation of on-goingness was semanticized" (Petré 2017, p. 231, see also his figure 1, p.232) and extended to main clauses and the present tense; with concomitant loss of cotextual features like adverbials defining tense and discourse status, the progressive semantics of the form was in principle established. However, the next stage, the shift from the "preferred past-tense niche to present-tense main clause" (Petré 2017, p. 232) and the massive extension to uses in the present tense, as observed from the end of the 16th century onwards, is also seen as due to new expressive possibilities. The form itself is felt to be new in the present tense and has the potential to be used to express "extravagant" meanings: emphasis, urgency, subjective involvement, marked by syntactic and lexical markers that force an emotional and involved interpretation (Petré 2017, p. 236):

(2)   you are now leaving all the pleasures of the world

(3)   Art thou delighting thy self in the society of the Saints

(4)   whither art thou going

As the experiencer of these emotions is in the present, with adverbials referring to the present time (typically "now", currently also used as intensifier), this configuration of present tense and presence of adverbials like "now" leads to a re-interpretive strengthening of the semanticization of the modern progressive meanings of the form also in the present tense that is now present even without supporting lexical or syntactic structures: "And indeed; while three out of four present-tense instances in the period 1501–70 in Petré (2016a) have the time adverb *now*, the presence of *now* has dropped to 10 percent in the period 1575–1640."

This contextual material in "extravagant" uses was a crucial context of use for extending the progressive meaning to the present tense in a process of the "coercion" of a structure with a previously static meaning of the structure into a progressive, dynamic, and no longer static use: "…it is an entire construction ([BE Ving]) that is made to conform to the meaning of its embedding structure, i.e. the clause containing NOW or HERE." (Petré 2017, p. 239)

This particular case of language development, together with the example from the history of *do*, highlights several theoretical points made in the previous discussion. The starting point for linguistic innovations, for a speaker innovation, seems to be some motivation in either form or function. The two cases adduced here, English *do* and the progressive, represent these two different types of initial

motivation for the speaker. While in the case of *do*, that segment of the development that triggers grammaticalization in questions and negation does not involve an expressive motivation, the case of the progressive seems to have involved two such motivations at different stages. The term "extravagance" was used here simply as a cover term for several possible types of expressive motivations in the source context (cf. the discussion in Traugott and Trousdale 2013, pp.124–127). These motivations, to the extent that they are meaning-related, are likely to be different in each individual case and especially in each individual usage case. Form-related motivations may also be more uniform (such as consonantal cluster avoidance or preferred prosodic patterns). This aspect throws a light, again, on the issue of polygenesis or monogenesis: given the wide range of potential semantic, pragmatic, and social motivations defining an onset context, in order to postulate polygenesis, we will have to postulate that such a motivation is present in several or many speakers in the same way in order to create what is corpus-statistically manifested as an initial bias. This cover notion of "extravagance" is also quite plausible from a semiotic point of view of constructional iconicity (another facet of naturalness): an extra element in form is construed by the hearer as an extra element in meaning. Such is the case with all marked syntactic constructions in English, like preposings or non-grammaticalized inversions (Dorgeloh and Wanner forthcoming).

Some more general theoretical points are illustrated by the two cases cited in some detail. What Traugott and Trousdale (2013, p. 27) describe as several "pre-constructionalization constructional changes" are really several false steps in the sense of local departures from norms. There are likely many such changes, and the considerations regarding their poly- or monogenetic character equally apply to these phenomena of neoanalysis. They all involve the same pragmatic processes of deviation from a norm, of sign-creating through inference processes. The act of norm-breaking in a single speech act prior to any semanticization has been conceptualized as "coercion" in the framework of construction grammar (Traugott and Trousdale 2013, p. 204f), but can be generalized to be included as a central initiating mechanism in processes of change. The process basically presupposes a measure of "underspecification", of redundancy or "free space", causing "inferential procedure or type-shifting" (Traugott and Trousdale 2013, p. 204), with "default […] compositionality […]overridden" and the semantic conflict resolved "online" (Traugott and Trousdale 2013, pp. 204, 205). In the final analysis, then, the old conflict between the structural categoricalness of compositionality and the fact of change has to be given up and pragmatics must come to the rescue.

As mentioned in §4, the transitional phase is often characterized as a phase of ambiguity or vagueness of use of a form in a transitional stage. From a structural and semantic point of view, in terms of construction grammar, we observe a sequence of partial processes of pre-constructionalizations, prior to final semanticization or

full constructionalization, which, in a more coarse and quantitative perspective, is reflected in a state of "ambiguous" forms. Traugott and Trousdale (2013, p. 201) suggest replacing the self-suggesting term "polysemy" with a concept of "heterosemy" that specifically captures the range and sequence of neoanalyses or new meanings in partial processes leading up to final semanticization. If this concept involves a semasiological viewpoint, it is appealing to extend our more pragmatic analysis to this ambiguity stage. Obviously speakers individually produce different new semiotizations (as observed by Petré 2018), creating a quantitative average impression as undifferentiated "ambiguity" and polysemy. To put it more provocatively, it appears that a structuralist and categorical and compositional meaning-maximal view has forced the theory of linguistic change into implausible assumptions about ambiguity. Arguably, indeterminacy only exists for the analyst but not for the speaker, and ambiguity is a construct in the eye of the generalization-seeking beholder. Petré (2018, p. 27) can also be followed in his assumption that contextual determination, and reduction of compositional share, is higher in less autonomous spoken language. This is also plausible in the case of the grammatical reinterpretion of *do* in second person contexts, as the prototypical context of spoken language.

For the individual speaker, then, there never is indeterminacy. Ambiguity there may be for the analyst trying to force a structural generalization. This more pragmatic view of the individual speech act presupposes with a vengeance a meaning-minimalist view, curtailing compositionality in order for an expression to combine with a multitude of contexts for different kinds of innovative semantics in different contexts, resulting from a semiological point of view, in a situation of polysemy such as observed in both the cases of *do* and the progressive. Consequently, these different meanings can only be identified by local interpretation of the individual occurrence. Quantitative corpus analysis is unhelpful here. At best, quantitative increases must be taken as reflecting increases in types and tokens of such micro-meanings.

When it comes to the choice of linguistic expressions to express the desired content, in order to be able to assume a measure of deviation from a norm, we must assume a measure of continuity (Traugott and Trousdale 2013, p. 204 "persistence" and "look-back" effects). What is often invoked is a notion like analogy, operating on both form and function. In fact, apart from grammaticalization, analogy is arguably the process most often adduced for grammatical change (Kiparsky 2012). In traditional approaches, analogy is discussed on the level of system change; in the present context of discussion, analogy is better conceived of as subsumed under a broader concept of similarity (Stein 1988) between the usage history of a form and the present expressive needs. What matters is that, again under the assumption of polygenesis, several or many speakers would be subject to the same similarity-constraining factors and would make, in a concrete speech act, the same

similarity-based judgments about which pre-existing expression is eligible, in addition to other motivations like iconicity, to be used in an act of norm-deviation. In other words, the local "directionality" must be shared.

Concepts like "reanalysis" and "restructurings" make these processes sound as if they were operating at an abstract level, when it is the activity, a process of semiotization, by the individual mind in the individual speech act, that is operating. The "reanalysis" is actually a hypothesis about an epiphenomenal process on an abstract level, after the real-time norm-breaking has been "perpetrated" in logically and temporally prior processes. It must not be forgotten that it is always speakers and hearers who work on this in a cooperative effort. It is speakers and hearers who cooperate, and it is the speaker who is speculating what formulation will get the hearer to "resolve the semantic conflict on-line" (Traugott and Trousdale 2013, p. 205), where we have tried to explicate "online-ness" in more pragmatic categories, with speakers monitoring how successful or otherwise they have been. It is always salutary for theories of language and linguistic change to have their theories grounded in real-time pragmatics. In addition, it is a moot point if a concept like "reanalysis" is to be predicated on an individual or a group of speakers. Again, the same reanalysis must be performed by several speakers independently of each other for reanalysis to take place on any quantitatively significant level. This online-ness in the interaction between speakers is represented in the concept of "cooptation" that describes "a cognitive-communicative operation whereby some fragment of linguistic discourse is transferred from one domain of discourse to another." (Heine, Kaltenböck, Kuteva and Long 2017, p. 1). The "coopted units" typically concern meanings that are concerned with "a dynamic, real-time, interactional tool for planning and processing information" (Heine, Kaltenböck, Kuteva and Long 2017, p. 6). It would appear that such a conception of language use would be most hospitable to a view of the onset situation that can accommodate the role of speaker-hearer interaction in performing an affordable and intentional minimal breach of a norm. In addition, it is significant that this process is observed at the level of the individual utterance and will generate discourse meanings, – a process that is well-established on the system level (e.g. Traugott, 1982) and has in fact also been documented for *do* by Stein (1985, pp. 294–298) and Stein (1990, pp. 31–42).

From such a perspective, "extravagance" as a motivation for the sea of speaker innovations is the intention of the speaker to get the hearer to construe a novel or additional meaning by using a norm-breaking form, which forces a "mismatch" (Traugott and Trousdale 2013, p. 201) between the remembered and abstracted usage history and the present context of usage. This, then is the initial false step in the onset context. In the case of *do* the initial surge of occurrences in all syntactic contexts (Stein 1990) seems to have involved many types of false steps with different but related types of "extravagant" uses of the form, much in the same way

as in the case of the early history of the progressive. But the case of *do* did not, at this first stage, have further consequences. In contrast, the phase leading to the present grammatical functions of the form (Stein 1990, pp. 143–194) seems to have involved phonotactically motivated onset contexts with "misuses" of a form that was not "locked in" to express propositional or grammatical meanings. So both cases discussed here involve several partial processes that do not involve one single directionality, but several local directionalities.

In this context it will be illuminating to look at major directional long-term trends such as those identified in work on historical change. Recent studies, (e.g. Kuteva, Heine, Hong, Long, Narrog, and Rhee 2019, pp. 6-7) have demonstrated that grammaticalization is a long-term process which appears to be non-reversible. Translated into the perspective offered in this paper, no one speaker has in mind to "weaken" the semantics or effect a category change in the long run. This in turn ultimately means that minimal deviations are performed by speakers only in one direction along an ordered cline of phenomena, where the individual speaker "knows" the direction of the next step in a sort of "natural" way relative to any given stage. Exactly what the nature of this knowledge in the individual is, and whether or not this knowledge is related to so-called "natural" tendencies remains to be seen in further study.

The hypothesis of the rise of the English progressive is, of course, another argument for caution, as put forward earlier (§ 4) and as formulated by Traugott and Trousdale (2013, pp. 17, 124f.) that linguistic change may develop in such a way that the original motivation in the first "onset" context may be different from the direction of later developments. In other words, it will not be possible to determine the initial motivation of a diachronically later stage. Just as the initial phonotactic motivation in the case of *do* cannot be gleaned from the syntactic effects in later stages, so morphosyntactic change cannot be explained by extrapolating backwards from looking at a later stage of the language. Local directionalities such as the effect of local "analogies" in microprocesses involved in local processes of constructionalization are plausible at this local level. But it is very difficult to imagine a "translation" of hypothesized large-scale morphosyntactic processes of change to operate in the same type of micro-process as described for local processes of pre-constructional changes. This, then, presents a difficulty or at least an explanatory challenge for any conception of directionality of linguistic change, especially if they are hypotheses derived from synchronic end-states. The whole range of development from onset to finished state cannot a priori be assumed without analyzing its microstructure, with only one directional common denominator that explains the whole range of development from onset to structural near-completion. There are obviously bound to be such cases, but the assumption of one-denominator directionality from conception or birth to maturity cannot be taken for granted. Such a directionality – and not

a teleology – from onset context to final structural result seems to be suggested at least plausible for a complex, but well documented process like grammaticalization, where only some of many incipient processes that are observed, but where only some are carried through to completed final stage (Hopper 1991).

Because of the potential breadth of the range of initial individual motivations, corpus-based methods can only go so far in either suggesting (corpus-driven) or testing (corpus-based) hypotheses for source contexts, and then only to the extent they can narrow down the context of production so as to isolate the individual context of production as much as possible. Naturally, the methodology will be more successful the closer it gets to an individual person. Given the complexities of production contexts, it is obvious that the hermeneutics of "understanding" the individual first occurrence must still be the optimal access to identify possible meaning motivations, followed by linguistic indicators, such as the occurrence of "now" in the case of the rise of the English progressive in its present grammatical function.

Another issue is why hearers cooperate in constructing the meaning intended by the speaker. Invoking a more general cooperative principle or a presumption of intended communicativity such as that postulated by Bach and Harnish (1979) leads to an inference or implicature about the most plausible intended meaning in a given context. This aspect of the whole process can also be conceived of as an act of closure. This, in turn, is the basis, on the more linguistic level, for the process that has been termed "coercion" and refers to a process "whereby the meaning of the lexical item conforms to the meaning of the structure in which it is embedded" (Petré 2017, p. 239). Coercion is an outflow of a coercive tendency of the mind to force closure.

Interactive pragmatics has taught us that each use of an expression is an innovative use: a sign, in a more pragmatic view, is something that arises in a fully specified context, different in each new act of speaking. And as language is used in such "every-time-different," fully-specified contexts, so linguistic change is in principle built in to language use, as each and every speech act is differently contextualized. So, in essence, each use of a form is really a false step. Some of these contexts may be just a little further away from the remembered previous uses, and be, in effect, felt to be deviating from a previously perceived norm. The speaker intends for the hearer to notice the slight deviation or extension to a context where it might not have been used before. To do so one might have an extra meaning motivation. Ultimately, each language use is really a norm deviation only if our view of what linguistic communication is has a fixity, an immutability of a synchronic view, as a presupposition. It is invariably the "right" step if it is seen as the essential creativity of language use as the default. Overlooking this creativity in the process of "meaning" is one of the shortcomings of a synchronically-oriented view of language applied in an attempt to "explain" language change.

6. Conclusion

The very discreteness of the analytical notion of "structure", as well as the absence of the roles of hearer and speaker, has resulted in a gross misconception that "somehow," by a process of telementation, meaning can travel bodiless from speaker to hearer; this mistaken view has clouded a realistic perspective on language change, on the "Seinsweise" of language. Weinreich, Labov, and Herzog (1968) and Coseriu (1958/1974) represent early steps in recognizing and incorporating variation in linguistic theory, opening up a perspective for further steps towards a realistic theory of language change. Milroy & Milroy (1985) and Milroy (1992, 1999) went on to construct a view of a socially plausible concept of diffusion, with great potential to account for processes involved in contact-induced change. A further radical step towards a realistic comprehensive view of the process would be to take a close look at where variants come from, where innovation really occurs. This step, the analysis of the onset context in the individual speech act of the individual, requires the inclusion, in morphosyntax, of semantics and pragmatics in a way historical linguistics has not been accustomed to providing. The fact that there are serious methodological difficulties for obtaining data for these very first innovations must not serve as an excuse for disregarding or discounting their role in creating the hotbeds for internally-caused change and in promoting a comprehensive view of language change as an essential explanatory element.

Acknowledgements

References

Bach, K. & R. M. Harnish (1979). *Linguistic communication and speech acts*. Cambridge: M.I.T. Press.

Bailey, C.-J. (1980). Old and new views on language history and language. In H. Lüdtke (Ed.), *Kommunikationstheoretische Grundlagen des Sprachwandels* (pp. 182–252). Berlin, New York: de Gruyter.

Bergs, A. (2006) Language change and the role of the individual in historical social network analysis. *Logos and Language. Journal of General Linguistics and Language Theory*. VI.2. 30–54.

Bybee, J. (2010). *Language, usage, and cognition*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511750526

Cicourel, A. (1981). Notes on the integration of micro- and macro-levels of analysis. In K. Knorr-Cetina, & A. Cicourel (Eds.), *Advances in social theory and methodology: Toward an integration of micro- and macro-sociologies* (pp. 51–80). Boston: Routledge and Kegan Paul.

Coseriu, E. (1958 / 1974). *Synchronie, Diachronie und Geschichte*. München: Fink. (First published as *Sincronía, diacronía e historia. El problema del cambio lingüístico*. Madrid: Gredos.)

Croft, W. (2000). *Explaining language change: An evolutionary approach*. Harlow: Longman/Pearson.

Diewald, G. (2009) Konstruktionen und Paradigmen, *Zeitschrift für Germanistische Linguistik*, 37 (pp. 445–468). https://doi.org/10.1515/ZGL.2009.03

Dorgeloh, H. & A. Wanner (forthcoming) *Discourse Syntax: English grammar beyond the sentence*. Cambridge: Cambridge University Press.

Drinka, B. (2017). *Language contact in Europe: The periphrastic perfect through history*. Cambridge: Cambridge University Press.

Ellegard, A. (1953) *The auxiliary do: The establishment and regulation of its use in English*. Stockholm: Almqvist & Wiksell.

Enfield, N. J. (2014) *Natural causes of language: Frames, biases, and cultural transmission* (Conceptual Foundations of Language Science 1). Berlin: Language Science Press. https://doi.org/10.26530/OAPEN_533873

Gal, S. (1979) *Language shift: Social determinants of linguistic change in bilingual Austria*. San Francisco: Academic Press.

Gal, S. (1989). Language and political economy. *Annual Review of Anthropology* 18, 345–67. https://doi.org/10.1146/annurev.an.18.100189.002021

Gries, S. (2012) Frequencies, probabilities, and association measures in usage-/exemplar-based linguistics: Some necessary clarifications. *Studies in Language*. 36.3, 477–510. https://doi.org/10.1075/sl.36.3.02gri

Haiman, J. (1985). *Natural syntax*. Cambridge: Cambridge University Press.

Harris, R. (Ed.) (2002). *The language myth in Western culture*. Richmond, UK: Curzon Press.

Heine, B., & Kuteva, T. (2005). *Language contact and grammatical change*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511614132

Hernández-Campoy, J. & García-Vidal, T. (2018). Style-shifting and accommodative competence in late Middle English written correspondence: Putting audience design to the test of time. *Folia Linguistica Historica* 39: 383–420.

Hilpert, M. 2013. *Constructional change in English: Developments in allomorphy, word-formation and syntax*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139004206

Hockett, C. (1965). Sound change. *Language*, 41(2), 185–205. https://doi.org/10.2307/411873

Hopper, P. J. (1987). Emergent grammar. In J. Aske, N. Berry, L. Michaelis, & H. Filip, eds., Berkeley Linguistics Society 13: *General Session and Parasession on Grammar and Cognition*, 139–157. Berkeley: Berkeley Linguistics Society.

Hopper, P. J. (1991) On some principles of grammaticization. In: E. C. Traugott & B. Heine (Eds.). *Grammaticalization. Vol. I. Focus on Theoretical and Methodological Issues*. (pp. 17–37.) Amsterdam: Benjamins. https://doi.org/10.1075/tsl.19.1.04hop

Hudson, R. (1997). Inherent variability and linguistic theory. *Cognitive Linguistics*, 8(1), 73–108. https://doi.org/10.1515/cogl.1997.8.1.73

Hundt, M., Mollin, S., & Pfenninger, S. (Eds.). (2017). *The changing English language: Psycholinguistic perspectives*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316091746

Keller, R. (1994). *On language change: The invisible hand in language*. London, New York: Routledge.

Keller, R. (1997). In what sense can explanations of language change be functional? In J. Gvozdanovic (Ed.), *Language change and functional explanation* (pp. 9–20). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110813753.9

Kiparsky, P. (2012). Grammaticalization as optimization. In D. Jonas, J. Whitman & A. Garrett, *Grammatical change. Origins, nature, outcomes* (pp. 15–51). Oxford: Oxford University Press.

Kuteva, T., Heine, B., Hong, B., Long, H., Narrog, H., and Rhee, S. (2019). *World lexicon of grammaticalization, 2nd extensively revised and updated edition*. Cambridge: Cambridge University Press.

Lüdtke, H. (Ed.). (1980a). *Kommunikationstheoretische Grundlagen des Sprachwandels*. Berlin: Walter de Gruyter.

Lüdtke, H. (1980b). Sprachwandel als universales Phänomen. In H. Lüdtke, (Ed.), *Kommunikationstheoretische Grundlagen des Sprachwandels* (pp. 1–19). Berlin: Walter de Gruyter.

Lüdtke, H. (1980c). Auf dem Weg zu einer Theorie des Sprachwandels. In H. Lüdtke, (Ed.), *Kommunikationstheoretische Grundlagen des Sprachwandels* (pp. 182–254). Berlin: Walter de Gruyter.

Luraghi, S. (2010). Causes of language change. In S. Luraghi, & V. Bubenik (Eds.), *The continuum companion to historical linguistics* (pp. 354–366). London, New York: Continuum.

Mayerthaler, W. (1981). *Morphologische Natürlichkeit*. Wiesbaden: Harassowitz.

Milroy, J. (1992). *Linguistic variation and change: On the historical sociolinguistics of English*. Oxford: Blackwell.

Milroy, J. (1999). Toward a speaker-based account of language change. In E. Jahr (Ed.), *Language change: Advances in historical sociolinguistics* (pp. 21–36). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110807653.21

Milroy, J., & Milroy, L. (1985). Linguistic change, social network and speaker innovation. *Journal of Linguistics*, 21, 339–384. https://doi.org/10.1017/S0022226700010306

Mufwene, S. (2001). *The ecology of language evolution*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511612862

Ohala, J. (1981). The listener as a source of sound change. In C. S. Maseck, R. A. Hendrick, & M. F. Miller (Eds.), *Papers from the parasession on language and behavior* (pp. 178–203). Chicago Linguistic Society.

Petré, P. 2017. The extravagant progressive: An experimental corpus study on the history of emphatic [BE Ving]. *English Language and Linguistics* 21: 227–250. https://doi.org/10.1017/S1360674317000107

Petré, P.. 2018. "The real-time dynamics of the individual and the community in grammaticalization," *Language* 94: 4, 1–35

Pienemann, M. (1998). *Language processing and second language development: Processability theory*. Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.15

Stein, D. (1985). Discourse markers in Early Modern English. In R. Eaton, O. Fischer, W. Koopman, F. van der Leek, (Eds.), *Papers from the 4th International Conference on English Historical Linguistics, Amsterdam, 10–13 April, 1985* (pp.283–302). Amsterdam: Benjamins. https://doi.org/10.1075/cilt.41.21ste

Stein, D. (1988) Semantic similarity between categories as a vehicle of linguistic change. *Diachronica* 5:1, 1–20. https://doi.org/10.1075/dia.5.1-2.02ste

Stein, D. (1990). *The semantics of syntactic change: Aspects of the evolution of do* in English. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110846829

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge: Harvard University Press.

Toolan, M. (2002). The language myth and the law. In R. Harris (Ed.), *The language myth in Western culture* (pp. 159–182). Richmond: Curzon.

Traugott, E. C. (1982). From propositional to textual and expressive meaning; Some semanticpragmatic aspects of grammaticalization. In W. P. Lehmann and Y. Malkiel (Eds.), *Perspectives on Historical Linguistics* (pp. 245–271). Amsterdam: Benjamins. https://doi.org/10.1075/cilt.24.09clo

Traugott, E.C. & Heine, B. (Eds.) (1991) *Approaches to Grammaticalization. Vol. I. Theoretical and methodological issues*. Amsterdam: Benjamins.

Traugott, E. C. (2010). (Inter)subjectivity and (inter)subjectification: A reassessment. https://doi.org/10.1515/9783110226102.1.29

Traugott, E. C. & G. Trousdale (2013). *Constructionalization and constructional changes*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199679898.001.0001

Weinreich, U., Labov, W., & Herzog, M. (1968). In: W. P. Lehmann & Y. Malkiel (Eds.). *Directions for Historical Linguistics*. Austin: University of Texas Press.

Wurzel, W. U. (1984). *Flexionsmorphologie und Natürlichkeit: Ein Beitrag zur morphologischen Theoriebildung*. Berlin: Akademieverlag.

Corpus linguistics & morphosyntax

The rich get richer

Preferential attachment and the diachrony of light verbs in Old Swedish

John D. Sundquist

Purdue University

This study investigates five common light verbs in an Old Swedish corpus of texts from 1225–1526, focusing on the increasing frequency with which light verbs combine with an ever-widening range of NP-objects. Tools of statistical analysis from network science and ecology are used to track changes in frequency and diversity. Results indicate that light verbs are a small, closed class of verbs with unique historical properties that set them apart from more lexically-specific verbs. The way in which they grow more frequent and diverse is typical of the "rich-get-richer" phenomenon of preferential attachment in studies of dynamic network growth.

Keywords: light verbs, Old Swedish, frequency, corpus linguistics, network science

1.   Introduction

In cross-linguistic research on the diachrony of light verb constructions (LVCs), recent studies have highlighted several unique historical properties of this small class of verbs.[1] Light verbs (LVs), as first pointed out in Jespersen's (1942) discussion of English verbs like *make* or *take*, are semantically deficient verbs that can be combined with a wide variety of object complements to form expressions that could otherwise be expressed by a simple verb (e.g., *make an offer* vs. *offer*, or *take a walk* vs. *walk)*.[2] Butt & Lahiri (2013) note that, in contrast to auxiliaries, LVs are

---

1.   For an overview of recent research on light verb constructions from a diachronic perspective, see Bowern (2008), Butt & Lahiri (2013), Elenbaas (2013), and Ronan (2014).

2.   Following Brinton and Akimoto (1999), I assume that the combination of a light verb and its complement is a type of complex predicate – a general term used to refer to multi-word verbal expressions including phrasal verbs (*to turn over*), verb+PP combinations (*to set in motion*)

"diachronically pertinacious" and are unlikely to be reanalyzed (26). In addition, light verbs like *make* or *take* are usually among the most frequently-used verbs throughout a language's history, and they exhibit remarkable continuity over time. The short list of the four or five most common LVs in Present Day English, for example, is identical to that of Old English, and the meanings of these verbs are very similar across many languages (Brinton, 2008, p. 46). Moreover, as Butt & Lahiri (2013) describe them, LVs act as verbal *passepartouts* that fulfill a wide variety of lexico-grammatical functions due to their semantic generality (24).

The goals of the study are twofold: to track the frequency and diversity of LV-NP pairings in Old Swedish texts and to use tools of statistical analysis of biodiversity to explain the historical trends. Empirical analyses of LVs in the history of English reveal that they tend to go through periods of increasing frequency (Elenbaas, 2013; Ronan, 2014) and combine with an increasingly diverse number of objects (Hitunen 1983, Kytö 1999). This study fills several gaps in the literature in its focus on early stages of Swedish which have remained largely unexplored. In contrast to the majority of historical linguistic research on light verbs that has focused on early English (e.g., Brinton and Akimoto, 1999; Claridge, 2000; Ronan, 2014; Elenbaas, 2013), the current study analyzes examples extracted from texts from the Early Old Swedish (1225–1375) and Late Old Swedish (1375–1525) periods.

This study also makes use of a novel approach to analyzing light verb data by implementing tools of analysis from network science and ecology research. We calculate the frequency of LVCs and track the growth of light verbs' co-occurrence networks with NP object complements over time using several statistical measurements of biodiversity.

2.   Background

2.1   Terminology

There has been a lack of consensus in diachronic or synchronic studies on what terminology to use when referring to LVs and the entirety of multi-word phrasal units that contain them.[3] In terms of what constitutes an LV, there is some agreement

---

or verb+Adj expressions (*to make clear*). Discussion in this study is limited to verbo-nominal complex predicates (Light Verb + NP, e.g., *make a list*) with transitive (both mono-transitive and di-transitive) verbs. I will refer to this sub-group of complex predicates as Light Verb Constructions, or LVCs. For further discussion of how the verbo-nominal type fits in with the broader category of complex predicates, see Brinton and Akimoto (1999), Claridge (2000), and Bowern (2008).

3.   These expressions have been referred to as complex verbal structures (Nickel, 1978), composite predicates (Brinton, 2008), expanded predicates (Algeo, 1995), *constructions à verbe support*

that they are, in general, different from their full verb counterparts; they have been described as "semantically deficient" versions of full verbs (Bowern, 2008, p. 163) or that they "have a rather general meaning and are semantically more lightweight than the same verb would have been in a normal context" (Allerton, 2002, p. 172). For the most part, they resemble their full verb counterparts grammatically yet are generic in meaning when combined with various abstract NP objects.

For the purposes of the present study, I limit the analysis to a small set of highly frequent, transitive verbs in Old Swedish (OSw) that are commonly referred to in previous research as light verbs. Many diachronic studies of early English corpora follow similar procedures and discuss one or more of the five typical light verbs (*make*, *take*, *give*, *have,* and *do)*, as in studies in Brinton and Akimoto's (1999) volume and Iglesias-Rábade (2001) or Elenbaas (2013). As we will see in the methodology section, the five Old Swedish verbs analyzed are similar to their English counterparts: *fa* 'get/receive', *giva* 'give', *göra* 'do/make', *hava* 'have', and *taka* 'take'.

Regarding LVCs as whole phrasal units, there is general agreement that the verbo-nominal type of complex predicates consist of an LV and a NP object that jointly contribute to the overall meaning of a phrase. Butt (2010, p. 49) states that such LVCs "predicate as a single unit, i.e., their arguments map onto a monoclausal syntactic structure." For example, the Swedish light verb *ge* 'to give' in (1) combines with the object *svar* 'answer', in comparison to sentence (2), with the simplex verb *svara* 'to answer':[4]

(1) *Förhoppningsvis kan vi ge svar så tidigt som möjligt.*
     hopefully   can we give answer so early as   possible
                                                          (Bloggmix 2008)

(2) *Han svarade att den risken finns.*
     he   answered that that risk   exists   (Åbo Underrättelser 2013)

The light verb *ge* expresses a generic meaning rather than the meaning of its full verb counterpart, but with the combination of the NP object *svar* 'answer', the meaning of the phrase as a whole is made clear. Various approaches have shown that LVCs like this function as a unit, with the NP carrying the semantic weight of

---

'support verb constructions' in studies of French (Giry-Schneider, 1987), and *Nominalisierungsverbgefügen*, a specific type of *Funktionsverbgefügen* (von Polenz, 1987).

4.   Examples were found using corpora available through the online concordance search tool, Korp, from *Språkbanken* (The Swedish Language Bank). Examples from Modern Swedish are cited according to the corpus names in Korp and the year in which the texts were written. Old Swedish examples will be cited according to the abbreviation of the text in which the example occurs, following the titles provided in the *Fornsvenska Textbanken* corpus. The texts are listed in the Appendix. For further information on Korp, see Borin et al. (2012).

the phrase and expressing the verbal action, while the semantically-deficient light verb functions merely as a marker of inflection. (Quirk et al., 1985; Claridge, 2000; Ronan 2014).

These views on classification in studies of English and other languages are reflected in both synchronic and diachronic studies of Swedish as well. Dura (1997), for instance, follows Ekberg's (1993) analyses of so-called *abstrakta övergångsfraser* 'abstract transitional phrases', focusing on the subset of *stödverb* 'support verbs', which she defines as semantically light verbs that occur in fixed, idiomatic expressions with bare NPs, as in *byta namn* 'change name, rename' or *ha fest* 'have (a) party'. In a similar way, Cinková (2009a, 2009b) undertakes an analysis of LVCs by taking the 20 most frequent verbs in Swedish as a starting point. Following Viberg (1990), she highlights so-called "basic verbs" like *göra* 'make/do' or *ta* 'take', among others, citing their simple phonological form, high frequency, broad cross-linguistic distribution, and the fact that they often have "secondary meanings" that differ from the full verbs' meanings (Cinková, 2009b, p. 122).

There has been less attention paid to LVCs in earlier stages of Swedish than in Modern Swedish. Although Delsing's (1998) study of support verbs is mainly a synchronic study, he provides useful data on some historical aspects of LVCs in earlier stages of Swedish. He lists three sub-groups of support verbs, exemplified here in (3), (4), and (5) (adapted from Delsing, 1998):

(3) *ha/få behov av* 'have a need for'
   föra samtal om 'carry on a conversation about'

(4) *få lust till* 'get in the mood for'
   lämna/få besked om 'give an answer concerning'

(5) *ta hand om* 'take care of'
   sluta fred med 'make peace with'

The expressions in (3) may be replaced by a simplex verb (*behöva* 'need', *samtala* 'converse') related to the predicate noun, while those in (4) could be paraphrased by an unrelated verb (*tycka om* 'like', *berätta* 'tell'), and those in (5) cannot be paraphrased easily by a single verb (Delsing, 1998). Much like Dura (1997), Delsing defines support verb constructions as only those that contain bare NPs–without NP-modification through indefinite, definite articles, quantifiers, or adjectives. Moreover, Delsing includes in his examples from Old Swedish those that contain a PP complement to the noun, as indicated by the modern equivalents in (3), (4), and (5).

2.2 Historical trends and the frequency of LVCs

Previous research on the history of English has shown that LVCs tend to increase over time (Elenbaas, 2013; Ronan, 2014). Some factors that have been found to favor the growth of LVCs include an increase in LVCs with zero-derived nouns (Iglesias-Rábade, 2001, p. 13) and the rise of indefinite articles (Elenbaas, 2013, p. 60). This latter trend has often been connected to grammaticalization and the rise of aspectual differences between LVCs and lexically-equivalent verbs (e.g., *bathe vs. take a bath*) (Brinton, 2008, p. 48). In spite of these changes, the five English LVs (*do*, *give*, *have*, *make*, and *take)* have remained the most frequent LVs in complex predicates throughout the history of English (Traugott, 1999).

In the same way, we see a trend towards greater expansion of LVCs over time as each verb occurs with a wider range of different NP objects. Hiltunen (1983, p. 28) describes the expansion of LVCs to include a greater diversity of NPs, particularly from the Old English to Middle English periods. Iglesias-Rábade (2001) found similar evidence of an increase in new LV-NP pairings in Middle English, pointing out the increase from 42 different LVCs with *taken/nimen* in the Middle English Dictionary in the 13th century to 148 in the 14th century. Kytö (1999) corroborates these findings for Early Modern English, presenting a detailed analysis of the type-token ratio of certain LVs as a way to track the relative diversity of LVCs. Although the type-token ratio of each LV varies, with *make* and *do* taking the greatest variety of nouns, Kytö highlights the high frequency of rare LVCs for all five LVs in which a particular NP occurs only once (1999, p. 172).

2.3 Possible explanations for historical trends among LVCs

Both external and internal linguistic explanations have been offered to explain these two trends. In terms of external linguistic factors that may have played a role in the increase of LVCs over time, several scholars suggest that some LVCs are calques while others are borrowings.[5] Traugott (1999) and Brinton (2008) also point out that LVCs in Old English and Middle English become more common as a result of influence during the period of contact with Old Norse. Ronan (2014, p. 31) also suggests that language contact plays a role in the increase in Middle English, since substantial increases in LVCs take place during the period of French-English contact. Despite these arguments in favor of the influence of other languages, however, others such as Traugott (1999) and Matsumoto (1999) are quick to point out that

---

5. See Traugott (1999), Iglesias-Rábade (2001), and Brinton (2008) for a helpful overview on the topic of foreign influence on the increase of LVCs in Old and Middle English.

many LVCs are native to English, and that we should be cautious when attributing too many to other languages.[6]

Internal linguistic factors include an increase in analyticity (Hiltunen, 1983; Kytö 1999) and the rise of the indefinite article in LVCs, e.g., *take a bath* (Brinton, 2008). Hiltunen (1983) suggests that the overall expansion might also be attributable to the loss of verbal prefixes and subsequent rise of periphrastic constructions as alternative structures. Alternatively, Brinton and Akimoto (1999) and Brinton (2008) suggest that the increased use of the indefinite pronoun allows for wider range of functions of CPs, including new expressions of limited duration (Wierzbicka, 1982). Thus, by the Late Modern English period, LVCs with indefinite NPs like *to make a joke* vs. *joke* become more frequent in a wider variety of contexts, expressing subtle aspectual nuances related to boundedness or completedness that were not possible during the Old English period before the rise of the indefinite article (Traugott, 1999, p. 246). Despite these possible explanations, however, Hiltunen (1999) points out that "a good deal of work still needs to be done" to understand the causes for the growth of LVCs (200).

3. Methodology

3.1 Corpus

Eight Old Swedish texts were selected for analysis, four texts from the period traditionally known as *äldre fornsvenska* 'Older Early Swedish' (here, OSw1), written prior to 1375, and four texts are from the *yngre fornsvenska* 'Younger Early Swedish' period (OSw2) from 1375–1526 (see the Appendix for a list of texts and word lengths and approximate dates).[7] In selecting these texts, I considered four factors in order to achieve some balance and representativeness of the corpus: genre, foreign influence, date of the manuscript/ composition of the text, and the total number of words available in the texts for each of the two periods under investigation. The eight texts include a mixture of genres, including laws, religious prose, and secular

---

6. Sundquist (2018) provides evidence that genre and foreign influence plays a significant role in the varying frequencies of LVCs in Old Swedish. Legal texts exhibit a higher overall frequency of LVCs but less diversity of NP-LV pairings, while religious prose texts with foreign influence from Latin or Low German exhibit a wider range of pairings but slightly lower frequency of LVCs.

7. Höder (2010) provides helpful discussion on periodization in Scandinavian with specific reference to early Swedish. I follow the traditional periodization in Pettersson (1996), abbreviating these two periods here with the English OSw1 and OSw2 for convenience. For more information on Old Swedish text selection, see Magnusson (2007).

prose. Since the bulk of OSw texts are translations from Latin or Low German, it was necessary to select a mix of native and foreign texts from each genre from each period. In terms of assigning texts to OSw1 or OSw2, I followed traditional dating in *Fornsvenska Textbanken* and considered previous studies that dated the texts on linguistic grounds, especially when the date of composition and the date of the manuscript were different or uncertain. In general, the grouping of texts into one of two periods is meant to provide only a relative dating of earlier vs. later texts for comparison rather than some kind of absolute dating. Lastly, I considered the total number of words, in order to make sure that the two periods had relatively similar- sized corpora (approximately 300,000 words in each).

3.2   Selection of LVCs

The issues regarding the classification of LVCs discussed in Section 2 were taken into consideration when selecting light verbs and their corresponding NP objects. I limited my analysis of OSw to five common transitive LVs in verbo-nominal expressions: *fa* 'get/receive', *giva* 'give', *göra* 'do/make', *hava* 'have', and *taka* 'take'. Moreover, I made these three assumptions in selecting relevant LVCs for analysis:

(6) a. LVCs contain a verb with low semantic specificity and an object complement, both of which are conceptualized as one action and form a single, meaningful phrasal unit

   b. LVCs contain an NP object complement (e.g., an action or abstract noun) with or without modification (e.g., indefinite or definite articles, quantifiers, adjectives, or post-nominal PPs)

   c. The LV is a semantically lighter version of a full, transitive verb that only contributes meaning to the phrase when it combines with an NP object complement

Following others like Claridge (2000) and Ronan (2014), I assume that the NP in an LVC is a deverbal noun or a noun that can be paraphrased by a single verb similar in meaning to the LV+NP combination at some point in the language's history. Moreover, I also follow other criteria for LVCs, like Quirk et al. (1985), Delsing (1998) for Swedish, or more recently Claridge (2000), who point out that not all pairings can be easily paraphrased by a single verb (e.g., *make an effort*). The main criterion for inclusion of certain LV-NP pairings is that the verb is semantically lighter than a full verb and, as Claridge (2000, p. 81) states, "does not override the importance of the noun".

3.3 Data collection procedure

To search for LVCs in the OSw texts, I used the concordance tool Korp (version 5.05) of *Spraakbanken* (the Swedish Language Bank). First, I selected the OSw collection and chose all twelve of the corpora in *Fornsvenska Textbanken*, limiting my analysis to the eight texts discussed in 3.1. I searched for all instances of each of the five light verbs (*fa*, *giva*, *göra*, *hava*, and *taka)* in these texts, looking for any sentences with abstract nouns that formed LVCs. Korp allows the user to search by "Similar Lemgram", a process whereby one can search for different inflected forms of the same lemma. However, because of the possibility that some lemgrams might not be detectable by Korp for various reasons, given the wide variety of grammatical and orthographic variation in OSw and the possibility of errors during the corpus transfer process, it was necessary to double-check all alternative spellings and inflected forms of each verb. After using these strategies to find relevant sentences with LVs, I manually went through the list of tokens to sort out any non-relevant instances.

For each instance in which one of the five verbs occurs with a nominal object that fits the criteria for selection, the sentence was entered into a database for further analysis. For example, upon searching for instances of the verb *giva* 'give', I included the following sentence with the NP *radh* 'advice':

(6) *Sidhan   gaff balaam radh   balac konunge*
    afterwards gave Balaam advice-ACC Balac king-DAT
    'Afterwards Balaam gave King Balac advice.'   (*Pent*)

However, an example with *giva* that was not included is in (7):

(7) þa   sculu bönder giuæ   byscupi   þrer   marcher
    then shall farmers give-INF bishop-DAT three marks-ACC
    'Then the farmers shall give the bishop three marks.'   (*ÄVgL*)

The NP object in (7) is a concrete noun which, when in position as one of the object complements to *giva*, does not fit the criteria of the LVCs outlined above. In other words, *giva* functions as a full verb in (7) and a light verb in (6), where it makes less semantic contribution to the CP when paired with the abstract noun *radh*.

3.4 Statistical analysis of LVC frequency and diversity

For both Old Swedish periods, I gathered data on the number of instances (tokens) in which each light verb occurs with NPs that meet the criteria in Section 3.2 along with the number of unique NPs (types) that occur with each verb. This allows

for calculation of the type-token ratio (TTR) as an indicator of the diversity of LV+NP pairings in each period (i.e., number of unique NPs divided by the total number of LVCs for that LV). In addition, in order to get an idea of the frequency of LVCs without influence from the total number of words in each period's corpus, I normalized the frequency with which each verb occurs in LVCs, calculating the number of tokens per 10,000 words.

Several empirical tests were conducted to measure the relative diversity of each verb in both periods. Biodiversity measurements are particularly useful in going beyond simple measurements of frequency to consider the relative distribution and internal structure of groups with wide degrees of variety. Using measurements from the study of ecology, I was able to create a statistical profile of each verb that took into account how frequency and diversity interact over time to shape the occurrence of each LV in LVCs. In much the same way that the richness or diversity of animal species in two different ecological communities can be measured, I examined each verb's richness, its diversity, and its evenness in both OSw period.[8]

Species richness is a way of measuring how many different species there are (here the number of unique NPs) relative to the sample size. This was calculated as Menhinick's index by dividing the number of unique NPs (types) by the square root of the number of LVCs with that verb (tokens). The higher the number of unique NPs, the greater the richness of the LV.

The most common measurements of diversity are the Shannon-Wiener index and Simpson's D index. In the current study, these indices reflect how diverse a verb is in terms of the number of unique LV-NP pairings. Shannon-Wiener is calculated according to the formula given in (8):[9]

$$(8) \quad H' = -\sum_{i=1}^{S} p_i \ln p_i$$

H' usually comes out as a figure between around 1.00 and 4.00 and is useful for measuring relative biodiversity of more than one ecological community. The higher the value for H', the greater the diversity of the community. In terms of the current study, the higher the number of unique LV-NP pairings in addition to a higher frequency of each those pairings, the higher the diversity according to this measurement.

---

8. For an overview of measurements of biodiversity, see Magurran (2013).

9. This equation can be described as follows: the proportion of species i relative to the total number of species (p_i) is calculated first, followed by the multiplication by the natural logarithm of this proportion ($\ln p_i$). The product is summed across species and multiplied by −1.

Simpson's D (or Simpson's Diversity Index) is another biodiversity index that can be used, especially if there are a few more dominant species (or here, a small number of highly-frequent LV-NP pairings). The formula for Simpson's D is given in (9):[10]

(9) $D = \sum_{i=1}^{S} \frac{n_i(n_i - 1)}{N(N-1)}$

Many biodiversity studies calculate Simpson's D and then subtract it from 1 in order to come up with a measurement that ranges between 0 and 1 that can be used for comparison, with values closer to 1.00 reflecting greater diversity. In terms of the current study, S is the number of unique NPs that may pair up with an LV, n_i is the number of occurrences of an NP, and N is the total of all NPs. The closer this value is to 1.00, the higher the diversity.

One last measurement of biodiversity that was used in this study, evenness, was calculated as Pielou's index (J'). J' refers to how close the number of organisms of a species is, or in our terms, how close the numbers are to each other for LV-NP pairings for any given LV. This index is calculated using the Shannon-Wiener index that is divided by the natural logarithm of the number of individual species (i.e., the number of unique NPs for an LV).

4.   Results

For OSw1 (1225–1375), the frequency data for each light verb are listed in Table 1:

Table 1. Frequency of LVCs according to each light verb, Types (unique NP complements), Tokens, Type-Token Ratio, and normalized frequency (per 10,000 words) in OSw1 (1225–1375)

Light verb	# Unique NPs	# LVCs	TTR	Frequency
giva 'give'	25	140	.179	4.67
göra 'do/make'	26	99	.263	3.30
hava 'have'	13	86	.151	2.87
fa 'get/receive'	15	52	.288	1.73
taka 'take'	8	44	.167	1.47
	87	421	.182	14.03

---

10. The equation can be described as follows: subtract 1 from the number of organisms of a species (n_i) and multiply by the number of organisms of that species; next subtract 1 from the total number of organisms of all species (N) and multiply by that same number. Divide and sum across all species.

The verbs are listed Table 1 in order of decreasing frequency (per 10,000 words). *Giva* is the most frequent verb in LVCs in this period in terms of the overall number of tokens, while *göra* has the highest number of unique NP complements with which it occurs. *Taka* is the most infrequent of the five verbs analyzed here and occurs with the fewest number of unique NPs. *Fa* exhibits the highest TTR at .288. The average TTR for all of OSw1 is .182 and the overall frequency of LVCs in this period is 14.21 per 10,000 words.

For the Late Old Swedish period (OSw2, 1375–1526), the frequency data are listed in Table 2:

Table 2. Frequency of LVCs according to each light verb, Types (unique NP complements), Tokens, Type-Token Ratio, and normalized frequency (per 10,000 words) in OSw2 (1375–1526)

Light verb	# Unique NPs	# LVCs	TTR	Frequency
göra 'do/make'	70	226	.310	6.88
hava 'have'	64	196	.327	5.97
giva 'give'	59	195	.303	5.94
fa 'get/receive'	44	137	.321	4.17
taka 'take'	20	47	.426	1.43
	257	801	.321	24.39

Although *göra*, *hava*, and *giva* have switched rankings in OSw2, all three are at the top of the list again, while *fa* and *taka* remain in fourth and fifth place, respectively. With the exception of *taka*, all of the verbs exhibit an increase in frequency between OSw1 and OSw2. The overall frequency of LVCs has increased from 14.21 to 24.39 LVCs per 10,000 words. *Göra* is the most frequent LV in terms of tokens and types in OSw2, while *taka* is the most infrequent. *Taka* exhibits a higher TTR than the other verbs in this period. Overall, all five verbs show an increase in TTR from OSw1 to OSw2, as confirmed by the overall rise in TTR for the whole period (.182 vs. .321).

Next, the biodiversity measurements were also calculated for each LV for each period, as shown in Table 3.

According to all four indices, each LV exhibits an increase in diversity from OSw1 to OSw2. For example, all five verbs, including *taka*, whose frequency dipped slightly, exhibit an increase in richness. The increases in evenness for all five suggest that at the same time that the number of unique NPs is growing, the frequencies of the individual NPs is becoming more similar. Both biodiversity indices (Shannon-Wiener and Simpson's D) exhibit varying degrees of change, but in all five cases, the values are increasing. When all four indices are considered together, along with the frequency and TTR figures, we see that LVs exhibit consistent growth and expansion over time.

Table 3. Richness (Menhinick's Index), Evenness (Pielou's Index), *Diversity* (Shannon-Wiener Diversity Index, Simpson's Diversity Index) for LVs in OSw1 (1225–1375) and OSw2 (1375–1526)

	Richness		Evenness		Shannon-Wiener index		Simpson's D	
	OSw1	OSw2	OSw1	OSw2	OSw1	OSw2	OSw1	OSw2
giva 'give'	2.058	4.225	.811	.893	2.578	3.644	.880	.932
göra 'do/make'	2.300	4.657	.788	.861	2.470	3.656	.885	.961
hava 'have'	1.617	4.511	.711	.858	1.924	3.553	.787	.959
fa 'get/receive'	1.769	3.673	.786	.903	1.952	3.395	.811	.966
taka 'take'	1.22	2.801	.431	.864	.896	2.544	.375	.912

These figures are particularly revealing when more lexically-specific (or "less light") verbs are considered. Consider the frequency data for two additional verbs used in collocations with NP object types in the Old Swedish corpus. *Bära* 'to bear' and *halda* 'to hold' occur in both OSw1 and OSw2, as in *bära plict* 'bear duty', *bära skam* 'bear shame', *halda wenskap* 'keep friendship' or *halda ordh* 'keep word/promise'. (See Tables 4 and 5).

Table 4. LVCs with the verb *bära 'bear'* in OSw1 and OSw2, Types (unique NPs), Tokens, Type-Token Ratio, and Normalized Frequency (per 10,000 words)

Period	#Unique NPs	# LVCs	TTR	Frequency
OSw1	8	45	.178	1.50
OSw2	6	26	.231	.79

Table 5. LVCs with the verb *halda 'hold'* in OSw1 and OSw2, Types (unique NPs), Tokens, Type-Token Ratio, and Normalized Frequency (per 10,000 words)

Period	#Unique NPs	# LVCs	TTR	Frequency
OSw1	9	52	.173	1.73
OSw2	7	47	.149	1.43

For both verbs, the numbers of LVCs and types of unique NPs decrease between OSw1 and OSw2, although the TTR for *bära* does increase. In terms of the diversity indices, both verbs exhibit a different pattern compared to that of five LVs:

Values for Shannon-Wiener and Simpson's D for *bära* and *halda* decrease from OSw1 to OSw2, with *bära* dropping more drastically than *halda*. Both diversity values for these two verbs are also lower overall than those of most of the LVs in either period. Moreover, the declining richness values reflect the diminished variety of NPs for both verbs in OSw1 vs. OSw2. In terms of evenness, *bära* decreases while

Table 6. Richness (Menhinick's Index), Evenness (Pielou's Index), *Diversity* (Shannon-Wiener Diversity Index, Simpson's Diversity Index) for *bära 'bear'* and *halda 'hold/keep'* in OSw1 (1225–1375) and OSw2 (1375–1526)

	Richness		Evenness		Shannon-Wiener index		Simpson's D	
	OSw1	OSw2	OSw1	OSw2	OSw1	OSw2	OSw1	OSw2
bära 'bear'	1.193	1.177	.699	.676	1.463	1.211	.799	.603
halda 'hold/ keep'	1.248	1.021	.781	.855	1.715	1.664	.814	.787

halda increases. In other words, the statistical profiles for these two verbs in LVCs indicate that they are becoming less frequent, less diverse over time than LVs, and their evenness values either rise or fall.

6.   Discussion

These findings provide additional support for the claims of Butt & Lahiri (2013) that LVs as a class behave differently over time when compared to other verbs. Moreover, they share several properties with each other as a small, closed class. Findings from the statistical analyses here indicate that LVs not only increase in frequency and diversity, but they also share properties in the way that they change over time as a group. As in the OSw examples, LVs are usually among the most frequent verbs, expand their combinatorial range, and grow in terms of diversity, richness, and evenness. Other lexically-specific verbs (e.g., *bära* or *halda*), on the other hand, are less frequent, sometimes stagnate, and at other times drop in frequency, or may decrease in diversity. Their behavior is more varied (i.e., they do not seem to fall together as a class that shares similar properties) and they exhibit overall less robust changes in frequency.[11]

    This phenomenon shares similarities with the notion of the "rich get richer" that is well documented in several scientific disciplines.[12] As Albert and Barabási

---

11.   Ideally, one could investigate this conclusion further with a wider range of lexically-specific verbs. The corpus in this study, however, is limited in terms of how many verbs could be used for analysis. Lexically-specific verbs like *halda* or *bära* provide enough examples for data analysis, since they co-occur with a number of abstract nouns in both periods and in all texts. An analysis of additional verbs would be made possible by expanding the corpus to include more texts or other periods.

12.   See Easley & Kleinberg (2010) for a helpful introduction to network science as it is applied in multiple disciplines.

(2002) note, there is a common pattern in dynamic network analyses known as preferential attachment whereby large nodes (or central "hubs") are linked to more frequently and at a faster rate over time than those that are infrequent. As Easley and Kleinberg (2010) point out, the proliferation of websites on the internet provides a concrete example: those websites that are already popular are linked to more often and grow faster than those that are originally less popular (479).

The analogue with the current study is that LVs co-occur with more and more unique NPs over time, and their "networks" of co-occurrence grow more diverse than those of other verbs precisely because they are already more frequent and diverse at the onset. In other words, LVs act more like central hubs with associations to a wide variety of NP objects. Over time, this network grows more diverse and more quickly than that of other verbs that tend to exhibit more unpredictable patterns of growth or stagnation.

The implications for these findings are that the growth of LVCs in a language's history is generally more predictable than a long-term decline over time. While the internal and external linguistic causes of LVC-expansion might help explain temporary upticks in frequency over shorter periods of time, LVCs continue along a natural trajectory of growth in the long-term. Thus, language contact might be an adequate explanation for some short periods of dramatic change, for instance. However, the prediction based on the results here is that these increases in frequency and diversity will take place even without external influence when the long-term development of LVCs is considered.

Interestingly, these results relate closely to findings in several first language (L1) studies on learnability and the acquisition of LVs. Building on Pinker's (1984) early research on L1 verb acquisition, Clark (1995, p. 30) notes that transitive LVs like *make* or *do* are acquired early and used frequently as a type of "all purpose" verb in an initial stage before children acquire more specific verbs. Ninio (1999) describes transitive LVs as "path-breaking" verbs: due to their flexibility in combining with a variety of NP objects, LVs function as prototypical verbs for children to learn general principles of transitivity (620). Mauoene et al. (2011) also note that typical LVs have a higher number of different NP-associations among adults than "heavy" lexically-specific verbs, and those verbs with the widest variety are often identified as those that are acquired earliest by children (129).

However, there is evidence that it is their high frequency rather than any special properties of LVs that might play the most pivotal role in their acquisition. Theakston et al. (2004) point out that once frequency effects of LVs are removed in statistical analyses, any kind of privileged properties of LVs, like their semantic generality, prove to be insignificant predictors for the early acquisition and overuse of LVs in child data (90). In a similar way, and in close relation to the current study, Hills et al. (2009) employ network analysis tools to describe early acquisition of

nominal objects. Their findings indicate that the L1 acquisition of nouns follows the principles of preferential attachment, or more specifically, what they call "preferential acquisition": nouns are more easily acquired that are already more connected to other highly-connected words in the adult input, like LVs (729). In other words, by nature of their status as frequent and diverse verbs with a wide variety of NP co-occurrences in adult speech, the network of nouns that can be associated with LVs grows more rapidly.

These aspects of learnability and L1 acquisition shed light on the unique diachronic properties of LVs. Recall that Butt & Lahiri (2013) described these verbs as *passepartouts* that fulfill a number of lexical or grammatical functions (24). It is by nature of their lexical flexibility and generic meaning that they are frequently used in a wide variety of contexts and combine with many different objects. Empirical results of the study here, as well as some of the studies on L1 learnability, indicate that it is this frequency and diversity that fuel the growth of the verbs in the first place and allow them to maintain a privileged status as a verb that functions in many lexico-grammatical contexts.

6. Summary and conclusion

Results of this study of OSw LVs indicate that LVCs with the most common transitive light verbs increase in both frequency and diversity between 1225 and 1526. Empirical tests of frequency and type-token ratio were conducted, and several biodiversity index values were calculated. In comparison to more lexically-specific verbs, the five LVs in OSw behaved in a uniform way as a small class, increasing in richness, evenness, and diversity, not to mention in their overall frequency and range of NP-object pairings. By comparison, two lexically-specific verbs decreased in frequency and diversity, and they were more varied in their statistical patterns of development.

These findings were discussed in terms of dynamic networks. In particular, the notion in network science that the "rich get richer" was discussed in the context of light verbs. As highly frequent verbs that co-occur with a wider variety of NP object complements than other verbs, LVs have a privileged status, and as a result, they are more apt to continue naturally on a path towards greater diversity and frequency. Internal and external linguistic mechanisms of change might explain some periodic rises or surges, but the implications of this study are that LVs will continue to expand and grow as they do by nature of their unique status.

Future research may explore additional aspects of LVs to evaluate the possibility of other network properties common in cases of preferential attachment. For instance, as Easley & Kleinberg (2010) point out, such networks are scale-free

and exhibit power law distributions with a long tail with a high number of rare data-points (486). Expanding the corpus to include more texts and examining a wider array of verbs for comparison would allow one to evaluate the predictions here and explore the distribution of types and tokens beyond the current small dataset. A broader and deeper investigation might allow one to examine the more infrequent and rare LV-NP pairings to describe more adequately the historical aspects of this unique set of verbs as they change over time.

Abbreviations

ACC	accusative	NP	noun phrase
DAT	dative	OSw	Old Swedish
INF	infinitive	OSw1	Older Early Swedish
L1	first language	OSw2	Younger Early Swedish
LV	light verbs	PP	past participle
LVC	light verb constructions	TTR	type-token ratio

References

Algeo, J. (1995). Having a look at the expanded predicate. In B. Aarts, & C. F. Meyer (Eds.), *The verb in contemporary English: Theory and description* (pp. 203–17). Cambridge: Cambridge University Press.

Allerton, D. J. (2002). *Stretched verb constructions in English*. London: Taylor & Francis.

Albert, R. & Barabasi, A. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74, 47–97. https://doi.org/10.1103/RevModPhys.74.47

Borin, L., M. Forsberg & J. Roxendal. (2012). Korp – the corpus infrastructure of Språkbanken. *Proceedings of LREC 2012*, 474–478. Istanbul: ELRA.

Bowern, C. (2008). The diachrony of complex predicates. *Diachronica*, 25, 161–185. https://doi.org/10.1075/dia.25.2.03bow

Brinton, L. J. & M. Akimoto (Eds.) (1999). *Collocational and idiomatic aspects of composite predicates in the history of English*. Amsterdam: John Benjamins Publishing. https://doi.org/10.1075/slcs.47

Brinton, L. J. (2008). Where grammar and lexis meet. In E. Seoane, & M. J. López-Couso (Eds.), *Theoretical and empirical issues in grammaticalization* (pp. 33–53). Amsterdam: John Benjamins Publishing. https://doi.org/10.1075/tsl.77.04bri

Butt, M. (2010). The light verb jungle: Still hacking away. In A. Mengistu, B. Baker, & Mark Harvey (Eds.), *Complex predicates in cross-linguistic perspective* (pp. 48–78). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511712234.004

Butt, M. & Lahiri, A. (2013). Diachronic pertinacity of light verbs. *Lingua*, 135, 7–29. https://doi.org/10.1016/j.lingua.2012.11.006

Cinková, S. (2009a). A contrastive lexical description of basic verbs. *The Prague Bulletin of Mathematical Linguistics*, 92: 21–62.

Cinková, S. (2009b). Words that matter: towards a Swedish-Czech colligational dictionary of basic verbs. Ph.D. dissertation. University of Prague.

Claridge, C. (2000). *Multi-word verbs in early Modern English: A corpus-based study*. Amsterdam: Rodopi. https://doi.org/10.1163/9789004333840

Clark, E. V. (1995). *The lexicon in acquisition*. Cambridge: Cambridge University Press.

Delsing, L.-O. (1998). Support verbs and the argument rule. In K. van Durme & L. Schløser (Eds.), *Studies in Valency IV* (pp. 63–88). Odense: Odense University Press.

Dura, E. (1997). *Substantiv och stödverb*. Göteborg: Institutionen för svenska språket.

Easley, D. & Kleinberg, J. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge: Cambridge University Press.
https://doi.org/10.1017/CBO9780511761942

Ekberg, L. (1993). Verbet ta i metaforisk och grammatikaliserad användning. *Språk och stil*, 3, 105–139.

Elenbaas, M. (2013). The synchronic and diachronic status of English light verbs. *Linguistic Variation*, 13, 48–80. https://doi.org/10.1075/lv.13.1.02ele

Giry-Schneider, J. (1987). *Les prédicats nominaux en français: les phrases simples à verbe support*. Geneva: Librairie Droz.

Hills, T. T., Maouene, M., Maouene, J., Sheya, A. & Smith, L. (2009). Longitudinal analysis of early semantic networks: preferential attachment or preferential acquisition? *Psychological science*, 20(6), 729–739. https://doi.org/10.1111/j.1467-9280.2009.02365.x

Hiltunen, R. (1983). *The decline of the prefixes and the beginnings of the English phrasal verb: The evidence from some Old and Early Middle English texts*. Vol. 160 Turku: Turun yliopisto.

Hiltunen, R. (1999). Verbal phrases and phrasal verbs. In L. J. Brinton, & M. Akimoto (Eds.), *Collocational and idiomatic aspects of composite predicates in the history of English* (pp. 167–206). Amsterdam: John Benjamins Publishing. https://doi.org/10.1075/slcs.47.45hil

Höder, S. (2010). *Sprachausbau im Sprachkontakt: Syntaktischer Wandel im Altschwedischen*. Heidelberg: Universitätsverlag Winter.

Iglesias-Rábade, L. (2001). Composite predicates in Middle English with the verbs nimen and taken. *Studia Neophilologica*, 73, 143–163. https://doi.org/10.1080/003932701753401456

Jespersen, O. (1942). *A modern English grammar on historical principles VI*. Copenhagen: Allen and Unwin.

Kytö, M. (1999). Collocational and idiomatic aspects of verbs in Early Modern English. In L. J. Brinton & M. Akimoto (Eds.), *Collocational and idiomatic aspects of composite predicates in the history of English* (pp. 167–206). Amsterdam: John Benjamins Publishing.
https://doi.org/10.1075/slcs.47.53kyt

Magnusson, E. (2007). *Gränsöverskridande koordination. Syntaktisk förändring i äldre svenska*. Göteborg: Acta Universitatis Gothoburgensis.

Magurran, A. E. (2013). *Measuring biological diversity*. John Wiley & Sons.

Matsumoto, M. (1999). Composite Predicates in Middle English. In L. J. Brinton & M. Akimoto (Eds.), *Collocational and idiomatic aspects of composite predicates in the history of English* (pp. 59–96). Amsterdam: John Benjamins Publishing. https://doi.org/10.1075/slcs.47.24mat

Maouene, J., Laakso, A. & Smith, L. B. (2011). Object associations of early-learned light and heavy English verbs. *First Language*, 31, 109–132. https://doi.org/10.1177/0142723710380528

Nickel, G. (1978). Complex verbal structures in English. *International Review of Applied Linguistics*, 6, 1–21. https://doi.org/10.1515/iral.1968.6.1-4.1

Ninio, A. (1999). Pathbreaking verbs in syntactic development and the question of prototypical transitivity. *Journal of child language*, 26, 619–653.
https://doi.org/10.1017/S0305000999003931

Pettersson, G. (1996). *Svenska språket under sjuhundra år: en historia om svenskan och dess utforskande*. Lund: Studentlitteratur.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Polenz, P. von. (1987). Funktionsverben, Funktionsverbgefüge und Verwandtes. Vorschläge zur satzsemantischen Lexikographie. *Zeitschrift für germanistische Linguistik*, 15, 169–189. https://doi.org/10.1515/zfgl.1987.15.2.169

Quirk, R., S. Greenbaum, G. Leech & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.

Ronan, P. (2014). Light verb constructions in the history of English. In K. Davidse, C. Gentens, L. Ghesquière & L. Vandelanotte (Eds.), *Corpus interrogation and grammatical patterns* (pp. 15–34). Amsterdam: John Benjamins Publishing.

Sundquist, J. D. (2018). A diachronic analysis of light verb constructions in Old Swedish. *Journal of Germanic Linguistics*, 30(3), 260–306.

Theakston, A. L., Lieven, E. V., Pine, J. M. & Rowland, C. F. (2004). Semantic generality, input frequency and the acquisition of syntax. *Journal of child language*, 31, 61–99. https://doi.org/10.1017/S0305000903005956

Traugott, E. C. (1999). A historical overview of complex predicate types. In L. J. Brinton and M. Akimoto (Eds.), *Collocational and idiomatic aspects of composite predicates in the history of English* (pp. 239–260). Amsterdam: John Benjamins Publishing. https://doi.org/10.1075/slcs.47.74clo

Viberg, Å. (1990). Svenskans lexikala profil. *Svenskans beskrivning*, 17, 10–1.

Wierzbicka, A. (1982). Why can you have a drink when you can't *have an eat? *Language*, 58, 753–799. https://doi.org/10.2307/413956

Appendix. Old Swedish sources

Texts are listed alphabetically with total number of words analyzed along with approximate dates of composition and manuscripts in square brackets.

Early Old Swedish (ca. 1225–1375)

Leg   *Fornsvenska legendariet (*according to *Codex Bildstenianus)*. In *Svenska fornskriftsällskapets skrifter* 7:1–2, ed. by G. Stephens, 1847. Stockholm: Norstedt. [124,907 words; ca. 1350]

Pent   *Fem Moseböcker på fornsvenska*: *Pentateuchparafrasen efter MB I B*. In *Svenska fornskriftsällskapets skrifter* 60, ed. by O. Thorell, 1959. Stockholm: Norstedt. [145,150 words; ca. 1330, MS ca. 1526]

UL   *Upplandslagen* (according to Holm B 199). In *Svenska Fornskriftsällskapets Samlingar* 70, ed. by Sam Henning, 1967. Stockholm: Kungliga biblioteket. [46,620 words; 1297, MS ca. 1300s]

ÄVgL   *Äldre Västgötalagen*. In *Samling af Sweriges gamla lagar* 2, ed. by Hans Samuel Collin and Carl Johan Schlyter, 1827. Stockholm: Norstedt. [15,010 words; ca. 1220s, MS ca. 1280]

Later Old Swedish (1375–1526)

Birg *Birgittas uppenbarelser I: Bok 1–3* (according to Holm A 33). In *Svenska fornskriftsällskapets skrifter* 14:1, ed. by Gustav Edvard Klemming, 1857. Stockholm: Norstedt. [106,740 words; ca. 1380, MS ca. Mid-1400s)

BJ *Barlaam och Josaphat*. In *Svenska fornskriftsällskapets skrifter* 28, ed. by Gustav Edvard Klemming, 1887. Stockholm: Norstedt. [27,029 words; 1440s]

KL *Konung Kristoffers Landslag. Samling af Sweriges gamla lagar* 12, ed. by Carl Johan Schlyter, 1869. Stockholm: Z. Haeggström. [54,936 words; 1442]

Tröst *Själens tröst*. In *Svenska fornskriftsällskapets skrifter* 59, ed. by Samuel Henning, 1954. Stockholm: Norstedt. [145,473 words; ca. 1420]

Expletives in Icelandic

A corpus study

Hannah Booth
University of Konstanz

Various claims have been made concerning expletives in historical Icelandic (Eythórsson & Sigurðardóttir, 2016; Hróarsdóttir, 1998; Rögnvaldsson, 2002). However, previous studies are limited to certain centuries and/or text types. This paper assesses the status of expletives throughout all stages of attested Icelandic (1150–2008) using the Icelandic Parsed Historical Corpus (IcePaHC). The IcePaHC data indicates that the expletive is already attested in impersonal constructions in Old Icelandic, contra previous accounts. I also claim that *þar* 'there' was historically available as an expletive in presentational constructions and argue against a recent claim that the emergence of expletive *það* is a contact phenomenon. Strikingly, postfinite expletives are attested historically, which challenges the standard cross-Germanic account for the diachrony of expletives.

Keywords: syntactic change, historical corpora, Germanic, Icelandic, expletives, cataphora

1. Introduction

The rise of expletives is a diachronic development common to the Germanic languages.[1] All present-day Germanic languages exhibit expletives (see Vikner, 1995 for an outline). However, in earlier stages of Germanic expletives are typically absent in contexts where they occur today. The emergence of expletives has been well researched for many of these languages: e.g. Breivik (1983) and Ingham (2001) on English; Lenerz (1985) and Axel (2007) on German; Falk (1993) on Swedish; Faarlund (1990) and Kinn (2016) on Norwegian. By comparison, the Icelandic diachrony has received less attention, and previous studies have been limited to certain centuries, text genres and/or construction types (Hróarsdóttir, 1998;

---

1. "Expletive" in this sense is interchangeable with "pleonastic" or "dummy", i.e. an element which has no semantic content, but is there for syntactic reasons.

Rögnvaldsson, 2002; Eythórsson & Sigurðardóttir, 2016). Present-day Icelandic has one main form which functions as an expletive: það, formally identical to the 3SG. NT referential pronoun.[2] The development whereby expletive það emerged in the history of Icelandic is the focus of this paper. I present a quantitative corpus-based study, using the Icelandic Parsed Historical Corpus (IcePaHC, Wallenberg et al., 2011). This allows me to assess the status of það in a range of contexts throughout all attested stages of Icelandic (1150–2008), thereby capturing a diachronic breadth and detail not achieved in the literature to date.

The structure of the paper is as follows: in Section 2, I outline the background concerning expletives in Icelandic and historical Germanic; Section 3 presents the methodology and findings of the corpus study; Section 4 discusses these findings; Section 5 concludes the paper.

2.   Background

2.1   Three types of það

For the purposes of this study, I differentiate between three types of það in Icelandic:

1.   Anaphoric það (ANAPH), e.g. (1).
2.   Cataphoric það (CATPH), e.g. (2).
3.   Expletive það (EXPL), e.g. (3).

(1)   [Þetta   andlit]$_i$, ég   hafði   einhvern tíma   séð
      DEM.ACC face.ACC   I.NOM have.PST one.ACC time.ACC see.PST.PTCP
      það$_i$ áður…
      ANAPH before
      'This face, I had seen it one time before…'   (2008, Ofsi.258)[3]

(2)   a.   **Það**$_i$ var   samt ljóst [að   við   þyrftum
           CATPH be.PST still clear COMP we.NOM need.PST.SBJV
           meiri mannskap]$_i$.
           more crew
           'It was still clear that we would need more crew.'   (2008, Ofsi.733)

---

2.   There is also a second form which can function as an expletive in present-day Icelandic: *hann*, formally identical to the 3SG.MASC pronoun, which typically appears in constructions with weather predicates (Eythórsson & Sigurðardóttir, 2016).

3.   Unless otherwise stated, all examples are from IcePaHC and referenced: (Year, Text. CorpusID). Further information about the texts and editions used can be found in the IcePaHC documentation: https://github.com/antonkarl/icecorpus [last accessed 26.09.2018].

b. **Það**$_i$ má segja [að mér hafi lukkast að
   CATPH may say.INF COMP I.DAT have.PRS.SBJV succeed.PST.PTCP to
   fleygja sál minni…].
   throw-away.INF soul my
   'One may say that I have succeeded in throwing away my soul…'

   (1985, Margsaga.945)

(3) a. **Það** var að vora.
   EXPL be.PST to become-spring.INF
   'It was becoming spring.'   (2008, Mamma.1066)

b. **Það** var töluverður snjór yfir öllu.
   EXPL be.PST considerable.NOM snow.NOM over all
   'There was a considerable amount of snow over everything.'

   (2008, Ofsi, 772)

This paper will discuss cataphoric *það* and expletive *það*. Cataphoric *það* occurs in constructions with a clausal argument ("extraposition", e.g. Thráinsson, 1979), and cataphorically refers to this argument.[4] By contrast, expletive *það* is neither anaphorically nor cataphorically referential.

For expletive *það*, I distinguish between impersonal constructions which lack an overt subject, e.g. (3a), and presentational constructions which have a postfinite discourse-new subject, e.g. (3b).[5] The impersonal category includes certain predicates which some claim license "quasi-arguments" (non-referential arguments), distinct from "true expletives" (non-referential non-arguments), see Falk (1993, p. 67); Chomsky (1981, pp. 324–327). For authors who recognise quasi-arguments, *það* with the weather predicate in (3a) would qualify as such (see Vikner, 1995, pp. 224–227). However, as I claim in Section 2.2, expletive *það* does not qualify as a subject but is a structural placeholder for the prefinite position. As a non-subject, *það* cannot qualify as a quasi-argument.[6] As such, I classify constructions with weather predicates together with other impersonal constructions where *það* is a positionally restricted expletive, such as impersonal passives, e.g. (4).

(4) **Það** var dansað alla nóttina.
   EXPL be.PST dance.PASS.PTCP all night
   'It was danced all night.'   (Thráinsson, 2007, p. 266)

---

4. The classic account by Thráinsson (1979, p. 181) claims that, in contexts like (2), prefinite *það* can be referential or expletive, while postfinite *það* is always referential. For this paper, I assume that any *það* which co-occurs with a clausal argument later in the sentence is cataphoric.

5. For evidence that the postfinite noun phrase in Icelandic presentational constructions qualifies as a syntactic subject, see Zaenen et al. (2017).

6. For a similar view of expletive *það*, see Eythórsson & Sigurðardóttir (2016).

The presentational category captures both "existentials" with copular *vera* 'be', e.g. (3b), as well as presentational constructions with other verbs, e.g. unaccusative *koma* 'come' in (5).

(5) **Það** kæmi   bara tóm   þvæla.
     EXPL come.PST.SBJV only empty.NOM nonsense.NOM
     'There would come only empty nonsense.'   (2008, Mamma.754)

For an overview of subtypes of impersonal and presentational constructions in modern Icelandic, see Thráinsson (2007, pp. 310–311).[7] A particularly fine-grained classification scheme is presented in Booth (forthcoming), which takes into account both present-day and historical Icelandic data.

2.2 Position of expletive *það*

The positional distribution of expletive *það* has attracted much attention in the literature (e.g. Sells, 2005; Sigurðsson, 2007). It is well known for present-day Icelandic that expletive *það* is restricted to the clause-initial prefinite position. Compare the impersonal passive in (6a) which has a clause-initial *það*, with (6b), where the clause-initial position is occupied by an adjunct and *það* is ungrammatical.

(6) a. **Það** var   dansað   í gær.
         EXPL be.PST dance.PASS.PTCP yesterday
         'It was danced yesterday.'
     b. Í gær   var   (*það*) dansað
         yesterday be.PST (*EXPL) danced.PASS.PTCP
         'Yesterday it was danced.'

On the basis of the data in (6), I assume that *það* does not qualify as a subject, since Icelandic subjects typically invert in fronting contexts like (6b) (see also Thráinsson, 1979, pp. 480–481; Platzack, 1983; Maling, 1988). Compare (6) with the Swedish data in (7), in which expletive *det* behaves like a subject.

(7) a. **Det** dansades   i går.
         EXPL dance.PST.PASS yesterday
         'It was danced yesterday.'
     b. I går   dansades   **det**.
         yesterday dance.PST.PASS EXPL
         'Yesterday it was danced.'

---

7. *Það* also occurs in cleft constructions, but I do not treat clefts in this paper.

In this paper, I refer to the positionally restricted Icelandic type as a "prefinite expletive", and the Swedish type as a "subject expletive".[8]

2.3 Historical context

In earlier Icelandic, expletive *það* is typically absent in the impersonal and presentations constructions outlined in Section 2.1. The absence of a clause-initial *það* results in verb-initial (V1) structures, e.g. (8); cf. (3) above.[9]

(8) a. Tekur   nú að hausta.
       begin.PRS now to become-spring.INF
       'It now starts to become autumn.'   (1310, Grettir.48)[10]

  b. Var þá gleði   mikil   í kóngs   höll.
       be.PST then joy.NOM great.NOM in king.GEN hall
       'There was then great joy in the king's hall.'   (1480, Jarlmann.48)

Rögnvaldsson (2002) observed that the first instances of expletive *það* in presentational and impersonal constructions appear in c.1500. Hróarsdóttir (1998) examines 18th and 19th century informal letters and observes an increase in the frequency of *það* in authors born after 1810. However, these two studies are restricted to certain centuries, and therefore offer only a snapshot of the diachrony. Moreover, both Rögnvaldsson and Hróarsdóttir only count the number of sentences which feature expletive *það*, without taking into account the total contexts in which the expletive could potentially occur, including where it is absent, e.g. (8). As I discuss in Section 3.1, IcePaHC allows me to take this into account. Concerning cataphoric *það*, e.g. (2), Faarlund (1990, p. 72) and Rögnvaldsson (2002) state that cataphoric *það* is already frequently found in Old Norse/Icelandic (1150–1350), a claim which I test in the corpus study in Section 3.

---

8. Faarlund (1990, pp. 63–65) refers to the positionally restricted type as an "expletive topic". I prefer "prefinite expletive", since "expletive topic" implies information structural properties which are not relevant in describing a structural filler; being non-referential, expletives cannot be topics.

9. An anonymous reviewer suggests that the examples in (8) involve so-called "Narrative Inversion". However, as I understand it, Narrative Inversion refers to V1 declaratives which have a topical subject in the immediately postfinite position (see e.g. Sigurðsson, 1990). Since the examples in (8) are topicless, neither of them qualifies as Narrative Inversion.

10. All IcePaHC texts are normalised to modern orthography, regardless of their date. I follow this normalisation throughout.

As noted in Section 1, the development of expletives is common to Germanic. Based on studies of Germanic languages other than Icelandic, the pathway of change in (9) has been proposed, which I refer to as the "Prefinite First Hypothesis".[11]

(9)   **Prefinite First Hypothesis**
      prefinite expletive > subject expletive

The Prefinite First Hypothesis states that positionally restricted prefinite expletives emerge before subject expletives historically. In this connection, it has been claimed that the initial rise of Germanic expletives is not conditioned by subject-related considerations, but by structural considerations concerning the verb-second (V2) constraint (Richards & Biberauer, 2005). In this paper, I test whether the Prefinite First Hypothesis holds for Icelandic.

3.   Corpus study

3.1   Methodology

The basis for this study is data from the Icelandic Parsed Historical Corpus (Ice-PaHC, Wallenberg et al., 2011). IcePaHC contains approximately 1,000,000 words spanning ten centuries (1150–2008), thereby covering all attested stages of Icelandic.[12] IcePaHC thus allows one to examine change across the centuries which many pre-existing studies contrasting Old Icelandic with present-day data do not capture.[13] IcePaHC follows the Penn treebank format established for historical English (e.g. Kroch & Taylor, 2000; Santorini, 2010) and is compatible with the CorpusSearch query language (Randall, 2005). All content is lemmatised, part-of-speech tagged and annotated for constituency.

IcePaHC treats both cataphoric *það* and expletive *það* as an expletive ("ES"). Moreover, both overt and "null" expletives – identical constructions where *það* is

---

11. See Breivik (1983), Ingham (2001) on English; Lenerz (1985) on German; Falk (1993) on Swedish; Faarlund (1990) on Norwegian; more general overviews are Haiman (1974), Richards & Biberauer (2005).

12. As an anonymous reviewer points out, certain texts dated to 1150 are extant in manuscripts from the 14th century; their dating is an approximation.

13. IcePaHC has some limitations: the texts included represent a very small sample of attested historical Icelandic, and certain genres are over-represented and others under-represented in individual periods. Some texts are also translations of source texts in other languages. Nevertheless, the advantages offered by the syntactic annotation outweigh such issues.

absent, e.g. (8) – are distinctively tagged.[14] I take the tagging at face value, in that I do not look for additional sentences which qualify as cataphoric/expletive *það* contexts but are not annotated as such. I restrict the study to matrix declaratives.

I isolated the cataphoric/expletive *það* contexts outlined in Section 2.1 via CorpusSearch queries. I then manually examined all examples, checking the construction type assignment and confirming internal category coherence. I excluded 150 examples which are misannotated as cataphoric/expletive *það* contexts. The data extracted by this process is shown in Table 1. I split the IcePaHC diachrony into five periods, as noted in the first column.

Table 1. Cataphoric/expletive *það* contexts in IcePaHC, 1150–2008

Period	Cataphoric contexts	Expletive contexts		All
		Impersonal	Presentational	
1150–1350	366	365	90	821
1351–1550	161	237	88	486
1551–1750	172	151	93	416
1751–1900	158	204	100	462
1901–2008	197	225	135	557
All periods	**1054**	**1182**	**506**	**2742**

All examples in Table 1 were coded for the presence/absence of *það* via the search queries. Additionally, I tagged manually for further relevant properties: predicate type; verb position; position of *það* if present.

In this section, I compare the proportion of instances in which *það* is present in the clause-initial prefinite position resulting in V2, e.g. (2)–(3), to instances in which *það* is absent, resulting in V1, e.g. (8). In comparing the frequency of *það* across the five periods, I conduct χ^2-tests to examine whether the value for an individual period deviates significantly from the mean value across the whole corpus ($p < 0.05$*, $p < 0.01$**, $p < 0.001$***; "ns" marks a value which does not significantly deviate from the mean).

3.2 Cataphoric *það*

The IcePaHC findings confirm previous claims (Faarlund, 1990; Rögnvaldsson, 2002) that prefinite cataphoric *það* is already robustly attested in constructions with a clausal argument in Old Icelandic (1150–1350), at a frequency of 77.0%, see Table 2.

---

14. For the IcePaHC policy on expletives, see: http://www.linguist.is/icelandic_treebank/ Expletives [last accessed 26.09.2018].

Table 2. Frequency of prefinite cataphoric *það* in constructions
with a clausal argument in IcePaHC, 1150–2008

Period	Prefinite *það* (V2)	No *það* (V1)	Total	% *það*	χ^2
1150–1350	124	37	161	77.0%	ns
1351–1550	49	22	71	69.0%	*
1551–1750	52	24	76	68.4%	*
1751–1900	75	18	93	80.6%	ns
1901–2008	109	10	119	91.6%	***
All periods	**409**	**111**	**520**	**78.7%**	

Cataphoric *það* is frequent through all periods and there is an increase visible in
the data for 1901–2008, where *það* occurs at a frequency of 91.6%. The lowest fre-
quencies are observed for the periods 1351–1550 and 1551–1750, coinciding with
a genre bias: religious and biographical texts are disproportionately represented in
these two periods, compared to the other periods where narrative texts (i.e. sagas/
modern fiction) dominate. This may account for the lower frequencies, but more
research on texts beyond IcePaHC would be required to confirm this.

Examples (10)–(11) illustrate the use of cataphoric *það* from the earliest period
(1150–1350).

(10) **Það**$_i$ er ráð húsfreyja [að taka vel við gestum]$_i$.
     CATPH be.PRS advice housewife to receive.INF well PTCL guests
     'It is good advice, housewife, to receive guests well.'   (1310, Grettir.658)

(11) **Það**$_i$ vildi ég [að þú ræddir ekki um]$_i$.
     CATPH wish.PST I.NOM COMP you.NOM talk.PST NEG about
     'I wished that you didn't talk about it.'   (1275, Morkin.1280)

A particular context in which cataphoric *það* occurs is in constructions like (12),
where the main verb in the matrix clause is in present participle form (see Faarlund,
2004, pp. 133–134 for discussion).

(12) En **það**$_i$ er vitanda [að þá göfgum vér réttlega
     but CATPH be.PRS know.PRS.PTCP COMP then honour.PRS we rightly
     postula guðs alla]$_i$…
     apostle god.GEN all
     'But one is to know that we then rightly honour all of God's apostles…'
                                                          (1150, Homiliubok.304)

The majority of constructions like (12) in my dataset are attested in *Homiliubok*.
This text is known to have a Latin background, though the exact nature of this
connection remains unclear (for discussion, see Weenen, 1993). Weenen (1993,

p. 183) discusses constructions like (12) under "Latin constructions". She gives the Latin in (13a) as the parallel construction for (13b).[15]

(13) a. Sed quærendum   nobis   est…
          but seek.PASS.PTCP we.DAT be.PRS
          'But we are to seek…'

      b. En þess$_i$ er   oss   leitanda   [að…]$_i$
          but CATPH.GEN be.PRS we.DAT seek.PASS.PTCP COMP
          'But we are to seek that…'[16]

Strikingly, although there is no cataphoric pronoun in the Latin original (13a), cataphoric það (þess) is present in the Icelandic translation in (13b).

The *Ordbog over det norrøne prosasprog* ('Dictionary of Old Norse Prose') lists similar examples from *Homiliubok* with the Latin parallel, e.g. (14).[17]

(14) a. Sciendum   est…
          know.PASS.PTCP be.PRS
          'It is to be known…'

      b. En það$_i$ er   vitanda   [að…]$_i$
          but CATPH be.PRS know.PRS.PTCP COMP
          'But it is to be known that…'[18]

Data like (13)–(14) indicate that the use of cataphoric það in such contexts was a robust feature of Old Icelandic, and was employed even in sentences based on a source construction with no such cataphoric pronoun.

3.3   Expletive *það* in impersonal constructions

The literature states that expletive það does not occur in impersonal constructions prior to c.1500, at which point það is attested in impersonal passives, and with weather verbs in c.1540 (Rögnvaldsson, 2002). However, the IcePaHC data challenges this claim (see Table 3).

---

15. Thanks to Tarrin Wills (p.c) for pointing this out.

16. The verb *leita* 'seek' takes a genitive argument as its object.

17. *Ordbog over det norrøne prosasprog.* [Dictionary of Old Norse Prose.] (n.d.) University of Copenhagen, https://onp.ku.dk [last accessed 26.09.2018].

18. http://onpweb.nfi.sc.ku.dk/webart/v/vi/88016731326cartpnfr.htm [last accessed 26.09.2018].

Table 3. Frequency of prefinite expletive *það* in impersonal constructions in IcePaHC, 1150–2008

Period	Prefinite *það* (V2)	No *það* (V1)	Total	% *það*	χ^2
1150–1350	6	157	163	3.7%	***
1351–1550	4	110	114	3.5%	***
1551–1750	2	64	66	3.0%	**
1751–1900	17	72	89	19.1%	ns
1901–2008	68	18	86	79.1%	***
All periods	**97**	**190**	**518**	**18.7%**	

While expletive *það* is relatively infrequent in the early data, there are six examples with the expletive from different texts in Old Icelandic (1150–1350), contra the literature. Examples are given in (15)–(17).

(15) **Það** er   nú að segja frá Hrafni. Hann kom   á Eyri og…
      EXPL be.PRS now to say.INF from Hrafn he.NOM come.PST to Eyrir and
      'Now one speaks of Hrafn. He came to Eyrir and…'   (1350, Finnbogi.1394)[19]

(16) **Það** var   snemma um   morgun.
      EXPL be.PST early   about morning
      'It was early in the morning.'   (1350, BandamennM.1196)

(17) **Það** er   mælt   um   sakir þær allar sem hér eru
      EXPL be.PRS speak.PASS.PTCP about things DEM all   REL here be.PRS
      taldar,   um   frumhlaup   og um sár   og um
      tell.PASS.PTCP about personal-assault and about injury and about
      víg   og lagalöstu   alla…
      manslaughter and law-evasions all
      'One speaks about all those things which are told here, about personal assault
      and about injury and about manslaughter and all evasions of the law…'
                                                         (1270, Gragas.334)[20]

Contra previous claims, the examples in (15)–(17) indicate that expletive *það* could occur in impersonal constructions in Old Icelandic, although the overall preference at this stage was for there to be no expletive.

---

19. The example initiates a new discourse and so it is reasonable to rule out the possibility that *það* is anaphoric.

20. This sentence introduces a new chapter, and so one can rule out the possibility that the clause-initial *það* has anaphoric reference to something in the preceding discourse.

Examples with expletive *það* in the period 1351–1550 are shown in (18)–(20).

(18) **Það** er nú sagt af einum ríkum manni og mikilhæfum…
EXPL be.PRS now say.PASS.PTCP of one rich man and talented
Hann tók sótt hættliga.
he.NOM take.PST sickness dangerous
'One now speaks of a rich and talented man… He caught a dangerous illness.'
(1475, Ævintyri.477)[21]

(19) **Það** skall ekki svo ganga ef ég kann hitta hana.
EXPL shall NEG so go.INF if I.NOM can meet.INF she.ACC
'Things will not go that way if I can meet her.' (1475, Ævintyri.655)

(20) **Það** er þvottdagur.
EXPL be.PRS Saturday
'It is Saturday.' (1540, ntjohn.193.331)

Strikingly, the early examples with the expletive are restricted to 'say'-type predicates, e.g. (15), (17)–(18), and inherently subjectless predicates (i.e. those which do not select a subject argument), e.g. (16), (19)–(20).[22] Only in the late 19th/early 20th century does expletive *það* appear in impersonal constructions with other types of predicate. (See Examples (21)–(23).) This generalisation to all predicate types accounts for the significant increase in expletive *það* in the data for 1901–2008 shown on Table 3.

(21) **Það** þurfti ekki að vitja um Hans…
EXPL need.PST NEG to check.INF about Hans
'It is not necessary to check up on Hans…' (1883, Voggur.81)

(22) … **það** var barið; Geimundur og Snjólaug komu
EXPL be.PST knock.PASS.PTCP Geimundur and Snjólaug come.PST
jafnsnemma til dyranna…
same-time to doors.DEF
'…someone knocked; Geimundur and Snjólaug came to the door at the same time…' (1902, Fossar.1623)

(23) **Það** má reyna að telja einhverjum öðrum en mér trú
EXPL may try.INF to tell.INF someone other than I.DAT truth
um það.
about DEM
'One may try to tell someone other than me the truth about that.'
(1908, Ofurefli.1638)

---

21. This sentence also begins a new chapter. (See footnote 20.)

22. See Booth (forthcoming) for discussion of inherently subjectless predicates in Icelandic.

3.4 Expletive *það* in presentational constructions

Rögnvaldsson (2002) found that expletive *það* is absent in presentational construc-
tions in early Icelandic, first appearing in c.1500, with the verbs *vera* 'be' or *verða*
'become'. This claim is supported by the IcePaHC data, where *það* is absent in the
period 1150–1350, see Table 4.

Table 4. Proportion of prefinite expletive *það* in presentational constructions
in IcePaHC, 1150–2008

Period	Prefinite *það* (V2)	No *það* (V1)	Total	% *það*	χ^2
1150–1350	0	39	39	0.0%	***
1351–1550	5	33	38	13.2%	***
1551–1750	4	21	25	16.0%	***
1751–1900	35	19	54	64.8%	ns
1901–2008	86	5	91	94.5%	***
All periods	**130**	**117**	**247**	**52.6%**	

Later in the diachrony, increases in expletive *það* are visible in the data for 1751–
1900 and 1901–2008. Interestingly, the frequency of *það* in presentationals for
1901–2008 is higher (94.5%) than the frequency observed for impersonal con-
structions (79.1%), cf. Table 3.

The first examples with the expletive occur in the period 1351–1550, e.g. (24)–
(26). For those examples which do not start a new chapter, I provide the preceding
context to show that *það* cannot be analysed as anaphoric.

(24) **Það** var einn ríkur maður, að hann
EXPL be.PST.SG one.NOM.SG rich.NOM.SG man.NOM.SG COMP he
trúði ei rétt eftir því sem honum bar…
believe.PST NEG rightly after DEM REL he.DAT deserve.PST
'There was one rich man, that he did not rightly believe in that which he
deserved…' (1475, Ævintyri.175)[23]

(25) **Það** var einn kvinna er fastaði við brauð og
EXPL be.PST.SG one.NOM.SG woman.NOM.SG REL fast.PST with bread and
vatn fyrir Marju messu Magdalena].
water for Mary.GEN mass Magdalene.GEN
'There was one woman who fasted with bread and water for Mary Magdalene's
mass.' (1475, Ævintyri.17)[24]

---

23. This sentence begins a new chapter. (See footnote 20.)

24. This sentence also begins a new chapter. (See footnote 20). *Einn* is masculine in form and so
unexpected as a modifier of *kvinna* (FEM); this is however the text as presented in IcePaHC.

(26) Þar   voru   og   rómverskir og   júðar og   þeir sem kallaðir
     there be.PST also romans   and jews   and DEM   REL call.PASS.PTCP
     voru   proseliti, en   það voru   þeir sem snúist   höfðu   frá
     be.PST proselytes but DEM be.PST DEM   REL turn.PST.PTCP have.PST from
     heiðnum siðum og   til júðverskra hluta. **Það** voru   og   þeir
     heathen ways   and to   jewish   things EXPL   be.PST.PL also DEM.NOM.PL
     er   sum   ritning kalla   Gethinn…
     REL some writings call.PRS Gethinn
     'There were also Romans and Jews and those who were called Proselytes, but
     that was those who had turned from heathen customs to Jewish ways. There
     were also those whom certain writings call Gethinn…' (1525, Georgius.694)

The five examples with the expletive in the period 1351–1500 are attested in texts
from 1475–1525, thus at the very end of the second period. All five occur with the
verb *vera* 'be', in line with Rögnvaldsson's (2002) finding. In all five, the verb agrees
in number with the postfinite subject noun phrase, as in present-day Icelandic.
The two texts in which the examples are attested are translations rather than native
Icelandic compositions: *Ævintyri* is a translation from English and *Georgius* from
Low German. I return to this observation below.

Of the four examples with expletive *það* in 1551–1750, two occur in texts
which – though not direct translations – are acknowledged to have influence from
German and/or Danish. This applies to (27) from *Indiafari*, recognised to have
some German influence, and (28) from *Vidalin*, which has possible German and
Danish influence.[25]

(27) Ei   minnist   ég múrshæðina,   en   víst   meina   ég það
     NEG remember.PRS   I   wall-height.DEF   but certainly think.PRS I   DEM
     fimmtán faðma   hæð   vera.   **Það** eru   margir
     fifteen fathoms.GEN height be.INF EXPL   be.PRS.PL many.NOM.PL
     kimar   niður við sjómálið…
     bilges.NOM.PL down   by   high-waterline.DEF
     'I do not remember the wall height, but I certainly think it to be fifteen fathoms
     in height. There are many bilges down by the high waterline…'
                                                   (1661, Indiafari.73.1287)

---

25. The *Indiafari* extract reports speech of German-speaking soldiers, see: https://github.com/
antonkarl/icecorpus/blob/master/info/1661.indiafari.bio-tra.info [last accessed 26.09.2018].
     For details on German and Danish influence in *Vidalin*, see: https://github.com/antonkarl/
icecorpus/blob/master/info/1720.vidalin.rel-ser.info [last accessed 26.09.2018].

(28) Hvað er   nú framar? Þú ert   forsmaður, líður   órétt,   enginn
what be.PRS now further you be.PRS manager   pass.PRS injustice no-one
vill   kannast   við þig, það er   ekki
wish.PRS reckon.INF with you EXPL be.PRS.SG nothing.NOM.SG
nýtt   í heiminum.
new.NOM.SG in world.DEF
'What more is there? You are the manager, you pass injustice, no-one wishes
to reckon with you, there is nothing new in the world.'   (1720, Vidalin.603)

The remaining two examples in 1551–1750 represent the first presentational con-
structions with *það* in native Icelandic compositions, shown here in (29) and (30).

(29) Þar   með gjörði hún so að inna: "Það vekjast   upp
there with do.PST she so to tell.INF EXPL awake.PRS.MID.PL up
harmar   í hjarta mínu…
sorrows.NOM.PL in heart   my
'Therewith she went to tell the following: there awakens sorrow in my heart…'
(1675, Armann.113.775)

(30) En síðan tóku   þeir að smá-dirfast,   báðu   þeir kónginn um
but then begin.PST they to small-dare.INF ask.PST they king.DEF about
Viðey, það væri   einn   klén   hestahólmi,
Viðey EXPL be.PST.SBJV.SG one.SG.NOM small.SG.NOM horse-holm.SG.NOM
svo sem vera   mætti fyrir eina tvo hesta eða þrjá…
so as   be.INF might for   one two horses or three
'But then they started to be slightly daring, they asked the king about Viðey,
there would be one small horseholm, such as could be for one, two or three
horses…'   (1680, Skalholt.127)

There are, however, earlier examples of presentational constructions with a clause-
initial *þar*, formally identical to the locative adverb *þar* 'there', which is plausibly
expletive, e.g. (31)–(32). Both *Græn* and *Gunnar* are considered native Icelandic
compositions.

(31) **Þar** var   suðurmaður   einn   *í ferð*   er
there be.PST.SG south-man.NOM.SG one.NOM.SG in company REL
Tyrkir hét.
Tyrkir be-called.PST
'There was one southern man in the company, who was called Tyrkir.'
(1250, Græn, Chapter 2)[26]

---

26. *Græn* refers to *Grænlendinga saga* (c.1250) which is one of the texts which was manually
examined for expletive contexts in Booth (forthcoming). The edition used was that available
on the online text repository *Icelandic Netútgáfan*: https://www.snerpa.is/net/index.html [last
accessed 26.09.2018].

(32) Þar   var   ein   hella   stór   *á   vellinum*.
there be.PST.SG one.NOM.SG rock-slab.NOM.SG big.NOM.SG on field.DEF
'There was a certain big slab of rock in the field.'   (1400, Gunnar.679)

In (31)–(32), there is no specified location in the preceding context to which *þar* could have anaphoric reference. Furthermore, each example has a location specified later in the sentence (shown in italics). These two observations I take as indication that *þar* is an expletive, rather than a locative adverb. This contrasts with present-day Icelandic, where only *það* 'it' is available as an expletive in presentational constructions. If one accepts such examples as involving an expletive *þar*, this would challenge standard accounts in two ways. Firstly, it is widely assumed that Old Icelandic had no expletive in presentational constructions (see Section 2.3). Secondly, the possibility that *þar* was available as an expletive in Icelandic presentational constructions historically is not considered in the literature.

Later examples with a plausibly expletive clause-initial *þar* also exist, e.g. (33)–(34). Again, each example features a locative later in the sentence.

(33) **Þar**   var   gnægð   grjóts   og   viður   *hér*
there be.PST.SG abundance.NOM.SG stone.GEN and wood.NOM.SG here
og   þar   við   sjávarströndina með björgunum.
and there by   sea-beach.DEF with cliffs.DEF
'There was an abundance of stone and some wood here and there by the beach with the cliffs.'   (1650, Illugi.563)

(34) **Þar**   bjó   *í Kaupinhafn* einn   slátrari   sá   er
there live.PST.SG in Copenhagen one.NOM.SG butcher.NOM.SG DEM REL
Kristján hét…
Kristján be-called.PST
'There lived in Copenhagen one butcher who was called Kristján…'
(1661, Indiafari.35.258)

If one accepts the analysis of *þar* as an expletive in (33)–(34), then these examples, together with examples like (29)–(30) above, indicate that there was a stage in Icelandic at which both *þar* 'there' and *það* 'it' were available as expletives in presentational constructions. Cases of competition between two expletive forms have been observed elsewhere in historical Germanic: see e.g. Faarlund (1990, pp. 70–72) and Kinn (2011) on Middle Norwegian; Brevik (1983, pp. 257–259, 324) and Pfenninger (2009, pp. 54–56) on English. I leave further examination of the Icelandic case of competition for future research.

4.   Discussion

4.1   A contact-induced change?

Returning to the status of *það* as an expletive, the earliest examples of presentational constructions with *það* in IcePaHC were found in translation texts, see (24)–(26), or in texts which are assumed to have some influence from other languages, see (27)–(28). This in line with the study by Rögnvaldsson (2002), who found the earliest examples of presentational constructions with *það* in the English translation *Ævintyri*, see (24)–(25). This observation leads Eythórsson & Sigurðardóttir (2016, p. 102) to claim that it is "likely that the filler *það* found its way into Icelandic due to foreign influence". This seems a rather bold leap and is unlikely for a number of reasons.

Firstly, expletive *það* was not wholly absent before this point. As I showed in Section 3.3, there are examples of impersonal constructions in Old Icelandic (1150–1350) where expletive *það* is present, e.g. (15)–(17). The later appearance of *það* in other contruction types does not, therefore, represent a complete innovation in the language.

Secondly, as I argued in Section 3.4, an expletive *þar* is attested in presentational constructions before the first appearances of expletive *það* in this context. So the availbility of an expletive in Icelandic presentational constructions was already a feature of the grammar, before examples of expletive *það* appear in translation texts.

Thirdly, the example in (18) with expletive *það* casts further doubt on the contact theory. This example comes from the *Ævintyri* text, the English translation which has been at the centre of the contact claim. (18) is an impersonal passive, a construction which is not grammatical in English; this instance of expletive *það* thus cannot be accounted for in terms of a wholesale borrowing from English.[27]

Finally, typological evidence shows that structural borrowings occur relatively rarely and typically require intense contact between speaker populations (e.g. Thomason & Kaufman, 1988), circumstances not documented for this period. Translation could be considered relatively intense contact, and it is plausible that the examples in (24)–(25) represent the first instances of expletive *það* in Icelandic presentational constructions, on the model of English. However, the claim that the expletive became a fully productive feature of Icelandic grammar from these sparse contact-related examples is implausible, based on our current understanding of language change, where change is more frequently acknowledged to work from the bottom-up, rather than in a top-down fashion. Moreover, as the results presented in Table 4 show, there is a significant time gap between the expletives in this text

---

27. I thank an anonymous reviewer for pointing this out.

(*Ævintyri*, c.1475), and the development whereby expletive *það* becomes frequent in presentational constructions (in the data for 1751–2008).

For these reasons, an analysis in terms of spread within the language seems more plausible than an exclusively contact-induced explanation for the emergence of expletive *það* in Icelandic. In Booth (forthcoming), I present such an account which argues that expletive *það* spread to new construction types – including presentational constructions where it out-competed the earlier expletive *þar* – on the model of earlier topicless constructions with cataphoric *það* and a clausal argument, e.g. (10)–(12).

4.2   A closer look at distribution

In Section 2.3, I introduced the Prefinite First Hypothesis which states that prefinite expletives emerge before subject expletives in historical Germanic. Though widely accepted, this hypothesis has not been tested for Icelandic. In present-day Icelandic, expletive *það* is restricted to the clause-initial prefinite position in impersonal and presentational constructions, thus behaving like a prefinite expletive. (See Section 2.2.) Given this distribution in the present-day language, if the Prefinite First Hypothesis holds for Icelandic, then one expects no instances of the expletive in postfinite position (i.e. as an unambiguous subject expletive) in the earlier periods.

I test this hypothesis against the IcePaHC data. Of all instances of expletive *það*, I examine to what extent these are restricted to the prefinite position.[28] Impersonal and presentational constructions are examined together. Contra the hypothesis, postfinite expletives do occur in all IcePaHC periods, although prefinite expletives are much more frequent than postfinite expletives, see Table 5.

Table 5. Position of expletive *það* in impersonal and presentational constructions in IcePaHC, 1150–2008

Period	Prefinite EXPL	Postfinite EXPL	Total	%Prefinite	χ^2
1150–1350	6	1	7	85.7%	ns
1351–1550	9	3	12	75.0%	**
1551–1750	6	3	9	66.7%	***
1751–1900	52	7	59	88.1%	ns
1901–2008	154	1	155	99.4%	**
All periods	**227**	**15**	**242**	**93.8%**	

---

28. The positional distribution of cataphoric *það* in the history of Icelandic is investigated in Booth (forthcoming).

I provide select examples of postfinite expletives in (35)–(38).

(35) Var   það síð dags.
     be.PST EXPL late day.GEN
     'It was late in the day.'   (1350, Finnbogi.635.548)

(36) En frá   keisaranum er   það að segja.
     but from emperor.DEF be.PRS EXPL to say.INF
     'But one speaks of the emperor.'   (1525, Erasmus.402)

(37) Margt   er   það líka, sem enn er   óljóst   í
     much.NOM.SG be.PRS.SG EXPL also REL still be.PRS unclear in
     ævisögu   jarðarinnar…
     biography earth.GEN.DEF
     'There is also much which is still unclear in the biography of the Earth…'
                              (1835, Jonasedli.141)

(38) Sömuleiðis var   það hneisa   fyrir landið,   ef höfuðsmaður
     same-time be.PST.SG EXPL shame.NOM.SG for   land.DEF if chieftain
     bæri   íslenzkum   veizlum   illa   söguna
     bear.PST.SBJV Icelandic.DAT celebrations.DAT bad.ACC story.ACC.DEF
     meðal hærri stéttar manna   erlendis…
     among higher classes men.GEN abroad
     'At the same time there was shame across the land, if a chieftain would bring
     to Icelandic celebrations the bad story among higher classes of men abroad…'
                              (1882, Torfhildur.132.1860)

Moreover, there are also examples of expletive *þar* in postfinite position in presentational constructions in the earlier data, e.g. (39)–(40).

(39) Talaði Víglundur það   einn tíma að   hann   vildi   að
     tell.PST Víglundur CATPH one time COMP he.NOM wish.PST COMP
     þau   byndu   sína   ást með fastmælum en Ketilríður
     they.NOM bind.PRS.SBJV their-own love with fast-talk   but Ketilríður
     gaf   sér fátt að: "Eru   þar," segir hún   "margir
     give.PST REFL few PTCL be.PST.PL EXPL say.PRS she.NOM many.NOM.PL
     hlutir   í móti".
     thing.NOM.PL against
     'Víglundur said one time that he wished that they would seal their love with
     an engagement but Ketilríður took coldly to this: "There are", she says, "many
     things against that." '   (1400, Viglundur.539)

(40) Gekk   hann   í móti   sínum   fóstbróður   með allmikilli gleði, og
      go.PST he.NOM towards his-own foster-brother with all-great joy   and
      var   **þar**   allmikill   fagnafundur…
      be.PST.SG EXPL all-great.NOM.SG joyful-meeting.NOM.SG
      'He went towards his foster brother with great joy and there was a greatly joyful
      meeting….'   (1480, Jarlmann.569)

These examples are evidence that *það* cannot be exclusively accounted for as a structurally motivated filler for V2. Nevertheless, given that the expletive overwhelmingly prefers the prefinite position in all periods, it does not seem reasonable to claim that *það* qualifies as a fully-fledged subject expletive in the historical data.

A final observation of the findings in Table 5 is that the positional distribution of expletive *það* appears to undergo a change as of 1901; in the data for 1901–2008, virtually all instances of *það* occur in prefinite position, in contrast to the earlier periods which show a more mixed picture.[29] The IcePaHC data therefore indicates that the restricted positional distribution usually claimed in the literature for *það* holds for modern Icelandic (post-1901), but not for earlier stages (pre-1901).

5.   Conclusion

The corpus study presented in this paper sheds new light on the status of expletives in historical Icelandic. Two claims in the literature were confirmed by the IcePaHC data: (1) cataphoric *það* is indeed frequently present in constructions with a clausal argument in the earliest texts; (2) expletive *það* was confirmed to be absent in presentational constructions prior to c.1475, when it begins to appear in translation texts. However, contrary to the literature it was shown that expletive *það* is already attested in impersonal constructions in Old Icelandic (1150–1350).

I argued that an earlier form *þar* 'there' was available as an expletive in presentational constructions and persists in later periods, thus overlapping with expletive *það*. The availability of these two options challenges previous accounts of historical Icelandic, which neglect the availability of *þar* as an expletive. Furthermore, in light of the overall corpus findings, a recent claim by Eythórsson & Sigurðardóttir (2016) that the emergence of expletive *það* is a contact phenomenon was argued to be implausible.

More broadly, the IcePaHC findings pose a challenge for the Prefinite First Hypothesis assumed by many for the diachrony of expletives in Germanic. The

---

29. The postfinite expletive for 1901–2008 is in a text from 1902, i.e. at the very beginning of the period.

corpus findings indicate that the Icelandic expletive did not start out exclusively as a prefinite expletive, contra the predictions of the hypothesis. The positional prefinite restriction usually claimed for the Icelandic expletive is thus in fact a relatively recent phenomenon (post-1901). In sum, the Icelandic data – together with other recent studies which pose similar problems (Axel, 2007 on German; Kinn, 2016 on Norwegian) – indicate that the Prefinite First Hypothesis should be revisited, particularly in light of the empirical opportunities offered by the current availability of historical syntactic corpora for many Germanic languages.

Acknowledgements

Abbreviations

ANAPH	anaphoric	NEG	negation
ACC	accusative	NOM	nominative
CATPH	cataphoric	PASS	passive
COMP	complementizer	PL	plural
DAT	dative	PRS	present
DEM	demonstrative	PST	past
GEN	genitive	PTCP	participle
IcePaHC	Icelandic Parsed Historical Corpus	REL	relative
INF	infinitive	SG	singular
MASC	masculine	SBJV	subjunctive

References

Axel, K. (2007). *Studies on Old High German Syntax: Left Sentence Periphery, Verb Placement and Verb-Second*. Amsterdam: John Benjamins. https://doi.org/10.1075/la.112

Booth, H. (forthcoming). Expletives and Clause Structure: Syntactic Change in Icelandic. Doctoral dissertation, University of Manchester.

Breivik, L. E. (1983). *Existential there: A Synchronic and Diachronic Study*. Bergen: Department of English, University of Bergen.

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris publications.

Eythórsson, T. & Sigurðardóttir, S. S. (2016). A brief history of Icelandic weather verbs: syntax, semantics and argument structure. *Working Papers in Scandinavian Syntax*, 96, 91–125.

Faarlund, J. T. (1990). *Syntactic Change: Toward a Theory of Historical Syntax*. Berlin; New York: de Gruyter. https://doi.org/10.1515/9783110854947

Faarlund, J. T. (2004). *The Syntax of Old Norse: with a survey of the inflectional morphology and a complete bibliography*. Oxford: Oxford University Press.

Falk, C. (1993). *Non-referential Subjects in the History of Swedish*. Lund: Department of Scandinavian Languages, University of Lund. https://doi.org/10.1016/0024-3841(93)90051-W

Haiman, J. (1974). *Targets and Syntactic Change*. The Hague: Mouton.

Hróarsdóttir, Þ. (1998). *Setningafræðilegar Breytingar á 19. öld: Þróun Þriggja Málbreytinga*. Reykjavík: Institute of Linguistics, University of Iceland.

Ingham, R. (2001). The structure and function of expletive *there* in pre-modern English. *Reading Working Papers in Linguistics*, 5, 231–249.

Kinn, K. (2011). Overt non-referential subjects and subject-verb agreement in Middle Norwegian. *Working Papers in Scandinavian Syntax*, 88, 21–50.

Kinn, K. (2016). Referential vs. non-referential null subjects in Middle Norwegian. *Nordic Journal of Linguistics*, 39(3), 277–310. https://doi.org/10.1017/S0332586516000068

Kroch, A. & Taylor, A. (2000). *The Penn-Helsinki Parsed Corpus of Middle English (PPCME2)*. Philadelphia: Department of Linguistics, University of Pennsylvania, second edition. http://www.ling.upenn.edu/hist-corpora/

Lenerz, J. (1985). Zur Theorie syntaktischen Wandels: das expletive *es* in der Geschichte des Deutschen. In W. Abraham (Ed.), *Erklärende Syntax des Deutschen* (pp. 99–136). Tübingen: Gunter Narr.

Maling, J. (1988). Variations on a theme: Existential sentences in Swedish and Icelandic. *McGill Working Papers in Linguistics. Special Issue on Comparative Germanic Syntax*, 168–191.

Pfenninger, S. E. (2009). *Grammaticalization Paths of English and High German Existential Constructions: A Corpus-based Study*. Bern: Peter Lang.

Platzack, C. (1983). Existential sentences in English, German, Icelandic and Swedish. In F. Karlsson (Ed.), *Papers from the 7th Scandinavian Conference of Linguistics*, (pp. 80–100). Helsinki: Department of General Linguistics, University of Helsinki.

Randall, B. (2005). *CorpusSearch2 User's Guide*. Philadelphia: Department of Linguistics, University of Pennsylvania. http://corpussearch.sourceforge.net/CS-manual/Contents.html

Richards, M. & Biberauer, T. (2005). Explaining Expl. In M. den Dikken & C. Tortora (Eds.), *The Function of Function Words and Functional Categories*, (pp. 115–153). Amsterdam: John Benjamins. https://doi.org/10.1075/la.78.06ric

Rögnvaldsson, E. (2002). ÞAÐ í fornu máli – og síðar. *Íslenskt mál*, 24, 7–30.

Santorini, B. (2010). *Annotation manual for the Penn Historical Corpora and the PCEEC*. Philadelphia: Department of Linguistics, University of Pennsylvania. http://www.ling.upenn.edu/hist-corpora/annotation/index.html

Sells, P. (2005). The peripherality of the Icelandic expletive. In M. Butt & T. Holloway King (Eds.), *Proceedings of the LFG '05 Conference* (pp. 408–428). Stanford, CA: CSLI Publications.

Sigurðsson, H. Á. (1990). V1 declaratives and verb raising in Icelandic. In J. Maling & A. Zaenen (Eds.), *Syntax and Semantics: Modern Icelandic Syntax* (pp. 41–69). San Diego, CA: Academic Press.

Sigurðsson, H. Á. (2007). Argument features, clausal structure and the computation. In T. Bhattachayra, E. Reuland & G. Spathas (Eds.), *Argument Structure* (pp. 121–158). Amsterdam: John Benjamins. https://doi.org/10.1075/la.108.10sig

Thomason, S. G. & Kaufman, T. (1988). *Language Contact, Creolization, and Genetic Linguistics*. Berkeley, CA: University of California Press.

Thráinsson, H. (1979). *On Complementation in Icelandic*. New York: Garland.

Thráinsson, H. (2007). *The Syntax of Icelandic*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511619441

Vikner, S. (1995). *Verb Movement and Expletive Subjects in the Germanic Languages*. Oxford: Oxford University Press.

Wallenberg, J. C., Ingason, A. K., Sigurðsson, E. F. & Rögnvaldsson, E. (2011). Icelandic Parsed Historical Corpus (IcePaHC), version 0.9. http://linguist.is/icelandic_treebank

Weenen, A. de Leeuw van (Ed.). (1993). *The Icelandic Homily Book: Perg. 15 4º* in the Royal Library, Stockholm. Reykjavík: Stofnun Árna Magnússonar á Íslandi.

Zaenen, A., Engdahl, E. & Maling, J. (2017). Subject properties in presentational sentences in Icelandic and Swedish. *Bergen Language and Linguistics Studies*, 8(1), 260–281. https://doi.org/10.15845/bells.v8i1.1335

PART VI

Languages in contact

Contact and change in Neo-Aramaic dialects

Geoffrey Khan
University of Cambridge

Aramaic, a Semitic language, has survived down to modern times as a spoken language in a large diversity of Neo-Aramaic dialects. This paper examines various aspects of contact-induced linguistic change in the subgroup of dialects known as North-Eastern Neo-Aramaic (NENA). These dialects have for many centuries been in contact with various other languages, including Semitic (Arabic) and non-Semitic (Kurdish, Persian, Armenian, Turkic languages). Various motivating factors can be identified for contact-induced change in the NENA dialects. These are sociolinguistic and internal systemic. When change occurs it often involves only partial convergence. Change sometimes results in imitations of the morphology of the contact language using internal morphological elements. Sociolinguistic and internal systemic factors can also inhibit change in a contact situation.

Keywords: Aramaic, Neo-Aramaic, contact-induced change, perfect, copula, word order, homophony, aspiration, ergative, Kurdish, Armenian

1.   The Neo-Aramaic dialects

The Neo-Aramaic dialects are modern vernacular forms of Aramaic, a Semitic language, which has a documented history in the Middle East of over 3,000 years, the earliest inscriptions being datable to approximately 1,000 B.C. The Neo-Aramaic dialects that have survived down to modern times are generally classified into four subgroups:

1. Central Neo-Aramaic (southeastern Turkey west of the Tigris)
2. North-Eastern Neo-Aramaic (northern Iraq, southeastern Turkey and western Iran)
3. Neo-Mandaic (southwestern Iran)
4. Western Neo-Aramaic (Syria)

Due to upheavals in the Middle East over the last one hundred years thousands of speakers of Neo-Aramaic dialects have been forced to migrate from their homes or have perished in massacres. As a result, the dialects are now highly endangered.

In this paper I shall focus on the Neo-Aramaic dialects of the NENA subgroup, which was spoken by Christian and Jewish communities. The dialects of this subgroup have in general undergone a more advanced degree of historical change than other subgroups. Within the NENA subgroup, moreover, historical change has resulted in considerable diversity across the various dialects (Khan, 2007, 2011).

One factor that has had a bearing on divergent development of dialects is the existence of social boundaries between communities. This is reflected most clearly by the fact that there are many differences in all levels of grammar and also in the lexicon between dialects spoken by Christians and those spoken by Jews in the same geographical area. This is exemplified in (1), in which selected features from the Christian and Jewish dialects of Urmi (northwestern Iran) are compared:

(1)		Christian Urmi (Khan, 2016)	Jewish Urmi (Khan, 2008a)
Phonology			
*θ		t	l
*bayθa 'house'		béta	belá
Morphology			
3ms gen. suffix		-u	-éu
1ms copula		-ívən	-ilén
Syntax			
Word order		basic VO	basic OV
Lexicon			
big		+ɉúra[1]	+rəwwá
hair		cósa	+məsyé
he went		xə́ʃle	zə́lle

It is important to note that the Christian and Jewish communities in the region were not separated from each other geographically or physically (there were no Jewish ghettos). There was, moreover, professional contact between them. The crucial factor that induced divergent linguistic development was the different group identities and social networks of the two communities (Milroy, 1987; Trudgill, 1986, 1989).

---

1.   The symbol [+] denotes suprasegmental pharyngealization of the word.

Another major factor that has given rise to dialectal diversity is the impact of languages in contact with Neo-Aramaic in Western Asia. Numerous languages, belonging to a variety of language families, are currently in contact with the NENA dialects. The major ones include Kurdish (Kurmanji and Sorani varieties), Persian, Armenian, Turkic and Arabic. The influence was almost completely unidirectional, since there is no clear evidence of convergence of these languages with NENA, apart from a few NENA loanwords in Kurdish (Chyet, 1995).

In this paper I shall examine various aspects of contact-induced change in the NENA dialects.

2.   Constraints on change

Firstly I should like to draw attention to factors that have blocked change.

2.1   Constraint on lexical transfer

One of these is a socially-motivated constraint on lexical transfer from contact languages. This is reflected by the almost total exclusion of Armenian loanwords from NENA. Many speakers of NENA dialects were in close contact with speakers of Armenian. This contact was particularly close in the Urmi region, where there were many intermarriages between the Christian NENA-speaking communities and the Armenian-speaking communities. This contact is reflected by borrowing of some features of phonology and syntactic structure, but there are virtually no Armenian loanwords in the NENA dialects of the Urmi region and no documented code-switching. This appears to be the result of a deliberate attempt to mark a boundary between the group identity of the NENA-speaking community from that of the Armenian-speaking community. The motivation is the need to preserve NENA group identity in a situation of intimate social connection and cultural homogeneity between the two groups, in which the boundaries between group identities are particularly under threat. No such attempt was made to exclude loanwords from socially more distant groups such as those of speakers of Kurdish or Turkic, who do not constitute a threat to the group identity of NENA-speakers.[2] The replication in NENA of Armenian phonological features and syntactic patterns would have been below speakers' "level of awareness" (cf. Silverstein, 1981).

---

[2] For a similar phenomenon of conscious lexical exclusion between speech communities in close contact in Amazonia see Epps and Stenzel (2013, p. 36) and Floyd (2013), and the general discussion about language contact in the region by Aikhenvald (2003).

2.2   Size of community and geographical location

Other factors that constrain convergence with contact languages include the size of the language community and the terrain of its geographical location. Some of the most conservative NENA dialects are those that were spoken by the Christian Tiyare tribes, who, until 1915, lived in a large group of NENA-speaking villages in the Hakkari mountains of southeastern Turkey (Gaunt, 2006, p. 414; Talay, 2008a, 2008b). The dialects of NENA-speaking communities that were smaller in size have typically converged to a greater extent with languages in contact. Some of the dialects that have undergone the most radical contact-induced linguistic changes are those spoken by communities of Jews in northeastern Iraq and western Iran. All of these communities were small tight social networks embedded within villages or towns where the majority of the population spoke other languages. Many were very small communities, in some cases consisting of only a handful of families in a village, e.g. Dobe, Ruwanduz, Rustaqa (Khan, 2002b). As for terrain, the Hakkari region where the Tiyare tribes lived was mountainous, which is likely to have contributed to their geographical isolation from contact languages and also made the networks between the Tiyare villages looser. A high density of population of NENA speakers in villages on the Urmi plain (northwestern Iran, Gaunt, 2006, p. 417), which were less geographically isolated, did not result in such linguistic conservatism, on account of higher exposure to contact and a geography that was more conducive to tighter networks. The linguistic conservatism of the Tiyare tribes appears to be the result of a combination of their geographical isolation and their large population of loose networks. This would correspond to the theory of socio-linguists that innovations are diffused less easily in large loosely networked communities than in small tightly networked communities (e.g. Nettle, 1999, p. 139; Trudgill, 2009, p. 102).

2.3   Potential homophony

A further factor that blocks contact-induced change is the need to prevent homophony in linguistic systems. An example of this can be found in the Jewish dialects on the eastern periphery of the NENA area, in western Iran and adjacent areas in northeastern Iraq. In the majority of NENA dialects the subject inflection of the past stem of verbs is expressed by a series of oblique suffixes. The verbal system of Jewish dialects on the eastern periphery has converged more closely with that of Iranian languages of the area and has replaced the oblique inflection of past stems with direct, nominative, person suffixes in intransitive verbs that do not have agentive actionality.[3] This has taken place in all non-agentive verbs except the copula

---

3.   For further details of this aspect of the NENA verbal system see Coghill (2016) and Khan (2017).

verb 'to be'. The reason for this is that a change from oblique to direct person suffixes would have resulted in homophony between the past and present copula:

(2) J. Sanandaj[4] (western Iran) (Khan, 2009)

past transitive:	*ʃte-le*	'he drank'
	drink.PST-OBL.3MS	
past intransitive:	*bxe-Ø*	'he wept'
	weep.PST-NOM.3MS	
past copula:	*ye-le*	'he was'
	be.PST-OBL.3MS	
present copula:	*ye-Ø*	'he is'
	be.PRS-NOM.3MS	

2.4 Change inhibited by contact

Finally, contact with another language can act as a constraint on change of a particular feature, if the contact language shares this feature with NENA. This applies to some NENA dialects that have been in contact with Arabic dialects for many centuries. On account of the genetic relationship between Arabic and Aramaic, both of which are Semitic languages, they share some features that are not found in languages of other genetic families in the area. In NENA dialects in contact with Arabic certain features that correspond in form to those of the Arabic dialects of the area are preserved whereas these features undergo change in NENA dialects that are not in contact with Arabic. An example of this is the development of the interdental consonants /θ/ and /ð/, which belonged to the historical consonantal inventory of the NENA dialects. In many dialects these have been eliminated, in most cases by merging them with stops or sibilants, on account of the fact that interdentals do not occur in the contact languages. On the Mosul plain in northern Iraq, however, where NENA dialects have been in close contact with Arabic dialects that have interdentals, the NENA interdentals are preserved. This applies, for example, to the NENA dialect of Qaraqosh (Mosul plain, Khan, 2002a), which preserves interdentals as distinct from stops. This is contrasted below in (3) with the Christian NENA dialect of Urmi (northwestern Iran, Khan, 2016), in which the interdentals have merged with stops:

(3)

Mosul Arabic[5]	Qaraqosh	C. Urmi
/t/, /θ/	/t/, /θ/	/t/
/d/, /ð/	/d/, /ð/	/d/

---

4. Christian and Jewish NENA dialects of a particular place are distinguished by the abbreviations C. and J. respectively.

5. Jastrow (1979).

Some other documented cases of contact inhibiting change involves, as here, the contact of closely related languages, e.g. Jastrow (2015), who discusses some cases of the conservative influence of Arabic on the phonology of Central Neo-Aramaic dialects, and Enrique-Arias (2010) on the lack of change in Spanish in Majorca due to contact with Catalan.[6] Dickey (2011), in his discussion of the conservative influence of German on the western Slavic verbal system calls this a process of 'replica preservation'.[7]

3. Systemic motivations for contact-induced change

3.1 Elimination of homophony

We have seen in § 2.3. how contact-induced change may be blocked in a particular item in a linguistic system if this change would have eliminated functionally significant morphological distinctions. Conversely, if functionally significant morphological distinctions have been lost by internal developments in a NENA dialect, then convergence with contact languages may take place to compensate for this.

An example of such systemically motivated borrowing can be identified in the paradigm of the possessive suffixes in NENA. The 2fs suffix is *-ax*. The historical form of the 2ms suffix is *-āx* and by the normal process of historical phonology the reflex of this in the modern NENA dialects should have been *-ax*, i.e. a homophone of the 2fs suffix. In order to resolve this homophony the long *ā* of the historical 2ms form *-āx* shifts to /o/, which results in the maintenance of paradigm distinction between the 2ms *-ox* and 2fs *-ax*. The shift of *ā* to a back rounded vowel is not a general feature of NENA but is a feature of languages spoken in areas adjacent to NENA to the west of the Tigris river. This applies to Kurdish and Armenian dialects of the area and also the Central Neo-Aramaic subgroup (Jastrow, 2011). The shift of *ā* > /o/ in the 2ms suffix *-ox*, which is a general feature of all NENA dialects, can be identified as a borrowing from the phonology of neighbouring languages that is motivated by a morphological system.[8]

---

6. I thank Sarah Thomason for the last reference.

7. I am grateful to one of the anonymous reviewers for this reference.

8. Cf. the work of Malkiel (1968, 1976) on the morphological motivations for "irregular" sound changes in Romance.

3.2   Enrichment of resources

In some cases a borrowing from contact languages can result in the enriching of the original linguistic systems in NENA dialects. An example of this is the development of a phonemic distinction between aspirated and unaspirated unvoiced stops in some NENA dialects. In earlier Aramaic, and Semitic in general, unvoiced stops are aspirated. In some NENA dialects, such as C. Urmi (Khan, 2016, vol. 1, 92–110), an additional series of unaspirated stops and affricates has developed through convergence with the phoneme inventories of Kurmanji Kurdish and Eastern Armenian (in (4) the unaspirated phonemes are distinguished with a circumflex diacritic above or below the letter):

(4)   Proto-NENA C. Urmi E. Armenian Kurmanji
        labials
        voiceless aspirated   *p^h /p^h/   /p^h/   /p^h/
        voiceless unaspirated –   /$\hat{p}$/   /$\hat{p}$/   /$\hat{p}$/
        voiced   *b   /b/   /b/   /b/
        dental/alveolar
        voiceless aspirated   *t^h, *θ /t^h/   /t^h/   /t^h/
        voiceless unaspirated –   /ṱ/   /ṱ/   /ṱ/
        voiced   *d, *ð   /d/   /d/   /d/

The unaspirated stops have developed in various contexts in C. Urmi and are put to functional use. One such case is found in the morphophonology of verbs. The past perfective inflection of NENA verbs begins with a consonant cluster. When the first consonant of the cluster is a fricative and the second is an unvoiced stop, the stop is unaspirated in all dialects by a well-known process of de-aspiration after fricatives, e.g.

(5)   Qaraqosh (Khan, 2002a)
        past perfective   present
        /$xp^h\partial rre$/ [xp̂ərre]   /$xap^h\partial r$/ [xapʰər]
        dig.PST.OBL.3MS   dig.PRS.3MS
        'he dug'   'he digs'
        root /x-p^h-r/ 'to dig'

In NENA dialects that have unaspirated stops in their phonological system, such as C. Urmi, the unaspirated allophone in the cluster of the past perfective forms is re-analysed as an underlying unaspirated phoneme, i.e. [xp̂] > /xp̂/. This has been licensed by convergence with the phonological system of languages in contact. As a result, the unaspirated stop has come to be analysed as a radical of the morphological root, i.e. /x-p̂-r/. In other inflections of the verb where the stop is separated from the fricative by a vowel, therefore, the stop is still unaspirated:

(6) C. Urmi   (Khan, 2016)

      past perfective   present

      /xṗərrə/ [xṗərrə]   /xaṗər/ [xaṗər]

      dig.PST.OBL.3MS   dig.PRS.3MS

      'he dug'   'he digs'

      root /x-ṗ-r/ 'to dig'

The reanalysis of the allophone as an underlying unaspirated stop has the function of regularizing the inflectional system and reducing complexity in the correspondence between surface phonetics and the underlying phonology. This situation reflects Dahl's observation (2004, pp. 42–45) that an increase in complexity of resources (in our case the addition of a new series of phonemes) can be sometimes matched by a decrease in complexity of the rules of a language system.

4. Partial convergence

4.1 Perfect construction

The various NENA dialects exhibit differing degrees of convergence with the contact languages. This applies, for example, to the convergence of the NENA verbal system with that of the Iranian languages. The verbal system exhibited by the NENA dialects reflects a major restructuring of the system found in earlier Aramaic due to contact with Iranian. This included the development of ergativity in some constructions, which is a feature of Western Iranian (Coghill, 2016; Jügel, 2015). The convergence of most NENA dialects with the Iranian ergative structures, however, was only partial. This is clearly seen in the development of the NENA perfect. In many NENA dialects an innovative perfect construction has developed consisting of a resultative participle and copula. This replicates a Kurdish model. In Kurdish the construction has ergative alignment, whereby the participle and copula agree with the object of a transitive clause:

(7) Kurmanji   (Thackston, 2006a)

      wî   *ez*   *dît-ime*

      3S.OBL 1S.NOM see.PTCP-COP.1S

      'He has seen me'

In most NENA dialects, by contrast, transitive perfect clauses have accusative alignment, in that the copula and participle agree with the subject and the object has oblique marking. NENA replicates the morphological configuration of the Kurdish model (participle + copula) but not the syntactic alignment or argument structure of the Kurdish construction, e.g.

(8) C. Barwar (northern Iraq)   (Khan, 2008b)
    a. *ʾaw   xəzy-əle-li*
        he.NOM see.PTCP.M-COP.3MS-OBL.1S
        'He has seen me'
    b. *ʾay   xziθ-əla-li*
        she.NOM see.PTCP.F-COP.3FS-OBL.1S
        'She has seen me'

Some Jewish dialects on the eastern periphery of the NENA area, however, exhibit greater degrees of convergence with the Kurdish model. In the Jewish dialects of western Iran, for example, the participle and the copula agree with the object in transitive perfect constructions, e.g.

(9) J. Sanandaj   (Khan, 2009)
    ʾo-gora   baxt-ăke   grəʃta-ya
    that-man woman-the pull.PTCP.FS-COP.3FS
    'The man has pulled the woman'

What both these perfect constructions across the various dialects have in common is the head of the construction consisting of participle and copula. This is a replication of the corresponding morphology of the head of the perfect construction of Kurdish. In this case, therefore, there is a greater tendency to replicate the head of the construction than its morphosyntax and argument structure. NENA dialects that replicate also the syntactic structure can be said to have converged to a greater extent with the contact language. This can be expressed in the following hierarchy of convergence, in which the symbol > is to be read 'shows a greater tendency to be replicated in a language contact-situation than':

(10) head constituent > morphosyntax and argument structure of head

One can possibly identify a parallel to this split between converged head constituent and non-converged morphosyntax in the distinction in codeswitching between embedded language and matrix language (Myers-Scotton, 1993). According to Myers-Scotton (1993, p. 83), all system morphemes which have grammatical relations external to the head constituent will come from the matrix language even if the head is from the embedded language, i.e. embedded language head constituent with matrix language morphosyntax.

    In the model of contact-induced change proposed by Matras and Sakel (2007), one may say that the head constituent acts as the "pivot" of the convergence and the "replica construction evolves around the new pivot in a way that generally respects various constraints of the replica language" (Matras & Sakel, 2007, p. 830).

4.2   Copula

In earlier Aramaic the indicative present copula in nominal sentences is expressed by an anaphoric pronoun, which is typically enclitic. This is illustrated, for example, by Syriac, which has paradigms of independent personal pronouns and a corresponding paradigm of enclitic forms of these pronouns, which function as copulas:

(11)   Syriac

	Independent pronoun	Enclitic pronominal copula
3ms.	*hū*	*-ū*
3fs.	*hī*	*-ī*
3mpl.	*hennōn*	*-ennōn*
3fpl.	*hennēn*	*-ennēn*
2ms.	*ʾatt*	*-att*
2fs.	*ʾatt*	*-att*
2mpl.	*ʾattūn*	*-ttūn*
2fpl.	*ʾattēn*	*-ttēn*
1s..	*ʾenā*	*-nā*
1pl.	*ḥnan*	*-nan*

Enclitic copulas are an areal feature of the region and are found in the languages that are in contact with NENA, e.g. Armenian, Turkic languages, Kurdish and Persian. An enclitic copula is found also in Arabic dialects spoken in southeastern Turkey and northern Iraq belonging to the so-called *qəltu* sub-group (Jastrow, 1978, pp. 131–136). In Armenian and Iranian languages the copula is in origin a verb, though in most cases it has been reduced to verbal subject person markers.

Examples of the correspondence between the Kurdish enclitic copula and the person inflection of the present verb are given here for the standard forms of Kurmanji (Northern Kurdish) and Sorani (Central Kurdish):

(12)	Kurmanji Kurdish[9]		Sorani Kurdish[10]	
	Enclitic copula	Present verb inflection	Enclitic copula	Present verb inflection
3s	*-e*	*-e*	*-a*	*-e(t)*
3pl	*-in*	*-in*	*-in*	*-in*
2s	*-î*	*-î*	*-î(t)*	*-î(t)*
2pl	*-in*	*-in*	*-in*	*-in*
1s	*-im*	*-im*	*-im*	*-im*
1pl	*-in*	*-in*	*-în*	*-în*

---

9. Thackston (2006a, pp. 30, 34).

10. Thackston (2006b, pp. 25–26).

Some Jewish dialects on the eastern periphery of the NENA area exhibit a very close convergence with this Kurdish model, in that the copula, although pronominal in historical origin, has acquired verbal person inflection. This is seen, for example, in the J. Sanandaj NENA dialect:

(13) J. Sanandaj   (Khan, 2009)

	Independent Pronoun	Indicative Enclitic Copula	Irrealis copula verb
3ms.	*ʾo*	*-y-e*	*hăw-e*
3fs.	*ʾo*	*-y-a*	*hawy-a*
3pl.	*ʾoni*	*-y-en*	*hăw-en*
2ms.	*ʾāt*	*-y-et*	*hăw-et*
2fs.	*ʾāt*	*-y-at*	*hawy-at*
2pl.	*ʾaxtu*	*-y-etun*	*hăwe-tun*
1ms.	*ʾana*	*-y-ena*	*hăwe-na*
1fs.	*ʾana*	*-y-an*	*hawy-an*
1pl.	*ʾaxni*	*-y-ex*	*hăw-ex*

The enclitic copula has lost its structural relationship to the independent pronouns in J. Sanandaj. The inflectional endings of the enclitic copula have become those of the inflection of present verbal forms (i.e. verbal forms with a base that is derived historically from the active participle of earlier Aramaic). More specifically, the inflection of the copula resembles that of the present form of the verb *h-w-y* 'to be' (given in the right column above), which is used suppletively to express the irrealis and the future tense.

   The majority of NENA dialects do not exhibit a complete levelling of the inflection of the copula with that of the person markers of verbs but rather only a partial convergence. There is variation in convergence to the Kurdish model within the copula systems of individual dialects. Three hierarchies in this internal systemic variation may be represented as follows (the sign > is to be read 'has a greater tendency to develop verbal morphology than'):

(14) i.   1st and 2nd person > 3rd person
      ii.   Negative polarity   > Positive polarity
      iii. Past tense   > Present tense

This is reflected by the fact that in the main body of NENA dialects the 1st and 2nd person forms of the positive indicative copula have verbal inflection but not the 3rd person forms (Khan, 2001, 2012) and in several NENA dialects the 3rd person form has developed verbal inflection in the negated copula and past copula but not in the positive present copula (Khan, forthcoming).

   The items on the left side of these hierarchies are generally regarded as the marked members of the categories in question. There is, therefore, a greater

tendency here for convergence with the contact language to occur in the case of the marked items. This phenomenon may be compared to the observed fact that in language contact situations there is often loss of complexity, in particular in the non-dominant language. It has been reported that marked, i.e. complex, constructions in the non-dominant language are lost in language contact (e.g., Clyne, 1992). This in turn could perhaps be correlated with the loss of complexity in situations of suboptimal acquisition of language (Dahl, 2004, p. 281). Borrowing of a feature from a contact language is in effect the loss of that feature in the replicator language. Marked forms, therefore, would be lost more readily, on account of their semantic complexity, than unmarked forms, which are underspecified semantically.

4.3   Word order

Various degrees of convergence with contact languages across the NENA dialects can be identified also in the order of objects in verbal clauses. Christian dialects and Jewish dialects west of the Zab have a basic word order of VO in verbal clauses. Jewish trans-Zab dialects have a basic order OV. The OV syntax of the Jewish trans-Zab dialects is an innovation that has come about through convergence with contact languages, such as Kurdish and Turkic, which also have a basic OV order.

The NENA dialects that have a basic VO order in verbal clauses may front the O argument for pragmatic purposes. This pragmatic fronting is more frequent in some dialects than in others. It is more frequent in dialects in Iran (e.g. C. Urmi) than in dialects in Iraq (e.g. C. Barwar). This is seen in (15), which presents relative percentages of the occurrences of OV and VO in equivalent samples from C. Barwar (Iraq) and C. Urmi (northwestern Iran). These are compared with J. Urmi (trans-Zab, northwestern Iran), which has a basic OV order:

(15)		VO	OV
	C. Barwar	80%	20%
	C. Urmi	50%	50%
	J. Urmi	7%	93%

The differences in frequency of OV between C. Barwar and C. Urmi can be interpreted as reflecting differing degrees of convergence with the basic OV syntax of contact languages. This convergence is clearly greater in C. Urmi.

The lack of clear distinction between the overall frequency of VO and OV in the C. Urmi dialect evokes the question as to how one should decide what the basic word order is in this dialect. From a comparative perspective of NENA one can identify in addition to frequency also two further criteria:

(16) Diagnostics for change of basic word form VO to OV
    i.   Frequency of OV
    ii.   The auxiliary verb shifts to a position after the head verb
    iii. The locus of pragmatic strategies involving the O argument shifts from the pre-verbal field to the post-verbal field

Certain verbal constructions in the C. Barwar, C. Urmi and J. Urmi dialects consist of a non-finite verbal form and a verbal copula that functions as an auxiliary (C. Barwar *hawe*, C. Urmi *ʾavə*, J. Urmi *hawe*). In C. Barwar and C. Urmi the auxiliary is placed before the lexical verbal form, whereas in J. Urmi it is placed after it, e.g.

(17) C. Barwar: AUX—VERB   *hawe*   *ʾəθya*
                                 be.AUX.3MS   come.PTCP.MS
                                 'He may have come.'
      C. Urmi: AUX—VERB   *ʾavə tiyya*
      J. Urmi   VERB—AUX   *ədya-hawe*

The order lexical verb – auxiliary is the pattern that is found in Kurdish, e.g. Sorani (Thackston, 2006b, p. 61)

(18) *hāti-bā*
      come.PTCP-COP.IRR.3MS
      'He may have come.'

The postverbal position of the auxiliary in J. Urmi, therefore, reflects a greater degree of convergence with Kurdish than the preverbal positions in C. Barwar and C. Urmi. Whereas C. Urmi frequently places object arguments before the verb, exhibiting thereby some degree of convergence with the OV syntax of Kurdish, the auxiliary position in this dialect does not converge with the syntax of Kurdish and is regularly placed before the verb. The difference between the two constructions is that the fronting of the object argument before the verb can be used as an optional pragmatic strategy that is not a grammaticalized syntactic pattern, whereas the position of the auxiliary cannot be changed for pragmatic purposes without the grammatical pattern changing. The position of the auxiliary vis-à-vis the head verb is a grammaticalized pattern and it only changes when there is a change of grammatical pattern. It reflects, therefore, a change in grammatical word order, i.e. basic word order. This correlates with an objectively verifiable increase in placement of the O argument before the verb and so it is jusitifiable to posit that the change in grammatical word order reflected by the changed auxiliary position is correlated with a change in basic grammatical order of the verb and its object complement. This is in line with the cross-linguistic tendency for auxiliary verbs to precede the lexical

verb in VO languages and follow the lexical verb in OV languages (Anderson, 2011, p. 297; Dryer, 2009; Greenberg, 1966, universal 16).

We may infer from this that convergence with the syntactic arrangement of the contact language is more likely if this order can be used with a pragmatic function without changing the basic grammatical word order:

(19)   pragmatic strategy > grammatical pattern

The ultimate change of the basic grammatical order to OV in a dialect such as J. Urmi would be the completion of a pathway of grammaticalization of pragmatic strategies. This would be a syntactic correlate of the process of semantic change through grammaticalization proposed by scholars such as Traugott and König (1991) and Bybee (2010, 2015), whereby meaning changes by the grammaticalization of features in the pragmatic context of use. Already Meillet (1921, pp. 147–148) referred to grammaticalization not only of lexical items but also of the shift from Latin "free" word order to French "fixed" word order. Givón (1979, pp. 207–33) argues that language change may involve the shift from a pragmatic mode of communication (e.g. topic – comment) to a syntactic mode (subject – predicate) by a process he calls syntacticization. Lehmann (2008) discusses the grammaticalization of information structure, whereby "pragmatic relations lose their specificity". In the case of word order change in NENA, the grammaticalization of information structure has been contact-induced, just as grammaticalization of lexical items may be contact-induced (Heine & Kuteva, 2002, 2003).

When OV becomes the basic grammatical order, as in J. Urmi, the preverbal field cannot be used as the locus of an effective pragmatic strategy. Instead use is made of the postverbal field for pragmatic strategies that are performed by object preposing in dialects with a basic VO order. In dialects with a basic VO order, for example, a topical referent is fronted to perform the pragmatic strategy of expressing the close cohesion of the event or situation with what precedes.

(20)   C. Barwar   (Khan, 2008b, p. 874)
   ʾu-ʃwíqle   zúze   díye táma zìlle.| *θéle   xa-xèna,*| *ʾan-zúze*
   and left.he money his   there went.he came.3 MS another   **DEM-money**
   ʃqil-í-le **(OV)**   *ʾu zìlle.*|
   take.PST-ABS.PL.-ERG.3MS and-went.he
   'He left his money there and went off. Another man came, **took the money** and went off.'

This pragmatic strategy of object fronting can be identified also in dialects that have a frequent OV order, but a basic grammatical order of VO, such as C. Urmi:

(21)   C. Urmi   (Khan, 2016, vol. 2, p. 335)
márrə   xa náʃa   ʾáttən   ɟu-⁺dùssa̱*ḵ*| *bədmáyələ*   ⁺*àlli.*| … *ʾo-náʃa*
said.he   a-man   there.is   in-prison   resembles.he   to.me   **DEM-man**
⁺*bəḵṯàl-u=lə.*| (OV)
kill.PROG-OBJ.3ms=COP.3MS
'He said "There is a man in prison who resembles me." … **He kills this man**.'

In dialects with a basic, grammatical order of OV, such as J. Urmi, on the other
hand, the same function is expressed by postposing the topical object after the
verb, e.g.

(22)   J. Urmi   (Khan, 2008a: 165)
əl-⁺hudaé   ⁺rába ⁺rába ⁺mjizìlu.| ⁺***rába   ⁺talàn***
OBJ-Jews   much much harassed.PST-ABS.3PL.-ERG.3PL (ov)   **much   plunder**
wádlu   *l-⁺hudaé.*| **(VO)**
do.PST-ERG.3PL OBJ-Jews
'They harassed the Jews a great deal. **They plundered the Jews a lot**.'

The use in J. Urmi of VO to express the pragmatic strategy that is expressed by
OV in C. Barwar and C. Urmi is diagnostic of basic word order change in J. Urmi
to OV. It distinguishes dialects where OV is basic from those in which, although
it is frequent, it is still a pragmatic strategy (e.g. C. Urmi). The strategy is to use a
structure that is chiastic to the basic grammatical word order of the language sys-
tem, even if this word order does not occur in the immediately preceding clause.

5.   Imitation of morphology

A number of contact-induced changes in NENA dialects have resulted in morpho-
logical forms that are identical to, or at least very similar to, the phonetic shape of
corresponding forms in the contact languages. An example of phonetic imitation
can be found in the development of a preverbal particle in present indicative verbs
in NENA dialects. This is a morphosyntactic construction that is found in most of
the contact languages of the area. What is of particular significance is that there is
convergence across the languages not only in morphosyntactic pattern but also in
the phonetic shape of the preverbal particle, as can be seen in (24):

(23)   NENA (Txuma)   *ki-ʃate*   'he drinks'
       Arabic (Azəx):   *kū-nəktəb*   'we are writing'
       Armenian (Muš)   *kə-sirim*   'I love'

In all these languages the particle has a similar phonetic shape but each has developed using diverse types of morphological material that is internal to each language: NENA *ki-* < **kā* 'here' + 3ms copula, Arabic *kū-* < demonstrative *k* + 3ms pronoun *uwe*, Armenian *kə-* < deictic particle 'behold' (Makaev, 1977).

Such cases of resemblance in phonetic shape of morphemes across languages in contact may be compared to the phenomenon of "homophonous diamorphs" that have been observed to occur in code-switching between genetically related languages (Muysken, 2000, pp. 133, 149). Clyne (1967) has suggested that the distinction between two codes may be neutralized at the point where they share a pair of homophonous diamorphs, since it is hard to classify such elements unambiguously in terms of either of the two codes. Code-switching between genetically related languages and, consequently, typically exhibiting a high degree of structural equivalence, referred to by Muysken (2000) as "congruent lexicalization", involves switches of all lexical categories, including function words, and morphological integration (Law, 2017). In the case of the phenomenon of imitation in NENA, the languages are not necessarily genetically related nor is there a historically inherited homophony across morphemes, but rather homophonous diamorphs and structural equivalence are created by a process of contact-induced change. The phenomenon is an adaptive mechanism (for the concept see Farrar & Jones, 2002, p. 12) involving levelling of surface phonetic shape but retention of distinct underlying morphological representations. In the J. NENA dialect there was no extensive codeswitching, but there was massive lexical replacement. It would appear, therefore, that codeswitching is not a necessary condition for the creation of such diamorphs.

6. Reflection of change in contact language

The final aspect of contact and change that I would like to draw attention to is illustrated in (24):

(24) a. Bahdini Kurmanji   (MacKenzie, 1961, pp. 210–211)
          yê   hatî
          EZ.MS come.PST.PTCP
          'He has come.'
      b. J. Betanure NENA   (Mutzafi, 2008, p. 79)
          ʾile   ʾəθya
          COP.3MS come.PST.PTCP.MS
          'he has come'

In the Bahdini dialects of Kurmanji Kurdish a relative/attributive particle, known as *ezafe*, underwent change to a copula, via a cleft construction (Haig, 2011). Different dialects of Bahdini spoken across northern Iraq and south-eastern Turkey exhibit a variety of degrees of transitional development, to the extent that the grammatical category of the *ezafe* particle is often not transparent. The fact that NENA dialects in a particular area replicate such *ezafe* constructions with an unambiguous NENA copula (as in 24b) can be taken as evidence of the underlying morphosyntactic structure of the construction in the contact language, indicating that the change has reached completion.

7.   Summary

In this paper I have presented an overview of some aspects of contact and change in the NENA dialects. A summary of the main points is as follows:

–   Various constraints on contact-induced change may be identified, including socially-motivated resistance to lexical borrowing, factors blocking diffusion such as the large size of communities and loose networks, prevention of homophony in morphological systems, and the existence of structural similarities between forms in the NENA dialects and corresponding forms in contact languages.
–   Contact-induced change may have an internal systemic motivation. If functionally significant morphological distinctions have been lost by internal developments in a NENA dialect, then borrowing from contact languages may be motivated to take place in order to compensate for this.
–   Contact can increase the complexity of resources, which are put to use to reduce the complexity of linguistic systems.
–   Historical change in NENA has advanced at different rates across the dialects. Different dialects exhibit different degrees of convergence. In many cases the convergence is partial. Hierarchies of features can be identified with regard to their relative tendencies to converge with contact languages.
–   The outcome of contact-induced change can be a form that is identical to or closely resembles the phonetic shape of the corresponding form in the contact language.
–   The replication by NENA of constructions in contact languages that are not transparent may cast light on the underlying structure of such constructions in the contact language.

Abbreviations

ABS	absolutive	PL	plural
AUX	auxiliary	PROG	progressive
COP	copula	PRS	present
DEM	demonstrative	PST	past
ERG	ergative	PTCP	participle
F	feminine	S	single argument of canonical
IRR	irrealis		intransitive verb
MS	masculine	SG	singular
NOM	nominative	1	first person
OBJ	object	2	second person
OBL	oblique	3	third person

References

Aikhenvald, A. Y. (2003). *Language contact in Amazonia*. Oxford: Oxford University Press.

Anderson, G. D. S. (2011). Auxiliary verb constructions (and other complex predicate types): A Functional–Constructional overview. *Language and Linguistics Compass*, 5(11), 795–828. https://doi.org/10.1111/j.1749-818X.2011.00311.x

Bybee, J. L. (2010). *Language usage and cognition*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511750526

Bybee, J. L. (2015). *Language change*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139096768

Chyet, M. L. (1995). Neo-Aramaic and Kurdish: An interdisciplinary consideration of their influence on each other. *Israel Oriental Studies*, 15, 219–249.

Clyne, M. (1967). *Transference and triggering*. The Hague: Nijhoff.

Clyne, M. (1992). Linguistic and sociolinguistic aspects of language contact, maintenance and loss: Towards a multifacet theory. In W. Fase, K. Jaspaert, & S. Kroon (Eds.), *Maintenance and loss of minority languages* (pp. 17–36). Amsterdam: Benjamins. https://doi.org/10.1075/sibil.1.04cly

Coghill, E. (2016). *The rise and fall of ergativitiy in Aramaic: Cycles of Alignment Change*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198723806.001.0001

Dahl, Ö. (2004). *The growth and maintenance of linguistic complexity*. Amsterdam: Benjamins. https://doi.org/10.1075/slcs.71

Dickey, S. M. (2011). The varying role of po- in the grammaticalization of Slavic aspectual systems: Sequences of events, delimitatives, and German language contact. *Journal of Slavic Linguistics*, 19(2), 75–230. https://doi.org/10.1353/jsl.2011.0009

Dryer, M. S. (2009). The branching direction theory of word order correlations revisited. In S. Scalise, E. Magni, & A. Bisetto (Eds.), *Universals of Language Today* (pp. 185–207). Berlin: Springer. https://doi.org/10.1007/978-1-4020-8825-4_10

Enrique-Arias, A. (2010). On language contact as an inhibitor of language change: the Spanish of Catalan bilinguals in Majorca. In A. Breitbarth, C. Lucas, S. Watts, & D. Willis (Eds.), *Continuity and change in grammar*. (pp. 97–118). Amsterdam: Benjamins. https://doi.org/10.1075/la.159.05enr

Epps, P. & Stenzel, K. (2013). Introduction. In P. Epps & K. Stenzel (Eds.), *Upper Rio Negro: Cultural and linguistic interaction in northwestern Amazonia* (pp. 13–52). Rio de Janeiro: Museu Nacional Museu do Índio – Funai.

Farrar, K. & Jones, M. C. (2002). Introduction. In M. C. Jones & E. Esch (Eds.), *Language change: The interplay of internal, external and extra-linguistic factors* (pp. 1–18). Berlin-New York: De Gruyter Mouton. https://doi.org/10.1515/9783110892598.1

Floyd, S. (2013). Semantic transparency and cultural calquing in the northwest Amazon. In P. Epps & K. Stenzel (Eds.), *Upper Rio Negro: cultural and linguistic interaction in northwestern Amazonia* (pp. 271–306). Rio de Janeiro: Museu Nacional Museu do Índio – Funai.

Gaunt, D. (2006). *Massacres, resistance, protectors: Muslim–Christian relations in Eastern Anatolia during World War I*. Piscataway, NJ: Gorgias Press.

Givón, T. (1979). *On understanding grammar*. New York: Academic Press.

Greenberg, J. H. (1966). *Language universals with special reference to feature hierarchies*. The Hague: Mouton.

Haig, G. (2011). Linker, relativizer, nominalizer, tense-particle. On the Ezafe in West Iranian. In F. H. Yap, K. Grunaw-Hårsta, & J. Wrona (Eds.), *Nominalization in Asian languages: Diachronic and typological perspectives* (pp. 363–390). Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.96.13hai

Heine, B. & Kuteva, T. (2002). *World lexicon of grammaticalization*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511613463

Heine, B. & Kuteva, T. (2003). On contact-induced grammaticalization. *Studies in Language*, 27, 529–572. https://doi.org/10.1075/sl.27.3.04hei

Jastrow, O. (1978). *Die mesopotamisch-arabischen qəltu-Dialekte* (Vol. 1). Wiesbaden: Harrassowitz.

Jastrow, O. (1979). Zur arabischen Mundart von Mossul. *Zeitschrift Für Arabische Linguistik*, 2, 36–75.

Jastrow, O. (2011). Ṭuroyo and Mlaḥsô. In S. Weninger, G. Khan, M. Streck, & J. Watson (Eds.), *The Semitic Languages: An International Handbook* (pp. 697–707). Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110251586.697

Jastrow, O. (2015). Language contact as reflected in the consonant system of Ṭuroyo. In A. M. Butts (Ed.), *Semitic languages in contact*. Leiden-Boston: Brill. https://doi.org/10.1163/9789004300156_013

Jügel, T. (2015). *Die Entwicklung der Ergativkonstruktion im Alt- und Mitteliranischen – Eine korpusbasierte Untersuchung zu Kasus, Kongruenz und Satzbau*. Wiesbaden: Harrassowitz.

Khan, G. (forthcoming). The diachronic development of the copula in Neo-Aramaic dialects.

Khan, G. (2001). Quelques aspects de l'expression d' "être" en néo-araméen. In A. Donabédian (Ed.), *Langues de Diaspora. Langues en Contact* (pp. 139–148). Paris: Ophrys.

Khan, G. (2002a). *The Neo-Aramaic dialect of Qaraqosh*. Boston, MA: Brill. https://doi.org/10.1163/9789004348585

Khan, G. (2002b). The Neo-Aramaic dialect of the Jews of Rustaqa. In W. Arnold & H. Bobzin (Eds.), *"Sprich doch mit deinen Knechten aramäisch, wir verstehen es!" 60 Beiträge zur Semitistik, Festschrift für Otto Jastrow zum 60. Geburtstag* (pp. 395–410). Wiesbaden: Harrassowitz.

Khan, G. (2007). The North-Eastern Neo-Aramaic Dialects. *Journal of Semitic Studies*, 52(1), 1–20. https://doi.org/10.1093/jss/fgl034

Khan, G. (2008a). *The Jewish Neo-Aramaic dialect of Urmi*. Piscataway: Gorgias.

Khan, G. (2008b). *The Neo-Aramaic dialect of Barwar* (Vols. 1–3). Leiden: Brill. https://doi.org/10.1163/ej.9789004167650.i-2198

Khan, G. (2009). *The Jewish Neo-Aramaic dialect of Sanandaj*. Piscataway, NJ: Gorgias Press.

Khan, G. (2011). North-Eastern Neo-Aramaic. In S. Weninger, G. Khan, M. Streck, & J. Watson (Eds.), *The Semitic languages: An internatioanl handbook* (pp. 708–724). Berlin-Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110251586.708

Khan, G. (2012). Remarks on the historical development of the copula in Neo-Aramaic. In F. Corriente, G. del Olmo Lete, Á. Vicente, & J.-P. Vita (Eds.), *Dialectology of The Semitic Languages Proceedings of the Iv Meeting on Comparative Semitics Zaragoza 11/6–9/2010* (pp. 25–31). Sabadell: AUSA.

Khan, G. (2016). *The Neo-Aramaic dialect of the Assyrian Christians of Urmi* (Vols. 1–4). Leiden-Boston: Brill. https://doi.org/10.1163/9789004313934

Khan, G. (2017). Ergativity in Neo-Aramaic. In J. Coon, D. Massam, & L. Travis (Eds.), *Oxford Handbook of Ergativity* (pp. 873–899). Oxford: Oxford University Press.

Law, D. (2017). Language mixing and genetic similarity. *Diachronica*, 34, 40–78. https://doi.org/10.1075/dia.34.1.02law

Lehmann, C. (2008). Information structure and grammaticalization. In E. Seoane & M. J. López-Couso (Eds.), *Theoretical and Empirical Issues in Grammaticalization* (pp. 207–229). Amsterdam: Benjamins. https://doi.org/10.1075/tsl.77.12leh

MacKenzie, D. N. (1961). *Kurdish dialect studies*. London, New York: Oxford University Press.

Makaev, E. A. (1977). The origins of the modal particle *ke* in Armenian. In N.Ya Gabaraev (Ed.), *Questions of Iranian and General Philology* (pp. 195–200). Tbilisi: Metsniereba.

Malkiel, Y. (1968). The inflectional paradigm as an occasional determinant of sound change. In W. P. Lehmann & Y. Malkiel (Eds.), *Directions for historical linguist* (pp. 20–64). Austin: University of Texas Press.

Malkiel, Y. (1976). Multi-conditioned sound change and the impact of morphology on phonology. *Language*, 52, 757–778. https://doi.org/10.2307/413294

Matras, Y. & Sakel, J. (2007). Investigating the mechanisms of pattern replication in language convergence. *Studies in Language*, 31(4), 829–865. https://doi.org/10.1075/sl.31.4.05mat

Meillet, A. (1921). *Linguistique historique et linguistique générale*. Paris: Champion.

Milroy, L. (1987). *Language and social networks*. Oxford: Blackwell.

Mutzafi, H. (2008). *The Jewish Neo-Aramaic dialect of Betanure (province of Dihok)*. Wiesbaden: Harrassowitz.

Muysken, P. (2000). *Bilingual speech: A typology of code-mixing*. Cambridge: Cambridge University Press.

Myers-Scotton, C. (1993). *Duelling languages: Grammatical structure in code-switching*. Oxford: Clarendon.

Nettle, D. (1999). *Linguistic diversity*. Oxford: Oxford University Press.

Silverstein, M. (1981). The limits of awareness. In *Working Papers in Sociolinguistics* (Vol. 84). Austin: Southwest Educational Development Laboratory.

Talay, S. (2008a). *Die Neuaramaischen Dialekte Der Khabur-Assyrer in Nordostsyrien: Einführung, Phonologie und Morphologie*. Wiesbaden: Harrassowitz.

Talay, S. (2008b). The neo-Aramaic dialects of the Tiyari Assyrians in Syria: with special consideration of their phonological characteristics. In G. Khan (Ed.), *Neo-Aramaic Dialect Studies* (pp. 39–63). Piscataway: Gorgias Press.

Thackston, W. M. (2006a). *Kurmanji Kurdish. A reference grammar with selected readings*. Renas Media. Retrieved from http://www.fas.harvard.edu/~iranian/Kurmanji/index.html

Thackston, W. M. (2006b). Sorani Kurdish: A Reference Grammar with Selected Readings. Retrieved from internal-pdf://Sorani-Grammar-2192906496/Sorani-Grammar.pdf

Traugott, E. C. & König, E. (1991). The semantics-pragmatics of grammaticalization revisited. In E. C. Traugott & B. Heine (Eds.), *Approaches to Grammaticalization* (Vol. 1, pp. 189–218). Amsterdam: Benjamins. https://doi.org/10.1075/tsl.19.1.10clo

Trudgill, P. (1986). *Dialects in contact*. Oxford: Blackwell.

Trudgill, P. (1989). Contact and isolation in linguistic change. In L. E. Breivik & E. H. Jahr (Eds.), *Language change: contributions to the study of its causes* (pp. 227–238). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110853063.227

Trudgill, P. (2009). Sociolinguistic typology and complexification. In G. Sampson, D. Gil, & P. Trudgill (Eds.), *Language complexity as an evolving variable* (pp. 98–109). Oxford: Oxford University Press.

Copying of argument structure

A gap in borrowing scales and a new approach to model contact-induced change

Carola Trips
University of Mannheim

This article deals with the copying of argument structure, pursuing two aims: first, I will discuss why the copying of verbs, and more specifically their argument structure, has never been truly dealt with in models of language contact, including borrowing scales and hierarchies. I will show that the reason lies in the use of the long-standing dichotomy between lexical and structural borrowing, which should be rethought. Second, I will propose an alternative approach using Johanson's integrative approach to language contact and code copying (2002, 2008), Holler's (2015) definition of integration conflicts on the level of argument structure, and a modified version of Myers-Scotton's (2002) Abstract Level Model. I will apply my approach to a qualitative empirical corpus study of Old French psych verbs copied to Middle English, focusing on the description, analysis, and theoretical modelling of the integration conflict that arises with the EXPERIENCER argument syntactically expressed by a *to*-PP on the model of Old French. Finally, I will compare my findings with similar findings from studies of language acquisition and suggest some generalisations.

Keywords: Middle English, Old French, language contact, argument structure, copying, integration conflicts

1.   Introduction

This article deals with the copying of argument structure and especially focuses on the English-French contact situation in medieval times, triggered by the Norman Conquest of 1066 and lasting more than three hundred years. The study presented here is part of the project *Borrowing of Argument Structure in Contact Situations* (BASICS)[1] and seeks to answer the question of how much the argument structure

of English verbs was affected by the copying of Old French verbs. It fills a gap in the vast research area of lexical and structural borrowing (see for example Prins, 1948; Orr, 1962; Rothwell, 1976, 1980; Matras, 2009) as it specifically investigates the nature of the copying of argument structure and concomitant possible syntactic changes in the grammatical system of English induced by French verbs.

Surprisingly, the copying of argument structure is a neglected field of study: if it is studied at all, it only ever marginally occurs in models of language contact and borrowing hierarchies. The long-standing traditional dichotomy between lexical and structural borrowing may be seen as a reason for this gap. Winford (2003, p. 61), for example, states that "[s]ome scholars, for example, Thomason and Kaufman (1988), distinguish between "lexical" and "structural" borrowing as though the two could in fact proceed independently of each other". Many other studies based on this allegedly clear-cut distinction are not hard to find (for an overview of the literature on borrowing, see for example Matras, 2009). It is evident, however, that verbs comprise both lexical and structural aspects that could basically be subject to processes of copying. In addition to their lexical properties, they define semantic and syntactic aspects of their arguments which shape the structure of a sentence. Thus, to fully account for processes of copying, it is not helpful to talk about "either lexical or structural". This is why I suggest that the terms lexical and structural borrowing should be abandoned and be replaced by a more adequate description along the lines of global copying in the sense of Johanson (2002, 2008). This approach serves well to define instances of the copying of argument structure and especially to identify mismatches or integration conflicts which reveal relevant aspects of contact-induced change.

The article is structured as follows: in Section 2, the state of the art of borrowing will briefly be discussed, focusing on borrowing hierarchies and the gap they display when it comes to verbs (and their argument structure). Section 3 introduces the concept of copying, an alternative term for borrowing, and defines the historical contact situation based on Johanson's work. In Section 4, I discuss criteria to define instances of verb integration and integration conflicts by taking a closer look at the studies of Eisenberg (2001, 2011), and Holler (2015) for Standard German. Section 5 combines the assumptions of Johanson, Eisenberg, and Holler and applies them to historical data; here I will focus on identifying instances of integration and integration conflicts in the class of Middle English psych verbs. In Section 6, I will sketch a multi-layered model similar to the one Myers-Scotton proposes and provide an analysis of my empirical findings. In Section 7, I will summarise my findings and complement them with insights from studies of language acquisition.

2.   Borrowing hierarchies and the place of verbs

As noted in the introduction, one reason why the copying of argument structure is a phenomenon that has not received sufficient attention is that the dichotomy between lexical and structural borrowing does not leave room for entities that exhibit both lexical and structural properties. Another reason might be that, as early as 1978, Moravcsik postulated a number of implicational typological constraints, one of which stated that "[a] lexical item whose meaning is verbal can never be included in the set of borrowed properties. This statement excludes a language that has borrowed the symbolic association of verb form and verbal meaning." The author based this constraint on data from the New Guinea language Enga, which borrows verbs by systematically adding indigenous verbs to it (p. 111). For example, the English verb *lose* is borrowed as *lúsa* and must be followed by the verb form *lengé* (*lúsa lengé* 'lose utter') which only ever occurs with nominal complements and never with verbal complements. So verbs can only be borrowed with a non-verbal, nominal meaning. In a similar vein, Hock & Joseph (1996, p. 257) claim that

> [a]lthough verbs are borrowed more easily than basic vocabulary, they nevertheless are not as readily borrowed as nouns. And if the need for borrowing a verb does arise, many languages instead borrow a nominal form of the verb and employ a native all-purpose verb such as *do* or *make* as a means of turning that form into the equivalent of a verb.

One of the most frequently cited works in the field of language contact is Thomason & Kaufman's (1988) monograph, which was the first to deal with language contact in the historical dimension in a systematic fashion. Unlike other borrowing hierarchies that suggest an implicational relationship between the borrowing of individual categories (cf. Matras, 2007, p. 153), Thomason & Kaufman claim that the degree of intensity of contact can be related to the likelihood of structural borrowing (see Figure 1).

Thomason and Kaufman's (1988) borrowing scale

Casual contact   Category 1: content words

Category 2: function words, minor phonological
                  features, lexical semantic features

Category 3: adpositions, derivational suffixes,
                  phonemes

Category 4: word order, distinctive features in
                  phonology, inflectional morphology

Intense Contact   Category 5: significant typological disruption,
                  phonetic changes

Figure 1. (From Matras, 2009, p. 156)

Scrutinising their borrowing scale reveals that although they thoroughly define five stages going from casual contact to heavy structural borrowing by giving very detailed information as to what is likely to be borrowed (see the original version in Thomason and Kaufman, 1988, p. 74–76), the authors do not mention the borrowing of the argument structure of verbs at all. However, if our survey is extended to other borrowing scales or hierarchies that have been proposed in the literature, we find a slightly different picture. The hierarchies suggested, for example, by Haugen (1969), Muysken (1981, p. 62), Appel and Muysken (1987, p. 170f), and Matras (2007), which predict the borrowability of word-classes on the basis of frequency, include verbs and state that verbs are less likely to be borrowed than nouns but more likely than, for example, subordinating conjunctions or inflectional affixes (see, for example, Muysken, 1981 and Matras, 2007). Field (2002) proposes a general hierarchy which is said to capture the likelihood of borrowing for relevant categories:

(1) content item (nouns > adjectives, verbs) > function word >
     agglutinating affix > fusional affix

Implicational hierarchies of this kind suggest a constraint on borrowing of lower-ranking categories (e.g. adjectives and verbs are not borrowed unless nouns are borrowed as well). Overall, however, we see in all of these papers that the distinction between lexical and structural borrowing is always implicit, and, moreover, it is not explicitly defined what verb borrowing actually means.

An exception may be seen in Winford's (2003) monograph on contact linguistics. Referring to these hierarchies the author comments on the reasons why some categories are more easily borrowed than others and postulates a number of constraints inhibiting or favouring the borrowing of verbs (p. 51ff). First, he mentions syntagmatic constraints, which relate to the morphological and syntactic properties of lexical categories and can be seen as an inhibiting factor. Since verbs govern other categories and assign case, they are less likely to be borrowed. Second, he sees the degree of morphological complexity in the paradigms of a lexical category as an inhibiting factor: verbs tend to be morphologically complex and central to the syntax of the sentence and are therefore less likely to be borrowed. However, close typological similarity in verbal structure can be seen as a favouring factor between the languages in contact.

Most of the time, when authors refer to verbs, they talk about the formal adaptation and accommodation of these items (which relates to Winford's second factor). This is, for example, implicit in one of the statements by Matras, when he notes that "… there appears to be a near-consensus view that the borrowing of verbs is made more cumbersome in some languages due to the widespread tendency of verbs to be morphologically more complex" (Matras, 2009: 175). In a recent

monograph on the borrowing of verbs by Wohlgemuth (2009), the focus is exactly on these formal aspects. The author suggests a loan verb integration hierarchy which is ordered according to the integrational effort involved in (formally) accommodating a borrowed verb. For him, the Light Verb Strategy requires the most integrational effort and Direct Insertion the least integrational effort.

Light Verb Strategy < Indirect Insertion < Direct Insertion
(< Paradigm Insertion)

Figure 2. (Wohlgemuth, 2009, p. 285)

Wohlgemuth also briefly comments on the role valency plays in the process of verbal borrowing and claims that it is of minor relevance cross-linguistically because it only governs some instances in a few languages and cannot be seen as an overarching principle. Concerning the role the donor and recipient language play in the process, Wohlgemuth states that the former is fairly irrelevant, although speakers who borrow a lexical item must have some competence in the donor language. In the following quote, the dichotomy between lexical and structural borrowing is, once again, prominent:

> Yet, for lexical borrowing, this competence needs [sic] not exceed the (sometimes rather fuzzy and fragmentary) knowledge of some vocabulary items and at least one of their lexical meanings. It is only for grammatical borrowing that the speakers borrowing a form need to have an understanding of the borrowed form's grammatical function in the donor language.   (Wohlgemuth, 2009, p. 272)

If we extend the survey to sociolinguistic work on (synchronic) bilingual code-mixing, we find a similar picture. In her recent monograph on borrowing in code-switching, Poplack (2018, p. 5–6) investigates language mixing of "overt lexical material from two or more languages".[2] She focuses on the process of integration, and hence how speakers adapt foreign items to the phonological, morphological, and syntactic patterns of their recipient language (p. 8).[3] She further asks whether a distinction should be made between borrowing and code-switching. Again, the dichotomy between lexical and structural borrowing is her basis; possible instances of non-integration are not addressed at all.

---

2. Poplack (2018) does not consider structural borrowing, semantic extensions, or other contact-induced changes which have no overt lexical footprint.

3. Another interesting sociolinguistic study is King (2000) investigating Acadian French. Although she addresses the copying of lexical verbs, she does not deal explicitly with their argument structure.

To summarise this survey of borrowing hierarchies and scales as well as papers on this topic, I identified the following main points:

- the dichotomy between lexical and structural borrowing is ubiquitous;
- in this dichotomy, the borrowing of verbs is predominantly classified as lexical borrowing;
- verbs are, thus, mainly treated as lexical entities;
- some of the hierarchies we discussed include verbs, some do not;
- their morphological shape seems to be especially relevant.

To fill the gap in borrowing scales and, in more general terms, research on borrowing, we need to define the borrowing of verbs in terms of argument structure (Comrie, 1993; Levin, 2018), so that we then can adequately model it theoretically. Further, we need a terminology which is not based on the traditional dichotomy but instead uses concepts which include both lexical and structural properties. We also need a method with well-defined criteria to identify degrees of integration and deviation. And finally, we need to apply a theory that accounts for the findings in an adequate way, i.e. an abstract model which has the potential to generalise findings and predict changes in argument structure.

There is one model which comes quite close to fulfilling the criteria listed above and which I have not included in my survey as it lies outside of the scope of the field of historical language contact. It is Myers-Scotton's Abstract Level model (Myers-Scotton & Jake, 1995; Myers-Scotton's 2002) which she developed to account for effects of bilingualism (code-switching, borrowing, convergence, attrition, formation of creoles) from a psycholinguistic perspective. The model is based on premises about the structure of the mental lexicon. The major premise is that all lemmas in the mental lexicon include three levels of abstract lexical structure: (i) lexical-conceptual structure (semantic and pragmatic information), (ii) predicate-argument structure (the mapping of thematic structure onto syntactic relations), (iii) morphological realization patterns (surface realizations of grammatical information; cf. Myers-Scotton, 2002, p. 19). Myers-Scotton assumes that language contact can affect all three levels to different degrees, which results in different forms of language mixing. One such result is the splitting and recombining of information on the level of predicate-argument structure. Example (2) illustrates convergence of English to Hungarian:

(2) a. Hungarian of English-Hungarian bilingual child
        mert te nem kel-sz   en-gem fel
        because you not wake.INTRANS.PRES.2S I.ACC up

'Because you do not wake me up.'

b. *Standard Hungarian:*
mert te nem kel-t-esz en-gem fel
because you not wake.TRANS.PRES.2SG I.ACC up
(Bolonyai 1999, p. 101 cited in Myers-Scotton, 2002, p. 204)

In English *wake up* can either be a transitive or intransitive verb depending on the presence or absence of an object complement. Causativity is part of the transitive verb. In Standard Hungarian, causativity is expressed by the suffix, *-t* which is added to the verb *kel* 'wake, rise'. The causative verb *kelt* subcategorizes for a direct object while *kel* does not. The example in (2) shows that the bilingual child uses the verb *kel* with a causative meaning but without the causative suffix *-t*. Convergence on all three levels of abstract lexical structure can be observed here: the specifications for predicate argument structure are altered for *kel*, the distinction between causativity and non-causativity is neutralised, and the morphological realisation of causativity is changed.

Before I apply a slightly modified version of this model to cases of copied argument structure, I will define the contact situation and the properties of copied verbs following Johanson's work.

3. Historical language contact and the copying of verbs

Johanson proposes an integrative approach to language contact which aims at dealing especially with contact-induced change within a code-copying framework. Generally, he assumes that language contact situations have to be seen as an asymmetrical dominance relation between a code A, the sociolinguistically dominated (or weak) code, and a code B, the sociolinguistically dominant and prestigious (or strong) code. Most frequently, more prestigious languages (foreign Model Codes) influence less prestigious languages (Basic Codes). If we apply these assumptions to the OF-ME contact situation we must make a distinction between two different stages: In the stage from ca. 1066 to the early 13th century, Old French (including Anglo-Norman[4] (AN)) clearly functioned as the (foreign) Model Code since OF was the sociolinguistically dominant and prestigious code that influenced ME, the less prestigious and sociolinguistically dominated Basic Code. This situation changed when at the beginning of the 12th century the language of instruction in school became AN. Ingham (2012: 9) proposes that "the period of greatest contact

---

4. This variety of French was used in England after the Norman Conquest for some time. For further information, see the Anglo-Norman Hub.

influence with English occurred when fluent bilingualism would have been commonplace among educated speakers, in the C13 and the first half of the C14". Thus, we see changes in the relation between the Basic Code and the Model Code which finally resulted in ME being the more dominant language again at the end of 15th century (cf. Trips and Stein (2019)).

Johanson uses the term *copying* as a cover term for borrowing, diffusion, transfer, replication and the like to stress that the copied material is never identical to its model, i.e. it is never a true replica. The reason for this is that the copies are always adapted to the Basic Code system, i.e. phonologically, lexico-semantically and morpho-syntactically modified in the direction of the Basic Code. In the process of insertional copying, elements from the Model Code are inserted into clauses of the Basic Code. The basis for inserting copies are structural and conceptual similarities between the two codes. This means that speakers, consciously or unconsciously, establish equivalence relations between elements of the Model Code and Basic Code (Johanson, 2008, p. 63). As soon as elements from the Model Code have undergone insertion, they are part of the Basic Code and subject to its internal processes.

In his framework, Johanson introduces the notion of global copying. Global copying happens when a unit (morpheme, morpheme sequence) is copied from a Model Code as a whole, i.e. a block of different properties (material, semantic, combinational in syntax and word structure, frequential). These units have a material shape and can be simple or complex, bound or free, lexical or functional. The insertion of global copies is based on equivalence relations:

> […] global copies are inserted into those slots – equivalence positions, insertion points – that their "equivalents" in the basic code may fill. The decisive criterion of equivalence is the speaker's subjective assessment of what he or she feels to be close enough. Lack of real typological equivalence does not prevent insertion.
>
> (Johanson, 2002, p. 294)

This implies that copying begins as a momentary act in (the mind of) individuals and this presupposes some degree of knowledge of the Model Code. Code copies may become habitualised and thus have lasting effects in the code of the individual and they may also become conventionalised, i.e. having an effect on the code of speech communities. Johanson notes that this process can be described as a path from momentary copies initiated by the individual to the conventionalization of these copies in the speech community (Johanson, 2008, p. 65). Crucially, as soon as a copy has reached the status of being conventionalised, the degree of the speaker's knowledge of the Model Code becomes irrelevant. Possible deviations from the Basic Code are established as new sets of form and sometimes even equivalent forms in the Basic Code are replaced. If this happens, linguistic change occurs.

4. The copying of verbs and integration conflicts

In this section, the notions of integration and integration conflicts that arise in the copying of argument structure will be discussed. Although Johanson (2002: 299) mentions the integration of copied elements and states that they are part of the process of conventionalisation, his assumptions about argument structure are not explicit enough. In contrast, Eisenberg (2001, 2011) provides a clearer picture about integration and integration conflicts of copied elements, and Holler (2015)[5] convincingly applies his assumptions to foreign verbs copied to Standard German. This is why these studies will be briefly discussed here before we take a closer look at the copying of OF verbs in ME times.

Holler (2015) investigates the copying of foreign verbs to Standard German. To identify their foreignness, we need criteria which will allow us to define the special syntactic and semantic properties of non-native argument structure. Holler approaches this topic by adopting Eisenberg's (2001, 2011) symmetrical model of gradual integration. Eisenberg generally assumes that when foreign words are copied to Standard German, they become integrated into this system and are then part of the grammatical system of German. If this is the case, we expect them to follow the rules of this system. This does not mean, however, that they behave in the exact same way as the native words of the core system do (for a similar assumption see Johanson in the previous section). Rather, they build a subsystem with specific properties, i.e., they show properties of both the foreign Model Code and Basic Code. As a diagnostic for the degree of integration, Eisenberg proposes to use observable integration conflicts which manifest themselves in the variation of forms and structures or in situations where non-native structures only partially adhere to the rules of the native grammatical system.

Based on a study of forty non-native verbs which were first selected from the DUDEN-Fremdwörterbuch (2007) and then extracted from the *Deutsche Referenzkorpus* (DeReKo) and manually annotated for subcategorisation frames and semantic roles, Holler sought to determine the degree of integration of these verbs. To do so, she used Primus's (1999) Case Hierarchy (Nominative > Accusative > Dative > Prepositional case) and her Semantic Role Hierarchy (AGENT > EXPERIENCER/RECIPIENT/BENEFACTIVE > PATIENT) for Standard German (Primus, 2012) and established respective hierarchies for the system of the copied verbs. The main results are that, first, arguments in the dative are never realised and prepositional case is preferred instead. Further, non-native verbs deviate from

---

5. Holler frequently refers to a study by Holler & Scherer (2010) which discusses the original study of the authors. Holler (2015) is a slightly revised version of it which will be cited throughout this section unless it is necessary to cite the older study.

native verbs in not realizing EXPERIENCER, RECIPIENT and BENEFACTIVE to the same degree; LOCATIVE, SOURCE and GOAL are preferred instead. For example, the copied verb *mailen* 'to mail' prefers to express the RECIPIENT with a PP (*Sie mailt einen Brief an ihn*; 'she mails a letter to him') whereas the native equivalent of this verb prefers to express the RECIPIENT with a dative NP (*Sie sendet ihm.DAT einen Brief*; 'she send him a letter'). What follows from this is that in the argument structure of non-native verbs a dative NP is never linked to the roles of EXPERIENCER, RECIPIENT and BENEFACTIVE, in contrast to native verbs. This empirical fact can be interpreted as a conflict of integration which shows that non-native verbs build a subsystem with different rules: the ordering prepositional case-dative is reversed, dative is much less preferred, and LOCATIVE, SOURCE and GOAL are preferred over EXPERIENCER, RECIPIENT and BENEFACTIVE.

Overall, Holler's findings speak in favour of the assumption that copied verbs are (only) partially integrated into the system of the Basic Code. There is evidence that they behave like the native verbs to some degree but they also build their own particular argument structures. Thus, although they are integrated into the Basic Code, they build a subsystem.

Concerning the question of how this conflict can be explained, Holler briefly comments on the possibility that this might be due to the provenance of the Model Code ("Dies könnte beispielsweise durch die Gebersprache mit verursacht sein"; 2015, p. 412) but she does not further elaborate this point. Comparing (Present-Day) English with German, one notes the striking difference of morphological case marking. From this, we could assume that an integration conflict will arise when a language that morphologically marks semantic roles like EXPERIENCER, RECIPIENT and BENEFACTIVE with the dative case integrates verbs of a language that does not morphologically mark case. In the next section I will further investigate this generalisation by comparing the system of OF, which did not morphogically mark case and ME, which was gradually losing this property (cf. Allen, 1995). I will focus on the syntactic realisation of the EXPERIENCER argument of psych verbs and identify instances of integrated and non-integrated verbs copied from OF to ME on the model of Holler.

5. Integrated and non-integrated verbs copied from Old French to Middle English

In this section I will provide a qualitative empirical study of the copying of argument structure in Middle English. I will focus on identifying instances of integration and integration conflicts in the class of psych verbs, and more precisely, on possible differences in the syntactic realisation of the EXPERIENCER of amuse-type

verbs in Levin's terms (1993). First, I will take a look at one particular psych verb of this class (*plesen* 'please') and then extend my analysis to a larger set of this class.

The data presented here are predominantly taken from the lemmatised version of the syntactically annotated *Penn-Helsinki Parsed Corpus of Middle English2* (PPCME2, Kroch and Taylor, 2000).[6]

Based on the lemmatisation of the corpus and an advanced search in the *Oxford English Dictionary Online*[7] (OED), I identified the native and copied members of the amuse-type class in ME. To compare the argument structures of OF copied verbs and their OF equivalents, I consulted a number of sources for Old French and Anglo-Norman like the *Syntactic Reference Corpus of Medieval French* (SRCMF, Prévost and Stein, 2013, ca 266000 words), the corpus *Modéliser le changement : les voies du français* (MCVF, Martineau, 2009), the *Altfranzösisches Wörterbuch* (TL, Blumenthal and Stein, 2002), and the *Anglo-Norman Dictionary* (AND, Rothwell and Trotter, 2001).

Applying Johanson's approach, I assume that OF verbs copied to ME are instances of global copies bringing along their properties as a block. Some examples of amuse-type verbs are *abaishen* ('to embarass'), *anoien* ('to offend'), *greven* ('to vex'), *offenden* ('to cause offence') and *plesen* ('to please'). I further assume that speakers/writers copied OF verbs on the basis of equivalence relations and I interpret Johanson's comment on the subjective assessment of closeness between the copy of the foreign Model Code and the target of the copying in the Basic Code as the semantic closeness of verbs. That is to say, the copying of verbs is semantically driven, an actuation of the copying process (see also Section 6). This copied meaning is present in the OF original and corresponds to a syntactic realisation proper to OF. In this early period, in its innovative English context, the semantics of the copied verb is maintained and expressed by its frame of semantic roles.

Next, I will take a closer look at the argument structure(s) of the verb *plesen*. Table 1 provides a comparison of the structures of this verb in OF/AN and ME.

At first sight it seems that all of the OF/AN argument structures match the ME ones. However, if we bring in differences in the Basic Code and foreign Model Code (see Section 3) one clear integration conflict can be identified. It is the argument structure "NP[nom] + PP" (marked in bold in the table) where the EXPERIENCER argument is expressed by a PP. One possible OF/AN example is *Il plet a Dieu* which occurs in ME as *He pleases to God* instead of the expected *He pleases God*. To

---

Table 1. A comparison of syntactic constructions of OF *plaire* and ME *plesen* based on dictionaries

Syntactic construction	OF/AN *plaire/plaisir*	ME *plesen*
	T&L, AND	OED, MED
transitive:		
NP-SBJ-STIM. + NP-OBJ-EXP. intransitive + PP:	yes	yes
NP-SBJ-STIM. + PP-EXP. intransitive:	yes	yes
NP-SBJ-STIM. impersonal:	yes	(yes)
(NP[expl]) + NP/PP-EXP. + S-STIM. impersonal:	yes	(yes)
(NP[expl]) + NP/PP-EXP. + [*de*/*that* + S]-STIM.	yes	yes
reflexive: NP-SBJ-STIM. + RFL	yes	yes

mark the EXPERIENCER in the form of a PP is a property particular to French (and other Romance languages like Italian) which has been labelled as "prepositional dative" (cf. Troberg, 2008). This type of dative is also considered a structural dative (as opposed to an inherent dative; cf. Zaenen and Maling, 1990; Woolford, 2006) which is the result of the loss of case marking in the history of French. Although ME was in the course of losing case marking as well, marking the EXPERIENCER by the prepositional dative was never a result of this development. In this way, the OF and ME systems differ, and this fact can be seen in the integration conflict identified here (see also Trips and Stein (2019)).

  Some examples from the PPCME2 given in (3) highlight this integration conflict:

(3) a. *For God wasted þe bones   of hem þat plesen to men;*
          for God wasted the petitions of them that please to men
          'Because God destroyed the petitions of those who please men.'
                                                              (EARLPS,63.2771, M2)[8]
     b. *and eek a womman ne   myghte nat plese   to many folk at oones.*
          and also a woman   NEG might   not please to many folk at once
          'And also a woman might not please many people at the same time.'
                                                              (CTPARS,321.C1.1413, M3)

Looking at the range of the argument structures found for the verb *plesen* (in the OED, MED) given in (4), we see that the other OF structures are known and part of the native system:

---

8.   The corpus is divided into the following four subperiods: M1 (1150–1250), M2 (12501350), M3 (1350–1420), M4 (1420–1500).

(4) a. intransitive
 Thei that cunnen plese   and glose …
 they that know   please and obscure
 'Those who know how to please and obscure…'
  (c1393 Gower Confessio Amantis (Fairf.) vii. 2172 (MED))

  b. impersonal
 Me plesis   not at   nowder of þies sulde   be sent þis message
 me pleases not that none   of them should be send this message.
 'It doesn't please me that none of them should receive this message.'
  (c1450 Alphabet of Tales (1904) I. 72 (MED))

  c. transitive
 þese   wymmen … plesyd   god with lytyl penaunce.
 these women   pleased god with little penance.
 'These women … pleased God with little penance.'
  (a1400(c1303) R. Mannyng Handlyng Synne (Harl.) 2004 (MED))

It is not clear whether the intransitive and the transitive structures were copied as a foreign structure or whether the verb *plesen* was integrated into the transitive and intransitive structure of the native system. What we can see on the level of syntax is that they do not produce integration conflicts.

  In Old English (OE) psych verbs like *lician* 'please' the EXPERIENCER was morphologically marked by a dative (cf. Allen, 1995, p. 69; Fischer and van der Leek, 1983, p. 346, Denison, 1993, p. 62, Möhlig-Falke, 2012, p. 6ff):

(5) *& him   wel licode his wurðfulnesse þa.*
 and him.DAT wel liked his dignity.NOM then
 'and he liked well his dignity then.' (ÆLet_4_[SigeweardB]:67.8)

At the beginning of the ME period we can (still) find evidence for EXPERIENCER-datives:

(6) *þis is min   loue sune þet me   wel likeð.*
 this is mine love son   that me.DAT wel likes
 'This is my dear son that I well like.' (LAMB1,141.280)

According to Allen (1995, p. 213–217) the morphological realisation of dative case, i.e. the nominal dative suffix, was lost by the middle of the fourteenth century. However, the OF verbs that were copied to ME were subject to a system that generally did not mark semantic roles by morphologically case marked arguments. Instead, the role of EXPERIENCER (as well as RECIPIENT and BENEFACTIVE) could syntactically be realised as PP (see Togeby, 1983), as we have just seen. So the integration conflict which arises with OF verbs like *plesen* can be explained as a conflict between a foreign Model Code which lacks morphological case marking and a

Basic Code which (still) exhibits morphological case marking to some degree. The variation found can be interpreted as a manifestation of this conflict as Eisenberg proposed (see Section 4).

There are other OF copied psych verbs that show the same integration conflict in ME. One such case is *displesen* (OF *desplaire*, AN *desplaisir*):

(7) *And as cassidore seith 'A man dredeth to do outrages whan he woot and*
    and as C.   says 'A man dreads to do outrages when he knows and
    knoweth that it displeseth to the juges   and the sovereyns.'
    knows   that it displeases to the judges and the sovereigns.'
    'And as C. says: 'A man fears to indulge in excess when he is aware and knows
    that it displeases the judges and the sovereigns.'   (CTMELI,230.C1.498)

Moreover, this pattern sometimes also occurs with native psych verbs, for example *liken* in (8) and *quemen* in (9):

(8) *þet is þe zoþe uayrhede/ hueruore þe zaule to god likeþ/ and to*
    that is the true beauty   wherefore the soul to God likes and to
    þe angles þet yzeþ þe herte.
    the angles that see   the heart
    'This is the true spiritual beauty because of which the soul pleases God.'
                                                            (AYENBI,81.1576,M2)

    French original:
    C'est la veraie beauté par quoi l'ame plet a Dieu et es anges qui voient le cuer.
    this is the true beauty by which the soul pleases to God and to the angels
    who see the heart   (SOMME-ch47-par16)

(9) *and makeþ ofte   / lete þet guod to done: and do þet kuead / uor to*
    and makes often   let that good to do   and do that evil   for to
    kueme kueadliche to þe   wordle.
    please wickedly   to the world
    '… and do evil to wickedly please the world.'   (AYENBI,26.403,M2)
    French original:
    et fait mout de foiz lessier le bien a fere et fere le mal pour plere mauvesement
    au monde.
    and makes many of times let the good to do and do the evil to please wick-
    edly to+the world   (SOMME-ch32-par687)

Interestingly, these cases predominantly occur in ME texts which are based on a French original, for example the *Ayenbite of Inwyt*, (1340, Gradon 1965) which is a fairly direct translation of the *Somme le Roi* (1279, Laurent 2008).

6. Towards a new approach

In this section, I propose a way to model the integration conflict identified in the previous section, based on Myers-Scotton's Abstract Level model (see Section 2). I will focus on the levels of lexical-conceptual structure and predicate-argument structure and will ignore morphological realisation patterns. The level of conceptual structure will be modelled on Jackendoff's (1976, 1983, 1990) predicate decomposition approach; the definition of argument structure used here follows Levin (2018).

As noted above, the first aspect of the process of copying is semantic actuation, which was defined in Section 5. When the OF/AN verb *plaire/plaisir* was copied, it came into ME as a global copy, bringing along its material, semantic, syntactic, (and frequential) properties. I further assume that speakers/writers initiated the copying process on semantic grounds: they felt that the verb could well be inserted in an "equivalence position" in their Basic Code and they added the verb to their semantic class of amuse-type verbs which contained native verbs like *liken* and *quemen*. Based on Jackendoff's approach, I assume the following Lexical Conceptual Structures (LCS) for these three ME psych verbs:

(10) a. Y likes X. $[_{STATE} REACT^+ ([Y], [X])]$
   b. X quemes Y. $[_{STATE} BE_{aff+} ([X], [Y])]$
   c. X pleses Y. $[_{STATE} BE_{aff+} ([X], [Y])]$

However, it was not only the lexical semantics that played a decisive role in the copying process but also, as Allen (1995) has suggested, the predicate-argument structure of the verb. Evidence for this assumption comes from her small-scale study of these three verbs. Although all of them are generally translated with 'please', there are some crucial differences to be found on the level of predicate-argument structure, especially between *liken* and *quemen*: although both verbs showed the case marking frame EXPERIENCER-dative THEME-nominative (in OE and partly in ME times), the THEME argument of *quemen* had to be expressed by an NP denoting a human entity whereas this was not the case for *liken*. In the case of *quemen* the focus lies on the THEME as 'volitional', and in the case of *liken* the focus lies on the reaction of the EXPERIENCER (Allen, 1995, p. 147). As a result, ME *quemen* occurs more and more frequently with the THEME in the nominative and with the function of the subject, whereas *liken* occurs more and more frequently with the EXPERIENCER in the nominative and with the function of the subject. The argument structures of the three verbs in (10) show the difference. We also see that the copied verb *plesen* matches the argument structure of *quemen*. Against this backdrop we can then explain why it was the verb *quemen* and not *liken* that was replaced by *plesen*: it not only had the same meaning as *plesen* but also an identical argument structure.

If we add a further level of Myers-Scotton Abstract Level model to our abstract representation of the structure of verbs, namely the predicate-argument level, we can model this case as an example of full integration. Here, the EXPERIENCER argument of *plesen* can occur in a transitive construction being expressed by an NP. Clearly, this structure is part of the system of native verbs, and thus the copied verb *plesen* fully integrates into the system of native verbs (like *quemen*). The following schema models this type of integration:

(11)   Basic Code:
Syntax: NP   NP
                                                   ↓   ↓
LCS: [STATE BE$_{aff+}$ ([X], [Y])]
Foreign Model Code:
Syntax: NP   NP
                                                   ↓   ↓
LCS: [STATE BE$_{aff+}$ ([X], [Y])]
Basic Code: He quemes the king.
Foreign Model Code: He pleases the king.
(AN: *Vus*$_{THE\text{-}sbj}$ *m'avez, dame, fait hunir Pur [vostre maveisté]*$_{EXP\text{-}obj}$ *plaisir*; source: (*TL, AND, Trist (D) 1307)*)

However, in the case where this verb is copied along with its intransitive use, i.e. with an EXPERIENCER in the form of a *to*-PP, we observed an integration conflict as this construction is not used in the native inventory. But since PPs are an option to express arguments in English, EXPERIENCERS with copied verbs can also be expressed in this, albeit marked, way. The following schema illustrates this:

(12)   Basic Code:
Syntax: NP   NP
                                                   ↓   ↓
LCS: [STATE BE$_{aff+}$ ([X], [Y])]
Foreign Model Code:
Syntax: NP   PP
                                                   ↓   ↓
LCS: [STATE BE$_{aff+}$ ([X], [Y])]
Basic Code: He quemes the king.
Foreign Model Code: He pleases to the king.
(AN: *ces [qui]*$_{THE\text{-}sbj}$ *volent [a Deu]*$_{EXP\text{-}obj}$ *pleisir e le suen regne deservir*; source: *Purg S Pat MARIE 43.*))

In the cases where the native psych verbs like *liken* and *quemen* started to exhibit an EXPERIENCER in the form of a *to*-PP, the native predicate-argument structure

adapted to the foreign Model Code, at least for some time. This type was triggered by the integration conflict discussed above, but did not have a long-term effect since this structure is clearly not an option today. The following schema models this case:

(13) Basic Code:
Syntax: NP   PP
                                             ↓   ↓
LCS: [STATE BE$_{aff+}$ ([X], [Y])]
Foreign Model Code:
Syntax: NP   PP
                                             ↓   ↓
LCS: [STATE BE$_{aff+}$ ([X], [Y])]
Basic Code: He quemes to the king.
Foreign Model Code: He pleases to the king.
(AN: *ces [qui]*$_{THE-sbj}$ *volent [a Deu]*$_{EXP-obj}$ *pleisir e le suen regne deservir*;
source: *Purg S Pat MARIE 43*.))

Overall, I think Myers-Scotton's model is very useful to make lexical semantic as well as syntactic integration visible. What remains to be shown is whether (i) all three levels were always affected, (ii) other types of verbs and maybe whole semantic verb classes were affected by the massive copying of OF verbs to ME, and (iii) this had long-term effects on the level of grammar and syntax.

7. Concluding remarks

The previous section has dealt with the copying of the amuse-type psych verbs and integration conflicts that arise when the EXPERIENCER argument is syntactically realised by a *to*-PP. We could extend this discussion to other types of verbs, e.g. verbs of transfer of possession, which are known to express the RECIPIENT argument in the same way. This phenomenon has been called the dative-alternation of ditransitive verbs (see also Levin, 1993: 45). Pinker (1989, p. 45–47), in the context of the acquisition of argument structure, was probably the first to relate the occurrence of this alternation to the origin of the verbs: He noted that native verbs like *give* can form both the double-object construction and the construction with a PP as RECIPIENT, whereas Latinate verbs like *donate* only allow the PP construction.

As an explanation for the difference, Pinker gives two causes: first, the loss of case marking, and second, contact with OF in ME times. Pinker states that verbs copied from OF retained their argument structure and this is why we do not find the dative alternation with ditransitive verbs of Latinate origin today.

Recent studies by Ambridge and collaborators (cf. Ambridge et al. 2012, 2014) dealing with dative argument-structure overgeneralisation errors in language acquisition have tackled this difference as well and provided interesting and actually quite surprising results. They investigated how children learn that certain possible generalisations are ungrammatical and focused on cases where a verb that may occur in the prepositional-object dative construction (*I said no to her*) is incorrectly used in a double-object construction (**I said her no*) and vice versa. Following others, they assume a morphophonological constraint which states that speakers are sensitive to the syllable stress patterns of Latinate verbs (disyllabic with second syllable stress, or trisyllabic) as well as to certain morphemes (e.g. *-ify*, *-ate*). The authors designed an experiment to investigate the psychological reality of this constraint by obtaining grammaticality judgements for sentences with novel Latinate and native-like verbs. More precisely, they sought to find out whether this constraint is psychologically real for adults, and if so, when it is acquired by children. The experiment included three groups of monolingual speakers: adults, children aged 5;4–6;4 and children aged 9;4–10. A number of native and Latinate verbs from the class of illocutionary communication and from the class of verbs of giving were selected (high frequency verbs, e.g. *show*, *give* vs. *explain*, *transport* and low frequency verbs, e.g. *teach*, *hand* vs. *declare*, *donate*) and novel native-like (e.g. *nace*) and novel Latinate (e.g. *orgulate*) verbs were created (for details see Ambridge et al. 2012: 59, Ambridge et al. 2014: 225). The results show that concerning Latinate verbs, the adults and older children prefer the prepositional object over the double object dative uses but no such preference can be shown for Germanic verbs. The younger children do not show this pattern, but instead they seem to exhibit a general dispreference for Latinate verbs, regardless of whether they occur in prepositional object or double-object dative sentences. Concerning the novel verbs, adults prefer prepositional object uses over double-object uses for Latinate verbs whereas they do not have a preference for the native-like novel verbs. Children do not seem to make a distinction at all. The authors conclude that the morphophonological constraint has a psychological reality for adults and can neither be attributed to attested usage nor verb semantics. The most interesting result is probably that adults productively apply the rule of prepositional object dative to novel Latinate verbs (like *orgulate*) that they could not have encountered before. The group of the older children respect this restriction for already existing Latinate verbs but they have yet to acquire a productive constraint.

Looking again at the results of my study, Holler's study, Pinker's claim, and the results of the experiments of Ambridge et al., it seems that what they have in common is that differences exist in the syntactic behaviour of native and nonnative verbs. More precisely, native (Germanic) verbs syntactically mark semantic roles like EXPERIENCER and RECIPIENT with an NP by default (and if they have

a case system, they morphologically mark this argument with a dative), whereas non-native (Latinate) verbs syntactically mark these roles with a *to*-PP by default. This difference may manifest itself as an integration conflict in syntax and can be investigated from different perspectives (synchronic, diachronic, acquisitional). Integration conflicts may be taken as a diagnostic of language contact between two language families like Germanic and Romance, which exhibit a number of typological differences (see also Talmy, 2000). While further multidisciplinary research in this direction is clearly needed, this route seems to be promising and fruitful.

Abbreviations

ACC	accusative	OBJ	object
AFF	affirmative	OF	Old French
AN	Anglo-Norman	PL	plural
DAT	dative	PRES	present
EXP	EXPERIENCER	RFL	reflexive
INTR	intransitive	SG	singular
IRR	irrealis	SBJ	subject
ME	Middle English	TRANS.	transitive
NOM	nominative	2	second person
NP	noun phrase		

References

Allen, C. L. (1995). *Case marking and reanalysis: grammatical relations from Old to Early Modern English*. Oxford: Clarendon Press.

Ambridge, B., Pine, J. M., Rowland, C. F. & Chang, F. (2012). The roles of verb semantics, entrenchment, and morphophonology in the retreat from dative argument-structure overgeneralization errors. *Language*, 88(1), 45–81. https://doi.org/10.1353/lan.2012.0000

Ambridge, B., Pine, J., Rowland, C., Freudenthal, D. & Chang, F. (2014). Avoiding dative overgeneralization errors: semantics, statistics or both? *Language, Cognition and Neuroscience*, 29(2), 218–243. https://doi.org/10.1080/01690965.2012.738300

Appel, R. & Muysken, P. (1987). *Language contact and bilingualism*. London: Arnold.

Blumenthal, P. & Stein, A., editors (2002). *Tobler-Lommatzsch: Altfranzösisches Wörterbuch. 4 CD-ROMs und DVD mit Begleitbuch*. Stuttgart: Steiner.

Comrie, B. (1993). Argument structure. In J. Jacobs, et al. (Eds.), *Syntax: An international handbook of contemporary research*, Vol. 1, (p. 903–914). Berlin: Walter de Gruyter.

Denison, D. (1993). *English historical syntax*. Longon: Longman.

Eisenberg, P. (2001). Die grammatische Integration von Fremdwörtern. Was fängt das Deutsche mit seinen Latinismen und Anglizismen an? In G. Stickel, (Ed.), *Neues und Fremdes im deutschen Wortschatz. Aktueller lexikalischer Wandel*, (pp. 183–209). Berlin: de Gruyter.

Eisenberg, P. (2011). *Das Fremdwort im Deutschen*. de Gruyter Studium. Berlin: de Gruyter. https://doi.org/10.1515/9783110235654

Engelberg, S., Meliss, M., Proost, K., and Winkler, E. (Eds.). (2015). *Argumentstruktur zwischen Valenz und Konstruktion*. Tübingen: Narr Francke Attempto.

Field, F. (2002). *Linguistic borrowing in bilingual contexts*. John Benjamins, Amsterdam. https://doi.org/10.1075/slcs.62

Fischer, O. & van der Leek, F. (1983). The demise of the Old English impersonal construction. *Journal of Linguistics*, 19, 337–368. https://doi.org/10.1017/S0022226700007775

Gradon, P., (Ed.). (1965). *Dan Michel's Ayenbite of Inwyt or Remorse of Conscience*, Vol. I of *The Early English Text Society 278*. [reprint of 1866]. Oxford: Oxford University Press.

Greenberg, J., (Ed.). (1978). *Universals of human language*, Vol. 1. Method and Theory. Stanford: Stanford University Press.

Grossman, E., Seržant, I. & Witzlack-Makarevich, A. (Eds.). (Forthcoming). *Journal of Language Contact. Special Issue on Valency and Transitivity in Contact*. Leiden: Brill.

Haugen, E. (1969). *The Norwegian language in America: A study in bilingual behavior*. Bloomington: Indiana University Press.

Highfield, A. & Valdman, A. (Eds.). (1981). *Historicity and variation in Creole studies*. Ann Arbor: Karoma.

Hock, H. H. & Joseph, B. D. (1996). *Language history, language change, and language relationship*. Berlin: de Gruyter.

Holler, A. (2015). Grammatik und Integration. Wie fremd ist die Argumentstruktur nicht-nativer Verben? In S. Engelberg, M. Meliss, K. Proost, & E. Winkler, (Eds.). *Argumentstruktur zwischen Valenz und Konstruktion*, (pp. 397–416). Tübingen: Narr Francke Attempto.

Holler, A. & Scherer, C. (2010). Zur Argumentstruktur entlehnter Verben. In C. Scherer, & A. Holler (Eds.). *Nicht-native Einheiten und Strukturen*, (pp. 183–198). Berlin/New York: de Gruyter.

Ingham, R. (2012). *The transmission of Anglo-Norman: Language history and language acquisition, Vol. 9 of Series Language faculty and beyond*. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/lfab.9

Jackendoff, R. (1976). Towards an explanatory semantic representation. *Linguistic Inquiry*, 7, 89–150.

Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.

Jackendoff, R. (1990). *Semantic structures*. Cambridge, MA: MIT Press.

Jacobs, J., von Stechow, A., Sternefeld, W., and Vennemann, T., (Eds.). 1993. *Syntax: An international handbook of contemporary research*, Vol. 1. Berlin: de Gruyter. https://doi.org/10.1515/9783110095869.1

Johanson, L. (2002). Contact-induced change in a code-copying framework. In Jones, M. C. & Esch, E., (Eds.). *Language change: the interplay of internal, external and extra-linguistic factors*, (pp. 285–313). Berlin: de Gruyter. https://doi.org/10.1515/9783110892598.285

Johanson, L. (2008). Remodeling grammar. Copying conventionalization, grammaticalization. In Siemund, P. & Kintana, N., (Eds.), *Language contact and contact languages*, (pp. 61–79). Amsterdam/Philadelphia: Benjamins. https://doi.org/10.1075/hsm.7.05joh

Jones, M. C. & Esch, E. (Eds.). (2002). *Language change: the interplay of internal, external and extra-linguistic factors*. Berlin: de Gruyter. https://doi.org/10.1515/9783110892598

King, R. (2000). *The lexical basis of grammatical borrowing. A Prince Edward Island French case study*. Amsterdam/Philadelphia: Benjamins. https://doi.org/10.1075/cilt.209

Kroch, A. & Taylor, A. (Eds.). (2000). *The Penn-Helsinki Parsed Corpus of Middle English, Second Edition (PPCME2)*. Philadelphia: University of Pennsylvania.

Laurent, F. (2008). *La Somme le Roi*. Société des Anciens Textes Français, Paris/Abbeville: Paillart.

Levin, B. (1993). *English verb classes and alternations. A preliminary investigation*. Chicago: University of Chicago Press.

Levin, B. *Argument Structure*. (http://www.oxfordbibliographies.com, consulted on 21 September 2018).

Martineau, F., (Ed.). (2009). *Le corpus MCVF. Modéliser le changement: les voies du français*. Ottawa: Université d'Ottawa.

Matras, Y. (2007). The borrowability of grammatical categories. In Matras, Y. & Sakel, J., (Eds.). *Grammatical borrowing in cross-linguistic perspective*, (pp. 31–74). Berlin: de Gruyter. https://doi.org/10.1515/9783110199192.31

Matras, Y. (2009). *Language contact*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511809873

Matras, Y. & Sakel, J. (Eds.). (2007). *Grammatical borrowing in cross-linguistic perspective*. Berlin: de Gruyter. https://doi.org/10.1515/9783110199192

Moravcsik, E. A. (1978). Language contact. In Greenberg, J., (Ed.), *Universals of human language*, Vol. 1 Method and Theory, (pp. 93–122). Stanford: Stanford University Press.

Muysken, P. (1981). Halfway between Quechua and Spanish: The case for relexification. In A. Highfield & A. Valdman (Eds.). *Historicity and variation in Creole studies*, (pp. 52–78). Ann Arbor: Karoma.

Myers-Scotton, C. (2002). *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198299530.001.0001

Myers-Scotton, C. & Jake, J. (1995). Matching lemmas in a bilingual language production model: Evidence from intrasentential codeswitching. *Linguistics*, 33, 981–1024. https://doi.org/10.1515/ling.1995.33.5.981

Möhlig-Falke, R. (2012). *The Early English impersonal construction. An analysis of verbal and constructional meaning*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199777723.001.0001

Orr, J. (1962). *Old French and Modern English idiom*. Oxford: Blackwell.

Pinker, S. (1989). *Learnability and cognition. The acquisition of argument structure*. Cambridge MA: MIT Press.

Poplack, S. (2018). *Borrowing. Loanwords in the speech community and in the grammar*. Oxford: Oxford University Press.

Primus, B. (1999). Rektionsprinzipien. In H. Wegener (Ed.), *Deutsch. kontrastiv. Typologisch-vergleichende Untersuchungen zur deutschen Grammatik*, (pp. 135–170). Tübingen: Stauffenburg.

Primus, B. (2012). *Semantische Rollen*. Heidelbert: Universitätsverlag Winter.

Prins, A. A. (1948). French phrases in English. *Neophilologus*, 32, 28–39, 73–82. https://doi.org/10.1007/BF01534790

Prévost, S. & Stein, A. (Eds.). (2013). *Syntactic reference corpus of Medieval French (SRCMF)*. Lyons: ENS; Paris: Lattice; Lyon/Stuttgart Universität Stuttgart.

Rothwell, W. (1976). The role of French in thirteenth century England. *Bulletin of the John Rylauds Library*, 58, 445–66. https://doi.org/10.7227/BJRL.58.2.8

Rothwell, W. (1980). Lexical borrowing in a medieval context. *Bulletin of the John Rylands Library*, 63, 118–43. https://doi.org/10.7227/BJRL.63.1.6

Rothwell, W. & Trotter, D. (Eds.). (2001). *Anglo-Norman Dictionary 2. Online Version*. <http://www.anglo-norman.net/>.

Scherer, C. & Holler, A. (Eds.). *Nicht-native Einheiten und Strukturen*. Berlin/New York: de Gruyter.

Siemund, P. & Kintana, N. (Eds.). (2008). *Language contact and contact Languages*. Amsterdam/Philadelphia: Benjamins. https://doi.org/10.1075/hsm.7

Stickel, G. (Ed.). 2001. *Neues und Fremdes im deutschen Wortschatz. Aktueller lexikalischer Wandel.* Berlin: de Gruyter.

Talmy, L. (2000). *Toward a cognitive semantics. Vol. 2: Typology and process in concept structuring.* Cambridge, MA: MIT Press.

Thomason, S. G. & Kaufman, T. (1988). *Language contact, creolization and genetic linguistics.* Berkeley: University of California Press.

Togeby, K. (1983). *Grammaire française*, vol. II. Copenhagen: Akademisk Forlag.

Trips, C. & Stein, A. (2018). A comparison of multi-genre and single-genre corpora in the context of contact-induced change. In R. Whitt, (Ed.). *Diachronic corpora, genre and language change*, Studies in Corpus Linguistics. Amsterdam/Philadelphia: Benjamins. pages 241–260.

Trips, C. & Stein, A. (2019). Contact-induced changes in the argument structure of Middle English verbs on the model of Old French. In E. Grossman, I. Seržant & A. Witzlack-Makarevich (Eds.). *Journal of Language Contact. Special Issue on Valency and Transitivity in Contact.* Leiden: Brill. pages 232-267.

Troberg, M. (2008). Dynamic two-place indirect verbs in French: A synchronic and diachronic study in variation and change of valence. University of Toronto, PhD Thesis.

Wegener, H. (Ed.). 1999. *Deutsch kontrastiv. typologisch-vergleichende Untersuchungen zur deutschen Grammatik.* Tübingen: Stauffenburg.

Whitt, R. (Ed.). (Forthcoming). *Diachronic corpora, genre and language change*, Studies in Corpus Linguistics. Amsterdam/Philadelphia: Benjamins.

Winford, D. (2003). *An Introduction to contact linguistics.* Oxford: Blackwell.

Wohlgemuth, J. (2009). *A typology of verbal borrowings.* (Trends in Linguistics. Studies and Monographs [TiLSM] 211). Berlin / New York: de Gruyter. https://doi.org/10.1515/9783110219340

Woolford, E. (2006). Lexical case, inherent case, and argument structure. *Linguistic Inquiry*, 37, 111–130. https://doi.org/10.1162/002438906775321175

Zaenen, A. & Maling, J. (1990). Unaccusative, passive and quirky case. In A. Zaenen & J. Maling (Eds.). *Modern Icelandic syntax*, (p. 137–152). New York: Academic.

Zaenen, A. & Maling, J., (Eds.). *Modern Icelandic syntax*. New York: Academic.

Contact-induced change and the phonemicization of the vowel /ɑ/ in Quảng Nam Vietnamese

Andrea Hoa Pham

University of Florida

This paper is an account of the emergence of the low, back, unrounded vowel /ɑ/ in the Quảng Nam dialect spoken in south central Vietnam. This vowel is not seen in any other dialects. The paper provides evidence for a trace of this vowel in two subdialects of Hà Tĩnh province, north central Vietnam (Phạm, 1997, 2014, 2016), and claims that the Quảng Nam /ɑ/ originated from Hà Tĩnh dialects through migration. It was brought to Quảng Nam by early settlers mainly during the 15th to 18th centuries, where the vowel was further internally restructured through various linguistic processes. The Quảng Nam /ɑ/, therefore, originates from two sources: dialect contact and internal restructuring.

Keywords: Vietnamese, phonemicization, dialect contact, vowel chain shift

1.   Introduction

Vietnamese is the largest Mon-Khmer language of the Vietic group in the Austroasiatic family, yet also the most innovative one, with its introduction of monosyllabification and the loss of affixes. About 4,000 years ago a part of the Proto-Vietic population migrated from Laos and north central Vietnam (NCV) to north Vietnam. Through contact with Tai-Kadai languages, these Proto-Vietic languages developed into Common Viet-Muong, around the 10th to 14th centuries. Through contact with Chinese, Common Viet-Muong split into the Vietnamese and Muong languages, around the 13th to 15th centuries. Vietnamese is spoken in the Red River delta and along the coastline regions. Muong languages are spoken in remote, mountainous areas. Muong languages and Vietnamese dialects in NCV still retain many old features from the Common Viet-Muong (Trần T. D., 2005). The Vietnamese and Muong languages have very similar phonological systems except that initial consonant clusters were lost in northern Vietnamese but are still present

© 2020 John Benjamins Publishing Company

in Muong languages (Barker and Barker, 1970; Nguyễn V. T., 2005). Due to a long history of contact between speakers of several languages and dialects, the interinfluences among them are extremely complex, and we still know very little about them. This paper teases out various factors that influence the emergence of the low, back, unrounded /ɑ/ in a Vietnamese dialect spoken in Quảng Nam, the land acquired from Champa Kingdom in the 15th century. The vowel /ɑ/ is not seen in any major dialects of Vietnamese.

"Middle Vietnamese" refers to the variety spoken in the northern and southern regions of Vietnam. It was first recorded in Alexander de Rhodes' 1651 *Annam-Lusitan-Latin dictionary* using a Romanized script (*Quốc ngữ* 'National Script'). The phonemic-based script was created by European Christian missionaries in Vietnam for evangelization purposes. The Vietnamese sounds reflected in this dictionary are almost the same as in present-day Vietnamese, except that initial clusters such as tl- or bl- have now disappeared in modern Vietnamese. The Romanized script was improved over time, and in Pigneaux de Béhaine's 1773 *Vietnamese-Latin dictionary* it was almost identical to the script used today. The script represents all contrastive sounds from all Vietnamese dialects. However, it has no symbol for the vowel /ɑ/ discussed in this paper. Rhodes arrived in Hanoi in 1620 and spent ten years in and around there, and a few years in Quảng Nam. Béhaine worked in southern Vietnam and prepared the 1773 dictionary based on southern dialects. In neither dictionary do we find a trace of the back, unrounded vowel /ɑ/.

The Vietnamese language described in the literature is primarily based on the Hanoi dialect, a representative of northern dialects. In studying the history of Vietnamese sounds, it is much more important to compare the dialects spoken in NCV with northern or southern dialects than to compare dialects spoken in northern with those in southern cities (Maspero, 1912; Nguyễn, T. C., 1997). This paper discusses the origin and phonemicization of the Quảng Nam /ɑ/ and compares it with related vowels from two sub-dialects spoken in the Hà Tĩnh province of NCV. The Hà Tĩnh data suggest that the Quảng Nam /ɑ/ was introduced by early settlers from NCV. The vowel was then phonemicized through linguistic processes. These processes are common in other innovative southern dialects. The Quảng Nam /ɑ/, therefore, originates from two sources: dialect contact and internal restructuring. Discussing the origin of the Hà Tĩnh /ɑ/ is beyond the scope of this paper.

Section 2 introduces the Quảng Nam /ɑ/. Section 3 provides the historical background on territories and migration. Section 4 documents a vowel similar to Quảng Nam /ɑ/ in the Hà Tĩnh subdialects. Section 5 shows how /ɑ/ was restructured in the Quảng Nam dialect, and became a phoneme, and why the vowel was not strictly an internal development. The paper concludes with Section 6.

2. The Quảng Nam vowel /ɑ/

Quảng Nam, circled, is located about 900 km south of Hanoi (#5 on Map 1) with a population of approximately 1.5 million as of 2016 (General Statistics Office of Vietnam).
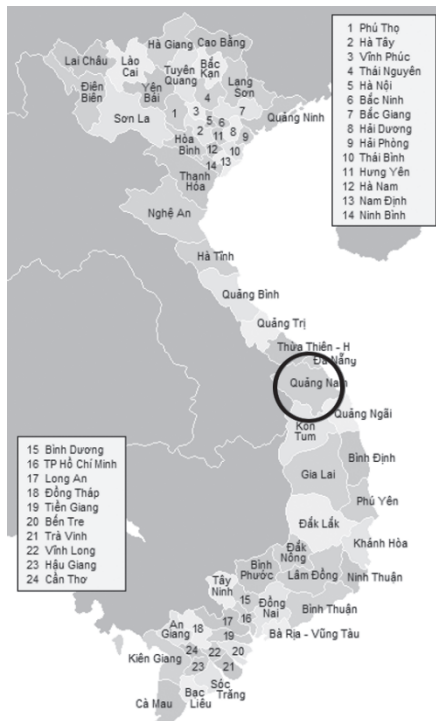
Map 1. Provinces of Vietnam

The Quảng Nam dialect has similar onset and tonal systems as those in other southern dialects. However, the Quảng Nam rhymes are very different with many peculiarities (Phạm A. H., 1997). The correspondences between vowels in the Hanoi and Quảng Nam dialects are shown in (1), focusing on the low, central /a/ and /ă/ in northern dialects, transcribed as *a* (long) and *ă* (short). Tones are indicated with numbers, superscripted. In open syllables, Quảng Nam and Hanoi vowels are similar except that the Hanoi long /a/ corresponds to the Quảng Nam /ɑ/ in both open and closed syllables (1a–f). In syllables ending with a glide, the Hanoi rhymes /ăw/ and /ăj/ correspond to Quảng Nam central [a] (1g–h). Hanoi short /ă/ corresponds to Quảng Nam long [a] before labial (1i), or [ɛ] before non-labial consonants (1j–k).

(1) Correspondences of the Hanoi and Quảng Nam vowels

	Vietnamese orthography	HANOI	QUẢNG NAM	Gloss
a.	*cá*	ka³	kɑ³	'fish'
b.	*làm*	lam²	lɑm²	'to do'
c.	*ngáp*	ŋap⁷	ŋɑp⁷	'to yawn'
d.	*bàn*	ban²	bɑɲ²	'table'
e.	*bác*	bak⁷	bɑk⁷	'uncle'
f.	*làng*	laŋ²	lɑŋ²	'village'
g.	*cau*	kăw¹	ka¹	'areca nut'
h.	*cay*	kăj¹	ka¹	'spicy'
i.	*năm*	năm¹	nam¹	'five'
j.	*lặn*	lăn⁴	lɛɲ⁴	'to dive'
k.	*nắng*	năŋ³	nɛŋ³	'sunny'

Descriptions of the Quảng Nam /ɑ/ vary widely in the literature. The IPA symbol for the vowel and its articulatory descriptions are provided, but these, unfortunately, do not include phonological or phonetic justifications. Except for Tohyama (2015), a master's thesis, these articles are only a few pages in length. The vowel is transcribed as [ˤa] in open syllables (Shimizu, 2013 and Tohyama, 2015). Vương H. L. (1998) and Trần T. T. (2003) correctly describe the vowel as low, back, and unrounded, and transcribe it with [ɑ]; however, neither author offers evidence or states if the symbol used is phonemic or phonetic. The vowel's distribution is not discussed or fully described. Trần T. T. (2003) presents the vowel only in open syllables where it corresponds to Hanoi rhymes [ăw] and [ăj], e.g., Quảng Nam [sɑ³] 'six' ~ Hanoi [săw³]. The vowel is also transcribed as [ɔ] or [o] before labial consonants (Trần T. T., 2003) or velar consonants (Tohyama, 2015), with a remark that in closed syllables /a/ changes to [ɔ] (Shimizu, 2013), and that the Quảng Nam sound changes are random (Trần T. T., 2003). In Section 5, I demonstrate that the symbols [o], [ɔ], and [ˤa] used in these studies to transcribe the Quảng Nam /ɑ/ indeed reflect an assimilation effect. In sum, it is at least possible to say that all symbols indicate either backness or rounding of the vowel. Phạm A. H. (1997, 2014, 2016) provide phonetic and phonemic evidence (e.g., distribution, cluster reduction, and acoustic properties) for this vowel being the low, back, unrounded phoneme /ɑ/.

Table 1 displays the Quảng Nam vowels. In Vietnamese short vowels occur only in syllables ending in a consonant or glide. Glides occupy the coda position. There are no phonemic diphthongs in Quảng Nam. Length is contrastive in several vowels. In closed syllables, /ɛ/ and /ɔ/ tend to raise to /e/ and /o/, respectively.

Vowel inventories in other Vietnamese dialects can differ from one another in the number of vowels with a length contrast, or in the presence vs. absence of diphthongs; but they do not have the vowel /ɑ/ that Quảng Nam does. If the Quảng

Table 1. The Quảng Nam vowel system

	Front		Central		Back	
					Unrounded	Rounded
High	i	ĭ	ɨ	ɨ̆		u ŭ
Mid	e		ɤ	ɤ̆		o
Low	ɛ	ɛ̆	a	ă	ɑ	ɔ

Nam /ɑ/ does not occur in any other dialects of Vietnamese, was it an innovation or acquired through contact? The following sections provide evidence that Quảng Nam /ɑ/ was introduced by early settlers from NCV. However, it has been restructured to become stable as a new phoneme. The Quảng Nam /ɑ/, therefore, comes from two sources: dialect contact and internal development.

3. Quảng Nam: The land and the people

The provinces from current Quảng Bình (NCV) southward, including Quảng Nam, used to be the territories of the Champa Kingdom, a country that existed from the second century CE until 1832. During the 11th to 19th centuries, the Champa lands were gradually absorbed by Vietnam, often through war. Map 2 shows Champa in 1402, after Quảng Nam fell under the authority of Vietnam. The dark line marks the boundary, Vietnam in the north, and Champa in the south.

The Vietnam court in the 16th century was ruled by the Trịnh family, under the name of the Lê monarchy. Conflicts arose between the Trịnh and Nguyễn families. During this period of conflict, a Nguyễn general, afraid of being eliminated by the Trịnh family, asked to be transferred to the deserted, remote southern region to serve the monarchy. Many high Nguyễn officials and other families from Thanh Hoá province followed him (Li, 1998). The Nguyễn family began to build another Vietnam, called the Inner State, in the south, as opposed to the Outer State in the north. Although the first provinces of Champa were annexed to Vietnam in 1069, the mass Vietnamese migrations southward between 1470 and 1835 are the most important forces that reorganized Vietnam into two states, with distinct cultural and economic characteristics (Li, 1998). Map 3 shows the Inner State in 1690: the dark line across the map marks the region labeled Quảng Nam in the south. In the 17th century, the name "Quảng Nam" comprised the current Quảng Nam, Quảng Ngãi and Bình Định provinces.[1]

---

1. The Quảng Nam /ɑ/ also occurs in the Quảng Ngãi dialect (Trần T. T. A., 2015).
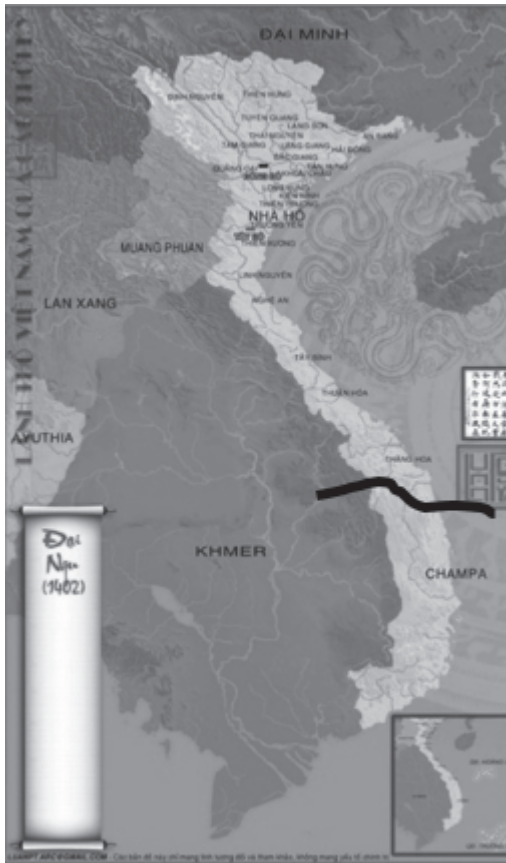
Map 2. Vietnam in 1402

Most Vietnamese immigrants in Quảng Nam originated from Thanh Hoá, Nghệ An, and Hà Tĩnh provinces in NCV. Several families in Quảng Nam claim, through genealogical records, that their ancestors were generals from Thanh Hoá who came to Quảng Nam on military duties in the late 15th century. Migration to the new land was not always voluntary. From 1475 the Lê royal court began to send prisoners to the far south. The more serious the crime committed, the further south one was sent. Most famines in the 16th and 17th centuries occured in the north central region, causing waves of migration to the southern lands (Li, 1998). After the victory of 1653–1657, Nguyễn Lord forced the residents from seven districts of Nghệ An to move to Quảng Nam (Ngô G. V. P., 2011, p. 78).

The Cham people are not ethnically homogenous, but rather an entity that consists of the Chams and other Austronesian and non-Austronesian tribes (Hickey,
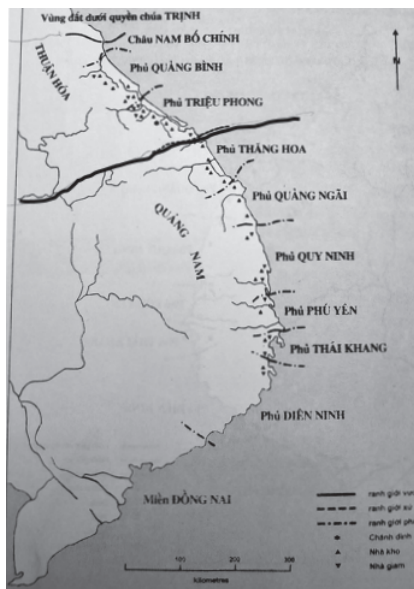
Map 3. The Inner State in 1690 (Li, 2016)

1982; Po Darma, 1987). The Cham people in Vietnam lived together in small pockets and spoke the dialect of eastern Cham (Ken, 2004; Li, 1998; Phan K., 1969). Cham is a non-tonal language of the Malayo-Polynesian branch in the Austronesian language family. The Vietnamese and Cham people adopted each other's cultural customs (Li, 1998; Ken, 2004); however, there is little evidence that the two languages influenced one another's sound systems, outside the claim that pitch was becoming the dominant feature of the tonal registers in eastern Cham, due to their contact with Vietnamese (e.g., Brunelle, 2009). The vowel /ɑ/ does not occur in any of the contemporary eastern Cham dialects. It is very unlikely that the Quảng Nam vowel /ɑ/ is a result of language contact with Chams.

If the Quảng Nam /ɑ/ did not originate from language contact with the Chams, could it be from dialect contact? Indeed, along with /ɑ/ and its allophonic variations, many other linguistic peculiarities in the Quảng Nam dialect were found in the two sub-dialects of Hà Tĩnh province. The following section describes the vowel /ɑ/ in the dialects spoken in the Hến and Kẻ Chay hamlets, Đức Thọ rural district, Hà Tĩnh.

4.   The Hà Tĩnh /ɑ/: Migration as a source of the emergence of the Quảng Nam /ɑ/

The data were collected by the author in 2016 in Hà Tĩnh, 300 miles north of the Quảng Nam province. A picture naming method was used, i.e., the consultants were asked to name the pictures presented to them. A total of 197 words representing all possible rhymes in contemporary Vietnamese were elicited (see Appendix). The photos were randomly arranged in two separate files for repetition. The first sub-dialect is spoken in the Hến hamlet of Trường Sơn village (the circle on Map 4), bordered by Nghệ An province in the north. The second subdialect is spoken in the Kẻ Chay hamlet of Đức An village (the square on Map 4).
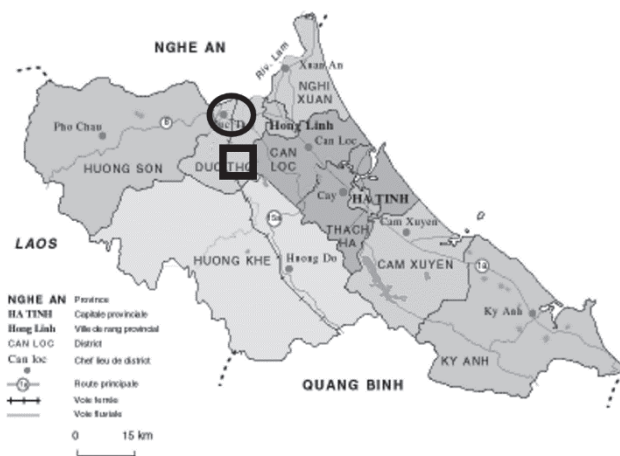
Map 4. Hà Tĩnh province

The data and analysis below focus on the vowel resembling the Quảng Nam /ɑ/ in Hến and Kẻ Chay dialects.

4.1   Hến dialect

The data were collected from one male and two female speakers (age: 50 ~ 70). Table 2 shows the vowel system in open (CV) and closed syllables (CVC) in the Hến dialect (Phạm A. H., 2016). There are three falling diphthongs, /iə/, /ɨə/, and /uə/. Note that the vowel /a/ is absent in open syllables. Length is contrastive only in /ɤ/. The low, back, unrounded vowel /ɑ/ occurs in all syllable types. In this dialect, [ɔ] often raises to [o], e.g., Hanoi *chó* /cɔ³/ 'dog', *nho* /ɲɔ¹/ 'grapes', *đỏ* /dɔ⁵/ 'red', *tỏi*

/tɔj⁵/ 'garlic', *bọt* /bɔt⁸/ 'bubbles' correspond to Hến [co], [ɲo], [do], [toj], and [bot], respectively.[2] The contrast between /e/ and /ɛ/ seems to be maintained in most cases, although /ɛ/ sometimes raises as well.

Table 2. The Hến vowel systems in open (CV) and closed (CVC) syllables

	CV			CVC		
	Front	**Central**	**Back**	**Front**	**Central**	**Back**
Diphthong	iə	ɨə	uə	iə	ɨə	uə
High	i	ɨ	u	i	ɨ	u
Mid	e	ɤ	o (ɔ)	e	ɤ ɤ̆	o (ɔ)
Low	ɛ		ɑ	ɛ	a	ɑ

In Figure 1, the vowel formants of [ɑ], [a], [ɔ] and [o] in the Hến dialect, from tokens where they are present, show that [ɑ] occurs in the vowel space near the two back vowels, [ɔ] and [o]. The long vowel /a/, seen only in closed syllables, is shown for comparison (Phạm A. H., 2016).
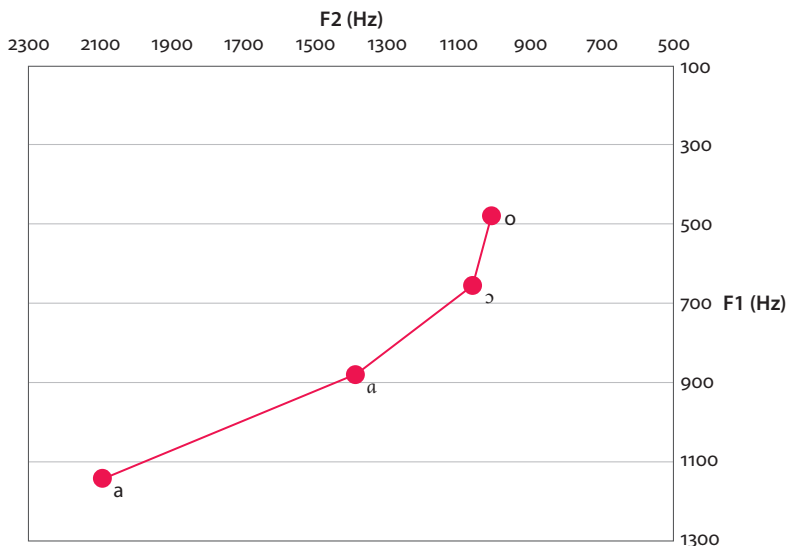
Figure 1. Vowel formants of [a], [ɑ], [ɔ], and [o] in the Hến dialect

---

2.   Hến and Kẻ Chay have tonal systems like those of dialects spoken in NCV and are completely different from the Hanoi and Quảng Nam tones. Tones are not discussed in this paper; therefore, they are not marked for the Hến or Kẻ Chay data.

Comparative data concerning /a/, /ă/, and /ɑ/ in the Hanoi, Hến and Quảng Nam dialects are given in (2). The Hanoi long /a/ corresponds to the Hến [ɑ] in all syllable types (2a–j), and Hanoi short /ă/ corresponds to the Hến long /a/ (2k–p). Thus the Hanoi short /ă/ vs. long /a/ contrast corresponds to a vowel quality difference of central /a/ vs. back /ɑ/ in Quảng Nam (2a–f). In syllables ending with a glide, Hanoi /aj/ and /aw/ correspond to Quảng Nam /ɨə/ and /o/, respectively (2g–j). Hanoi short /ă/ corresponds to Quảng Nam long /a/ or /ɛ/ (2k–p).

(2)		Vietnamese orthography	HANOI	HẾN	QUẢNG NAM	Gloss
	a.	*cá*	ka^3	kɑ	ka^3	'fish'
	b.	*ba*	ba^1	bɑ	ba^1	'three'
	c.	*làm*	lam^2	lɑm	lam^2	'to do'
	d.	*ngáp*	ŋap^7	ŋɑp	ŋap^7	'to yawn'
	e.	*bàn*	ban^2	bɑn	baŋ^2	'table'
	f.	*bác*	bak^7	lɑŋ	bak^7	'uncle'
	g.	*dài*	zaj^2	zɑj	jɨə^2	'long'
	h.	*mai*	maj^1	mɑj	mɨə^1	'tomorrow'
	i.	*áo*	ʔaw^3	ʔɑw	ʔo^3	'blouse'
	j.	*bão*	baw^6	bɑw	bo^5	'storm'
	k.	*bảy*	băj^5	baj	ba^5	'seven'
	l.	*sáu*	săw^3	ʃaw	ʃa^3	'six'
	m.	*năm*	năm^1	nam	nam^1	'five'
	n.	*bắp*	băp^7	bap	bap^7	'corn'
	o.	*măng*	măŋ^1	maŋ	mɛŋ^1	'bamboo shot'
	p.	*rắn*	zăn^3	tan	rɛŋ^3	'snake'

4.2   Kẻ Chay dialect

Kẻ Chay dialect is largely unintelligible to speakers of northern and southern dialects due to its different rhymes and grammatical particles. Speakers of this dialect often use the regional standard variety spoken in the Hà Tĩnh city, when talking to an outsider. The data presented in this paper were recorded from one speaker only, Mr. Đào, who was 80 years old at the time of recording. Due to the time constraint, the author could not stay longer to record more speakers. Although he represents the only source, Mr. Đào's speech shows consistency. On the one hand, in his speech there are features typical for many dialects in NCV, e.g., the preservation of non-spirantized initial medial stops and the initial tl- as remnants of Middle Vietnamese, the narrow tonal space and small tone inventory. On the other hand, certain ideosyncratic features in Mr. Đào's speech are seen either in Hến, another dialect of Hà Tĩnh, or in the Quảng Nam, a southern dialect. For example, [ɑ] is seen in Hến dialect, while the correspondence between [ow] and [o] is also shared in the Quảng Nam dialect. Until a larger sample of Kẻ Chay is available, Mr. Đào's speech will be referred to as Kẻ Chay dialect.

Kẻ Chay shares with Quảng Nam not only the low, back, unrounded [ɑ] but also many rhyme peculiarities. Figure 2 shows the graph of Kẻ Chay vowels in open syllables based on their F1 and F2 formants, with examples in the Vietnamese orthography (Phạm, A. H., 2019). Like in the Hến dialect, the central vowel /a/ does not occur in open syllables. The vowels [ɛ] and [ɔ] often raise to [e] and [o] or [u], respectively, e.g., Hanoi *beo* [bɛw] 'jaguar', *sét* [sɛt] 'lightening', *đỏ* [dɔ] 'red', *họp* [hɔp] 'meeting' correspond to Kẻ Chay [bew], [ʃet], [du], [hop], respectively. This confusion is reflected in a short distance between mid vowels, front and back, especially for the front vowels, where [e] and [ɛ] are almost merged. The loss of contrast between /e/ and /ɛ/, and /o/ and /ɔ/ in closed syllables is common in southern dialects (Hoàng T. C., 1989; Ngô T. M., 2017; Nguyễn T. H., 2016; Vương & Hoàng, 1994).
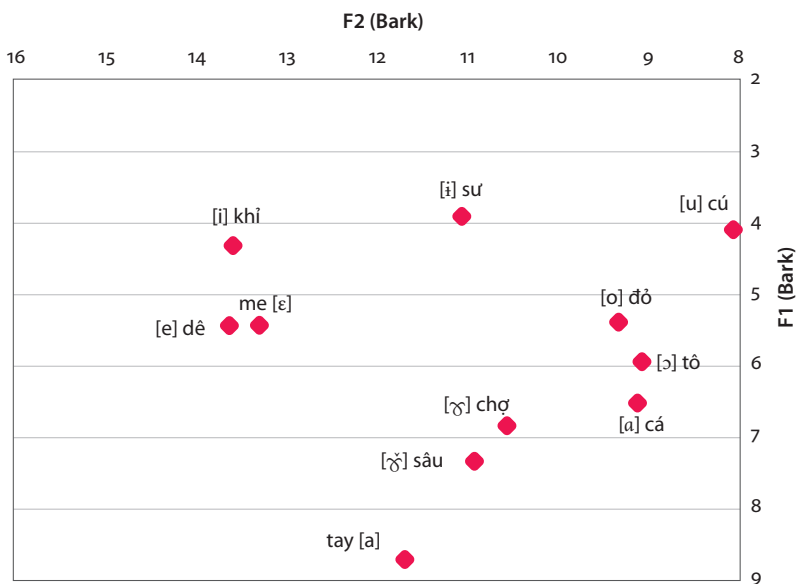
Figure 2. Kẻ Chay vowels

Table 3 shows the vowel inventories of the Kẻ Chay dialect. Four vowels have length contrast in closed syllables. As in the Hến dialect, central /a/ only occurs in closed syllables, and the back, unrounded /ɑ/ is present. Vowels [ɛ] and [ɔ] in Kẻ Chay are presented as variations of /e/ and /o/.

The most striking aspect about the vowel [ɑ] in the Kẻ Chay dialect is that this vowel manifests itself with many phonetic variants. The Hanoi long /a/ corresponds to Kẻ Chay [ɑ], e.g., *bác* [bɑk] 'uncle', *tai* [tɑj] 'ear'; to [ɔ] or [o], e.g., *tám* [tom] 'eight', *mai* [mɔj] 'yellow flowers', *gà* [ɣo] 'chicken'; sometimes is strongly labialized, e.g., *ba* [bˠa] 'three', *lan* [lˠan] 'orchid', *hai* [hˠaj] 'two'. In addition to qualitative auditory perception and distribution of the vowel, Pham A. H. (2019) provides the acoustic measurements of these variants in Kẻ Chay, corresponding to Hanoi

Table 3. Kẻ Chay vowels

	CV			CVC		
	Front	**Central**	**Back**	**Front**	**Central**	**Back**
High	i	ɨ	u	i ĭ	ɨ ɨ̆	u ŭ
Mid	e (ɛ)	ɤ	o (ɔ)	e (ɛ)	ɤ ɤ̆	o (ɔ)
Low			ɑ		a	ɑ

/a/, to show inconsistency. The formant values of F1, F2 and F3 of twelve syllables containing the vowel corresponding to Hanoi /a/ indicate the Kẻ Chay vowel is best analyzed as the low, back, unrounded /ɑ/, not the rounded /ɒ/. These variants in Kẻ Chay are similar to those used to describe the Quảng Nam /ɑ/, although the vowel occurs in different environments in the two dialects. Among these variations, /ɑ/ is chosen as a phonemic representative for the vowel in Kẻ Chay. Similar correspondences with Hanoi central /a/ are also observed elswhere in NCV. For example, in the Nghi Lộc dialect of Nghệ An province, the Hanoi /a/ corresponds to [ˀa], e.g., *bà* [bˀa] 'grandmother', *chát* [cˀat] 'bitter' (Vũ T., 2012). It is possible that /ɑ/ occurs in many still unknown subdialects in NCV.

Representative examples for the correspondences of vowels /a/ and /ă/ between the Hanoi, Hến, Kẻ Chay and Quảng Nam dialects are given in (3). Kẻ Chay appears with many variations. When the Hanoi syllables end in a glide /j/ or /w/, the glides disappear in the Quảng Nam dialect. The disappearance of glides is discussed in the following section.

(3)		Vietnamese orthography	Hanoi dialect	Hến dialect	Kẻ Chay dialect	Quảng Nam dialect	Gloss
	a.	*ba*	ba[1]	bɑ	bˀa	bˀɑ[1], bɑ[1]	'three'
	b.	*gà*	ɤa[2]	ɤɑ	ɤo	ɤˀɑ[2], ɤɑ[2]	'chicken'
	c.	*lan*	lan[1]	lɑn	lˀan	lɑŋ[1]	'orchid'
	d.	*tám*	tam[3]	tɑm	tɑm, tom	tɑm[3], tom[3]	'eight'
	e.	*tai*	taj[1]	tɑj	taj, tˀaj	tɨə[1]	'ear'
	f.	*mai*	maj[1]	mɑj	mɔj	mɨə[1]	'yellow' 'flowers'
	g.	*sao*	saw[1]	ʃaw	ʃo	ʃo[1]	'star'
	h.	*tắm*	tăm[7]	tam	tam	tam[7]	'to take a bath'
	i.	*máu*	măw[3]	maw	maw	ma[3]	'blood'
	j.	*tay*	tăj[1]	taj	taj	ta[1]	'hand'

Kẻ Chay and Quảng Nam dialects differ mostly in their tonal systems, and in the loss of alveolar consonants after non-front vowels in the Quảng Nam dialect. However, in addition to the presence of the low, back, unrounded /ɑ/ (Table 4, b), they share many other linguistic features: the correspondence between Hanoi

/aw/ and /o/ (Table 4, c); the unstable contrast between mid vowels (Table 4, d); the shortening of Hanoi long /a/ in close syllables (Table 4, e); the length contrast in high vowels (Table 4, f); a trace of the 17th century tl- (Table 4, g); and finally, the absence of three phonemic diphthongs (Table 4, h and see also Phạm A. H., 1997, 2019).

Table 4 summarizes striking similarities (in the shaded cells) between the Quảng Nam, Hến, and Kẻ Chay dialects, compared to the Hanoi dialect.

Table 4. Comparison of features of Hanoi, Hến, Kẻ Chay and Quảng Nam dialects

	Linguistic features	Hanoi (north)	Hến (north-central)	Kẻ Chay dialect (north-central)	Quảng Nam (central)
a.	[a] in CV	+	−	−	+
b.	presence of [ɑ]	−	+	+	+
c.	[aw] ~ [o] correspondence	−	−	+	+
d.	clear contrast in mid vowels, /e/ vs /ɛ/, /o/ vs /ɔ/	+	−	−	−
e.	contrast /a/ and /ă/ in CVC	+	−	−	−
f.	length contrast in high vowels	−	−	+	+
g.	trace of tl-	−	+	+	+
h.	three phonemic diphthongs	+	−	−	−

In all other Vietnamese dialects described in the literature, the central /a/ is always present but /ɑ/ is not, except for the Quảng Nam dialect. The vowel /ɑ/ is not seen in Proto-Vietic, Common Viet-Muong, Muong languages (Barker and Barker, 1970; Nguyễn V. T., 2005), or Middle Vietnamese.[3] According to the Hến and Kẻ Chay speakers, their ancestors came from Thanh Hoá province. The vowel /ɑ/ is not present in any description of the Thanh Hoá dialects.[4]

The absence of /a/ and the emergence of /ɑ/ in open syllables in the Hà Tĩnh subdialects seem connected. Backing of /a/ is common in sound change, which creates a vowel chain shift in Hà Tĩnh dialects. The chain shift was briefly mentioned in Đức H. (2016).

---

3. In Vietnamese around the 17th century, /a/ occurred in both Vietnamese and lexical borrowings from Chinese. Around 1,300 years ago, there was /a/ in the Viet-Muong group, but in Chinese borrowings, /a/ has at least two origins: */ɑ/ and */a/ (Nguyễn T. C., 1997, p. 168).

4. Backing of /a/ in Hà Tĩnh could be strictly linguistic-motivated, or as a result from contact with other Vietic languages in the region. Ferlus (1995) uses both [ɑ] and [ɒ] to describe the Cao Lao Hạ dialect of the Quảng Bình province, south of Hà Tĩnh. The vowel /ɑ/ is also reported in Arem, a Vietic language spoken in southwest of Quảng Bình (Ferlus and Trần, 2013). This vowel corresponds to Proto Viet-Muong *[ɔ:], and to contemporary Vietnamese [ɔ] and [uə].

(4)   a   => ɑ (ɔ, o)
       ɔ, o => u

Vietnamese [ɔ] is articulated as low as [ɒ], which is acoustically close to, and easily confused with, the low, unrounded, back vowel [ɑ]. Raising of [ɔ] might make it more distinct from [ɑ]. The backing of /a/ in the Hà Tĩnh subdialects is similar to the *trap-bath* split in English, where before certain consonants, [æ] was lengthened and then merged with /ɑ/. This lengthened vowel is variably realized as long, fronted [ɐ] or [a], or as rounded [ɒ] or [ɔ] in various accents of English (Wells, 1982).

4.3   Variations of /ɑ/

It is well known that sound change is manifested in dialect variation. The existence of variations in speakers' pronunciation "can create ambiguity and lead to the listener's misapprehension of the intended pronunciation norm, which is a changed pronunciation" (Ohala, 1989, p. 175). Whereas we still do not know how /ɑ/ was first acquired in the Hà Tĩnh dialects, the data show that the vowel is more stable in the Hến and Quảng Nam dialects than in the Kẻ Chay dialect, and that allophonic realizations of /ɑ/ are more abundant in the Kẻ Chay dialect than in the Quảng Nam dialect. The many variants of Kẻ Chay [ɑ] suggest that variants of the vowel are still competing with one another until one variant is selected and becomes regular.

   In open syllables Quảng Nam /ɑ/ might be pronounced as [ᵊɑ]. The diphthongization of vowels in open syllables is common in dialects from Quảng Nam southward. The vowel is pronounced lower than the corresponding vowel in northern dialects, and is colored with an onglide or offglide, which differs from the syllabic vowel in height, openness or roundness, e.g., /i/ is pronounced as [ᵊi], /e/ as [eʲ], /u/ as [ᵊu] (e.g., Lê N. M. H., 2016; Vương H. L., 1998; Vương & Hoàng, 1994). In the Kẻ Chay dialect, [ɛ] and [ɔ] are strongly diphthongized, with or without raising, e.g., *sen* [ʃ ʲɛn] 'lotus', *chó* [c ᵘo] 'dog', *tỏi* [t ᵘoj] 'garlic, *ngọt* [ŋᵘot] 'sweet'. Diphthongization of /ɑ/ in the Kẻ Chay and Quảng Nam dialects could be a mechanism to maximize contrasts and enhance perception. Before labial consonants, /ɑ/ can be articulated as [o] due to assimilation between the back vowel and the labial consonant.

5.   Phonemicization of /ɑ/ in the Quảng Nam

If, as argued above, the Quảng Nam [ɑ] descended from the Hà Tĩnh /ɑ/, how was this vowel integrated into the new dialect, where it becomes a new phoneme and contrasts with /a/ and /ɔ/ in all syllable types?

Phonemicization of the Quảng Nam /ɑ/ might be promoted by the following linguistically internal processes: (a) the missing /a/ in Hà Tĩnh open syllables was restored in the Quảng Nam dialect through a phonemic split of /a/, and (b) cluster reduction and feature preservation in the onset completed phonologization of the Quảng Nam /ɑ/.

5.1   Split of /a/

As shown earlier in (4), when the Hanoi short /ă/ precedes a glide (as in [măw³] 'blood' and [tăj¹] 'hand'), the glide disappears in Quảng Nam: [ma³], [ta¹]. The deletion of final glides after /ă/ in Quảng Nam is systematic and results in the reappearance of /a/ in Quảng Nam open syllables where it contrasts with the back /ɑ/: cf. Hanoi [kɑ³] 'fish' and [kăj¹] 'spicy' vs. Quảng Nam [kɑ³] 'fish' and [ka¹] 'spicy'. Final glide deletion is common in dialects spoken in central and south central Vietnam. For example, in some dialects in the Quảng Ngãi, Bình Định and Tuy Hoà provinces, a glide is deleted after /ɨ/, /ɤ/ or /ɨə/, e.g., Hanoi *trời* /cɤj²/ 'sky' corresponds to [ʈɤ²], Hanoi *mười* /mɨəj²/ 'ten' to [mɨ²] (Ngô M., 2017; Nguyễn T. H., 2016). These cases are very similar to those in Kẻ Chay dialect where Hanoi diphthongs [ɨə] and [uə], e.g, *cua* /kuə¹/ 'crab', *cưa* /kɨə¹/ 'to saw', *lúa* /luə³/ 'paddy' correspond to Kẻ Chay single vowels, [ku], [kɨ], and [lu] respectively.

The emergence of /a/ in Quảng Nam open syllable is likely to result from a split of /a/. Notice that the long /a/ in the two Hà Tĩnh varieties occurs only in closed syllables. When the conditioning environment disappeared in the Quảng Nam dialect (the loss of final glides) a phonemic split of /a/ was created. The long vowel /a/ emerged as a new phoneme in Quảng Nam open syllables, and so Quảng Nam /ta¹/ 'hand' is distinct from /tɑ¹/ 'we, inclusive' and in the presence of, for example, /tam¹/ 'tooth pick'.

Alternatively, the Quảng Nam /ɑ/ is developed internally from the split of a back, rounded vowel. However, this split is unmotivated because it did not happen with any other dialects.

5.2   Cluster reduction and feature preservation

In the Quảng Nam dialect, the back unrounded /ɑ/ contrasts with the central /a/ and other instances of /ɑ/ might have originated from feature preservation through simplification of syllable onsets. In this process, northern clusters composed of a consonant followed by /w/ reduced to a single segment. The consonant survives the process if it is alveolar or palatal (5a–c). The glide /w/ remains if the consonant is velar or glottal (5d–k). As expected, /ɑ/ contrasts with /a/ after a velar consonant

(5h–i, j). In (5k, l) the Hanoi short /ă/ before labial consonants corresponds to the Quảng Nam long /a/, as seen in syllables without the prevocalic /w/, e.g., Hanoi [năm] 'five' ~ Quảng Nam [nam].

(5) Cluster reduction in southern dialects

		HANOI	QUẢNG NAM	Gloss
a.	*chán*	cwaŋ^3	cɑŋ^3	'shocked'
b.	*thuế*	tʰwe^3	tʰe^3	'tax'
c.	*duyên*	zwiən^1	jiən^1	'charm'
d.	*huệ*	hwe^4	we^4	'lily'
e.	*oan*	ʔwan^1	waŋ^1	'wrongly accused'
f.	*hoa*	hwa^1	wɑ^1	'flower'
g.	*quên*	kwen^1	weŋ^1	'to forget'
h.	*qua*	kwa^1	wɑ^1	'to pass'
i.	*quay*	kwăj^1	wa^1	'to rotate'
j.	*quạu*	kwăw^4	wa^4	'irritated'
k.	*(lông) quặm*	kwăm^4	wam^4	'ingrown eyelash'
l.	*(oái) oăm*	ʔwăm^1	ʔam^1	'bizarre'

Cluster reduction is widespread in southern dialects. One characteristic of southeast Asian languages is that features seem to bounce back and forth between consonants and vowels (Matisoff, 1973). For example, in Bwe Karen, the labial feature freely moves between the cluster Cw and the vowel: lwi^1 ~ lu^1 'four', khwi^1 ~ khu^1 'nine', bwe^1 ~ bo^1 'how many', thwi^2 ~ chu^2 'dog' (Henderson, 1985, p. 3). Strikingly, in these examples when /w/ is removed from the cluster, a front vowel changes to a back vowel. In the Quảng Nam dialect, when a velar or glottal consonant is removed from the cluster, the feature [back] from the consonant is absorbed into the vowel /ɑ/. The preservation of [back] in cluster simplification in Quảng Nam can be viewed as a carryover type of assimilation in sound change. After reduction, the vowel in (5i–j) remains central in the Quảng Nam dialect. It could be because onset reduction occurred after final glide deletion, or examples such as (5i–j) could be a case of dominance of the coda, as in Mandarin. In Mandarin the vowel [ə] assimilates to the pre or post-nuclear glide. Where there is a conflict, assimilation to the coda glide prevails, e.g., /tuəi/ 'correct' > [twej], and /tiəu/ 'throw' > [tjow] (Lin, 2002). Another conflict in Mandarin is one between the fronting effect demanded by the coda [n], and the backing effect of the pre-nuclear [w] in syllables with the shape /Cuan/: the coda assimilation dominates over progressive assimilation (Wu and Kenstowicz, 2015). In the Quảng Nam dialect the coda offglides, /w/ and /j/, are more dominant than the pre-nuclear onglide /w/.

Because backing of central /a/ or splitting is a common sound change, can Quảng Nam /ɑ/ be accounted for strictly through internal development? Recall that

most early settlers in Quảng Nam came from Thanh Hoá (Đại Việt, 2001, p. 437; Khâm Định, 1998, p. 326). Phạm A. H. (2019) shows similarities in lexicon and grammatical constructions (e.g., shared personal and locative pronouns, particles, contracted forms, see Hoàng T. C., 1989; Phạm V. H., 1985, 1917), and pronunciation between the Kẻ Chay, Quảng Nam and Thanh Hoá dialects and suggests that the Quảng Nam dialect descended from the Thanh Hoá dialect and it integrated certain features from Hà Tĩnh dialects. Although the vowel /ɑ/ is absent in Thanh Hoá dialect, and central /a/ is seen in all syllable types, an analysis of a strictly internal development of the Quảng Nam /ɑ/ would assume the emergence of /ɑ/ is independent in Hà Tĩnh and Quảng Nam dialects, and would ignore completely the historical migration of Hà Tĩnh speakers and the presence of /ɑ/ in the Hà Tĩnh dialects. This hypothesis also treats other linguistic similarities between the Quảng Nam and Hà Tĩnh dialects, summarized in Table 4, as accidental. The internal development hypothesis would be worthy of consideration if a vowel similar to Quảng Nam /ɑ/ had not been discovered in those subdialects of Hà Tĩnh.[5]

The history of migration and similarities between the Hà Tĩnh, Quảng Nam, and Thanh Hoá dialects, including the vowel /ɑ/, provide strong evidence concerning the connection between the dialects of these regions that are geographically discontinuous.

6. Conclusions

This paper traces the Quảng Nam vowel /ɑ/ to its origin. A similar vowel is found in the Hến and Kẻ Chay dialects of Hà Tĩnh province. In these Hà Tĩnh subdialects, the backing of /a/ creates the back, unrounded vowel /ɑ/. The vowel was brought to Quảng Nam by early settlers from north central Vietnam. In the new land, Quảng Nam, a new dialect, was established. The Hà Tĩnh /ɑ/ brought by immigrants was integrated and became stable in Quảng Nam through linguistic restructuring. Although some features were lost in the process, the emergence of a new phoneme, /ɑ/, in the Quảng Nam dialect is one case where "speakers of various dialects colonized the new territories, brought their dialects into contact with each other to form a new variety" (Fernández-Ordóñez, 2012, p. 101). The Quảng Nam /ɑ/ is an example of a mixture of internal and contact factors in sound change..

The emergence of /ɑ/ in Vietnamese dialects can be explained through the dual nature of sound change (Kiparsky, 1995): (a) the various realizations of central /a/

---

5.   Observing some similarities between the Quảng Nam, Nghệ An and Hà Tĩnh dialects, Trần T. T. (2003) raises a general question whether early immigrants from Nghệ An brought their dialects with them to Quảng Nam.

in the Hến, Kẻ Chay and Quảng Nam dialects reflect phonetic variations inherent in speech, and (b) the selected variant ([ɑ]) is integrated into the linguistic system, and passed on to generations of speakers through acquisition.

Acknowledgements

References

Barker, M. & Barker, M. (1970). Proto-Vietnamuong (Annamuong) final consonants and vowels. *Lingua*, 24, 28–285.

Brunelle, M. (2009). Contact-induced Change? Registers in Cham dialects. *Journal of the Southeast Asian Linguistics Society*, 2, 1–22.

Đại Việt Sử Ký Toàn Thư. (2001). [Complete Annuals of Đại Việt]. *Finished in 1479 by Ngô Sĩ Liên, additional information by Vũ Quỳnh, Phạm Công Trứ, and Lê Hi in the 1697 edition*. Hanoi: Thời Đại published in 2011.

Đức, H. (2016). Ngôi làng nói tiếng kỳ lạ [A village speaking a strange dialect]. *VN Express*, Retrieved November 5, 2017. https://vnexpress.net/tin-tuc/thoi-su/ngoi-lang-noi-tieng-ky-la-3350582.html

Ferlus, M. (1995). Particularités du dialecte vietnamien de Cao Lao Hạ (Quảng Bình, Vietnam) [Peculiarities of the Vietnamese dialect of Cao Lao Hạ (Quảng Bình, Vietnam)]. *Dixièmes Journées de Linguistique d'Asie Orientale*. Paris. http://halshs.archives-ouvertes.fr/halshs-00922735

Ferlus, M. & Trần, T. D. (2013). Arem – French/English – Vietnamese A preliminary Lexicon. Paper presented at the 5th International Conference on Austroasiatic Linguistics at ANU, Canberra, September 4–5, 2013.

Fernández-Ordóñez, Inés. (2012). Dialect areas and linguistic change. Pronominal paradigms in Ibero-Romance dialects from a cross-linguistic and social typology perspective, in De Vogelaer, G. & Seiler, G. (Eds.) *The Dialect Laboratory* (pp. 73–106). Amsterdam/Philadelphia: John Benjamins.

Henderson, E. (1985). Feature Shuffling in Southeast Asian Languages. In S. Ratanakul et al. (Eds.), *Southeast Asian Linguistic Studies Presented to André-G. Haudricourt* (pp. 1–22). Institute of Language and Culture for Rural Development, Mahidol University.

Hickey, G. (1982). *Sons of the mountains: ethnohistory of the Vietnamese central highlands, 1954–1976*. New Haven: Yale University Press.

Hoàng, T. C. (1989). *Tiếng Việt trên các miền đất nước* [Vietnamese dialectology]. Hanoi: Khoa Học Xã Hội.

Ken DWT. (2004). Vietnam-Champa Relations and the Malay-Islam Regional Network in the 17th–19th Centuries. *Kyoto Review of Southeast Asia*, 5. Center for Southeast Asian Studies, Kyoto University.

Khâm Định Việt Sử Thông Giám Cương Mục. (1998). [History records of the Nguyễn Dynasty], *1884 by Quốc Sử quán triều Nguyễn*. Translated by Viện Sử học in 1960. Hanoi: Giáo dục published in 1998.

Kiparsky, P. (1995). The phonological basis of sound change. In J. A. Goldsmith (Ed.), *The handbook of phonology theory* (pp. 640–670). Oxford: Blackwell.

Lê, N. H. M. (2016). Âm sắc, trường độ và giải pháp cho hệ thống nguyên âm thổ ngữ Bình Định [Color, length, and solution for the vowel system in Binh Dinh]. *Ngôn Ngữ* [Journal of Vietnamese Linguistics], 11, 70–80.

Li, T. (1998). *Cochinchina, Southern Vietnam in the Seventeenth and Eighteenth centuries*. New York: Southeast Asia Program publications.

Lin, Y.-H. (2002). Mid vowel assimilation across Mandarin dialects. *Journal of East Asian Linguisitics*, 11, 303–347.

Maspero, H. (1912). Etude sur la phonétique historique de la langue annamite: les initiales [Studies in Annamese historical phonetics: initial consonants]. *Bulletin de l'Ecole Française d'Extrême-Orient*, 12, 1–127.

Matisoff, J. (1973). Tonogenesis in Southeast Asia. In L. Hyman (Ed.), *Consonant Types and Tone* (pp. 71–96). Los Angeles: USC.

Ngô, G. V. P. (2011). *Hoàng Lê Nhất Thống Chí*, witten by Ngô writers, i.e., first several chapters by Ngô Thời Chí (1753–1788). Hanoi: Văn Học.

Ngô, T. M. (2017). Đặc điểm chính âm và chính tả cuả ngôn ngữ tại địa phương Khánh Hoà và việc dạy học chính tả [Characteristics of phonetics and spelling of Khanh Hoa dialect]. *Paper presented at the 3rd International Conference in Vietnamese Linguistics*, April 24, 2017. Hanoi.

Nguyễn, T. C. (1997). *Giáo trình Lịch sử Ngữ âm tiếng Việt* [A course on Vietnamese Historical Phonology]. Hanoi: Giáo Dục.

Nguyễn, T. H. (2016). Một vài ghi nhận về thổ ngữ Tuy Phước, An Nhơn, Bình Định [Remarks on the An Nhon subdialect of An Nhon, Binh Dinh province], *Ngôn Ngữ* [Journal of Vietnamese Linguistics], 12, 22–32.

Nguyễn, V. T. (2005). *Ngữ âm tiếng Mường qua các phương ngữ* [The phonetics of the Mường language across its various dialects]. Hanoi: NXB Từ điển Bách khoa.

Ohala, J. (1989). Sound change is drawn from a pool of synchronic variation. In L. E. Brevik and E. H. Jahr (Eds.), *Language Change – Contributions to the Study of Its Causes* (pp. 173–198). New York: Mouton de Gruyter.

Phạm, A. H. (1997). The coronal-velar relationship in Vietnamese dialects. Master thesis, University of Toronto.

Phạm, V. H. (1985). Về một số đặc trưng của tiếng Thanh Hoá, thổ ngữ chuyển tiếp giữa phương ngữ Bắc Bộ và Bắc Trung Bộ [Remarks on characteristics of Thanh Hoá dialect, an intermediate dialect between dialects of north and north central regions]. *Ngôn Ngữ* [Journal of Vietnamese Linguistics], 4, 53–56.

Phạm, V. H. (2003). Tổ hợp phụ âm tl- ở thổ ngữ Đức An, huyện Đức Thọ tỉnh Hà Tĩnh [Initial cluster tl- in Duc An subdialect of Đức Thọ, Hà Tĩnh]. *Ngôn Ngữ* [Journal of Vietnamese Linguistics], 1&2, 14–17.

Phạm, V. H. (2017). Dấu vết cuả tiếng Thanh Hoá trong phương ngữ Nam Trung bộ trở vào [Traces of Thanh Hoá dialect in dialects from south central Vietnam southward]. Paper at *3rd International Conference in Linguistics*, Hanoi, Vietnam, April 22.

Phạm, A. H. (2014). Ngôn ngữ biến đổi và số phận của nguyên âm /a/ trong giọng Quảng Nam [Sound change and the destiny of /a/ in the Quảng Nam dialect]. *Ngôn Ngữ* [Journal of Vietnamese Linguistics], 6, 10–18. Hanoi.

Phạm, A. H. (2016). Sự biến âm trong vần tiếng Việt: thổ ngữ làng Hến, huyện Đức Thọ, tỉnh Hà Tĩnh [The Hến dialect of Đức Thọ District, Hà Tĩnh]. *Ngôn Ngữ* [Journal of Vietnamese Linguistics], 11, 7–26. Hanoi.

Phạm, A. H. (2019). Vietnamese dialects – a case of sound change through contact. In Duffel, Nigel, Trang Phan, and Tue Trinh (Eds.), *Interdisciplinary perspectives on Vietnamese Linguistics*, pp. 31–66. Amsterdam: John Benjamins.

Phan, K. (1969). *Việt sử xứ Đàng Trong* [History of Đàng Trong]. Saigon: Khai Trí.

Po, D. (1987). *Le Panduranga (Campa) 1802–1835. Ses rapports avec le Vietnam*, 2 vols. Paris: École Française d'Extrême-Orient.

Shimizu, M. (2013). Vị trí cuả tiếng Quảng Nam trong quá trình biến đổi âm cuối gốc lưỡi [The role of Quảng Nam dialect during the process of sound change of final velar consonants]. Paper at the *2nd International Conference on Linguistics*, Hanoi.

Tohyama, E. (2015). Tìm hiểu nét đặc trưng về mặt ngữ âm cuả thổ ngữ Quảng Nam tại Hội An và tìm hiểu dấu vết cuả nó trong cuốn Từ điển Việt – Bồ – La [Examining the phonetic characteristics of Hoi An subdialect of Quảng Nam in the Vietnamese-Portugese-Latin dictionary]. Paper given at *Chữ Quốc Ngữ – sự hình thành, phát triển và những đóng góp vào văn hoá Việt Nam* [Quoc Ngu Script- establishment, development and contributions to Vietnamese culture], October, Phú Yên.

Trần, T. D. (2005). *Giáo trình lịch sử tiếng Việt* [Lectures on the history of the Vietnamese Language]. Hanoi: Nhà Xuất Bản Đại học Quốc gia Hà Nội.

Trần, T. T. A. (2015). Hệ thống vần cái trong thổ ngữ Lý Sơn (Quảng Ngãi) [Vowel system in Ly Son dialect of Quang Ngai]. Master thesis. Ho Chi Minh City: KHXH & NV University.

Trần, T. T. (2003). Ngữ âm tiếng Tam Kì – Quảng Nam [Tam Ky subdialect of Quảng Nam], *Ngôn Ngữ và Đời sống* [Language and Life], 7, 11–15.

Vũ, T. (2012). Người Nghệ nói "tiếng Nghi [Nghe people speak the Nghi dialect]. *Tuoi Tre Online*, Retrieved 1/11/2012. http://tuoitre.vn/tin/chinh-tri-xa-hoi/phong-su-ky-su/20121101/nguoi-Nghệ-noi-tieng-nghi/518443.html

Vương, H. L. (1998). Vài nhận xét về đặc điểm cuả vần trong thổ âm Quảng Nam ở Hội An [Some remarks on characteristics of the rhymes of Hoi An subdialect, Quảng Nam]. *Một số vấn đề Ngôn ngữ học Việt Nam* [Some issues in Vietnamese Linguistics] (pp. 311–319). Hanoi: Đai học và Trung học chuyên nghiệp.

Vương, H. L. & Hoàng, D. (1994). *Ngữ Âm Tiếng Việt* [Vietnamese phonology]. Hanoi: Giáo dục.

Wells, J. C. (1982). *Accents of English*. New York: Cambridge University.

Wu, F. & Kenstowicz, M. (2015). Duration Reflexes of Syllable Structure in Mandarin. *Lingua*, 164, 87–99.

Appendix. Rhyming words in Hà Tĩnh

CV(G)

[i] *bí, ly, khỉ;* [e] *dế, ghế, dê, lê, khế;* [ɛ] *me, xe;* [ɨ] *lư, sư tử, sư;* [ɤ] *bàn thờ, bơ, chợ, mõ;* [ɑ] *ga, gà, ca, cá, cà, cả, bàn là, lá, mạ, mặt nạ, số ba, sả, tã;* [u] *cú, mũ;* [o] *xe thồ, tô, gỗ, hổ;* [ɔ] *chó, nho, đỏ*

[iə] *kiến, giếng, kiếm;* [ɨə] *dưa, dừa, cưa, mười, bướm, giường, lược;* [uə] *cua, luá, đuā, chuồn, cuốc, chuột*

[-j]: *bài, hai, mai, tai, tay, ăn mày, maý may, lông mày, bảy, xe máy, còi, sói, tỏi, voi*

[-w] *áo, táo, bāo, đào, sao, sáu, cau, máu, sâu, cá sấu, trâu, heo, mèo, lều, phễu*

CwV(C): *loa, hoa, đoàn, khăn quàng, cái khoan, khoai, quà, toán, dây chuyền, hột xoàn, cứu hoả*

CVC

[i] *tím, chín, vịt, lịch, xích, đinh, lính, chích*

[e] *nệm, sên, hến, nến, tết, bến xe, bệnh nhân, ếch, cung tên, lên*

[ɛ] *đèn, két, sen, chén, sét, kẻng, xẻng*

[ɑ] *tám, rạm, cam, thàp chàm, xe đạp, dây cáp, ngàn, bàn, bát lan, đàn, nhà sàn, cái sàng, hồng vàng, càng, khoai lang*

[a] *năm, mắm, trăm, cằm, tắm, bắp, cặp, trăn, ăn, sắn, rắn, khăn, mắt, banh, cánh, gánh, gạch, hành, nanh, sách, xanh, xe tăng, măng, xăng, trắng, trăng, răng, lăng*

[ɨ] *mực, mứt, sừng*

[ɤ] *cơm, ớt, lợn*

[ɤ̆] *hấp, mâm, ấm, nhẫn, Phật*

[u] *xúp lơ, bún, lụt, thúng*

[o]: *cốm, tôm, cốp, bốn, rồng, sông, ngỗng, móng tay, thùng xốp*

[ɔ] *cọp, phòng họp, ngọt, nón, bọt, tóc, lọng, ong*

The future markers in Palestinian Arabic

Internal or external motivation for language change?

Duaa AbuAmsha

University of Calgary

The main objective of this research is to explore the factors involved in the variable formation of the future in Palestinian Arabic (PA), and to examine whether it represents a case of ongoing grammatical change. The results show that middle-aged and younger Jaffan and Gazan speakers living in Gaza express future time reference in ways that are absent in the Jaffan dialect still spoken in Jaffa and are also not characteristic of the speech of older Jaffan and Gazan speakers in Gaza. Middle-aged and younger Gazans and Jaffans living in Gaza primarily express 'future' within the verbal paradigm by means of a prefix ḥa- attached to the non-past stem. The study concludes that both internal and external factors are involved in the development of the future marker in PA.

Keywords: Palestinian Arabic, future time reference, language change and variation, internal factors, external factors, Gazan dialect, Jaffan dialect

1. Introduction

The current paper explores expressions of future time reference in Palestinian Arabic (PA). Different speakers of PA spoken in Gaza City express future time reference in morphologically distinct ways; further, these differences correlate with speakers' dialect background and age. The pilot study described here is designed to answer the following research questions: does the variation in the expression of future time reference represent a case of ongoing grammatical change in PA? If so, to what extent are internal and external factors involved in the change? What is the role played by the individual speaker?

To establish the range within which PA dialects spoken in Gaza City vary in their morphological expression of future time reference, data were collected using interviews from speakers of two urban dialects: Gazan and Jaffan. For the study, sixteen female speakers representing four different age groups were interviewed,

eight speakers with Gazan dialect background and eight with Jaffan dialect background. The recordings were transcribed, and the verb forms which emerged were organized into paradigms according to their morphological form and function. A control group of four Jaffan dialect speakers who still live in Jaffa, Israel, were also interviewed using the same questions.

The results show that the verbal morphology of the middle-aged and younger Jaffan and Gazan speakers living in Gaza exhibits properties with respect to the way they mark 'future' that are absent in the Jaffan dialect still spoken in Jaffa and are not characteristic of the speech of older Jaffan and Gazan speakers in Gaza. Data from Jaffan speakers still living in Jaffa show that future time reference is expressed outside the verbal paradigm by means of a particle *ra:ḥ*, or the quasi-verb *bidd-* 'want'. However, middle-aged and younger Gazans and Jaffans living in Gaza primarily express 'future' within the verbal paradigm by means of a prefix *ḥa-* attached to the non-past stem.

To account for the variation in the expression of future time reference among PA speakers, I adopt the code-copying framework as developed by Johanson (1999, 2002, 2008), according to which internal factors are conceived as tendencies that may become the object of external factors of change. It recognizes the role of internal factors, but at the same time stresses the role of extra-linguistic profiles of speakers in adopting and developing innovative forms as discussed in the next section. I also draw on ideas from Matras's framework for language contact and the role of the individual speaker (2009). My analysis also relies on the principles of grammaticalization theory (cf. e.g. Hopper and Traugott, 2003; Bybee et al. 1994).

The paper argues in favour of a contact-induced approach to language change which has been shown to be appropriate when applied to other languages, particularly the Anglophone world, and has been adopted in different contexts. However, there is less research examining the role of contact in language change in Arabic-speaking contexts. Moreover, a review of the literature shows that most, if not all, studies on dialect contact and change in Arabic-speaking settings have focused on the role of contact in phonological change, for example, Al Wer (1991, 2002) on dialect contact between PA and Jordanian Arabic in the City of Amman, and Horesh (2014), which studies the shift in the phonemic inventory of PA spoken in Jaffa due to contact with Hebrew. To my knowledge, no research has examined the role of contact in morphological change in Spoken Arabic. Thus, the current study will shed needed light on the effects of dialect contact on morphological systems in general and in the Arabic-speaking world in particular.

The structure of the paper is as follows: Section 2 introduces background information about Arabic, the linguistic, social, and demographic situation in Gaza City, the setting of the field study, and an overview of the PA Arabic verb. Section 3 is dedicated to the topic of language change, dialect-contact, and the code-copying

framework adopted in this paper. In Section 4, I outline the study design. In Section 5, I present my findings about the expression of future time reference in urban PA. In Section 6, I examine the influence of dialect contact on the development of a future marker in Gazan and Jaffan dialects spoken in Gaza City and propose an explanation for this development. In Section 7, I conclude by discussing the results, considering the question of linguistic vs. external and extra-linguistic factors and predictors of contact-induced language change.

2.   Background information

This section provides background information about the Arabic language in general, spoken Palestinian Arabic, the Arabic verb, and the verb morphology of PA. I also describe the historical and current linguistic, social, and demographic situation in Gaza.

2.1   The Arabic language

The term "Arabic", a branch of the Semitic language family, refers to a set of linguistic systems, which despite manifesting substantial differences at different linguistic levels, exhibit sufficient mutual homogeneity to be classified as varieties or dialects of a single language. A distinction is made between Standard Arabic (SA) and spoken Arabic. Modern Standard Arabic (MSA) is the standard formal variety of Arabic in use today in writing and formal speech across the Middle East and North African countries. Colloquial or spoken Arabic consists of a number of local dialects which exhibit various linguistic features that make them distinguishable and even mutually unintelligible in extreme cases. These vernaculars are used for everyday speech and learned at home as first languages, while the formal language is learned later at school. Palestinian Arabic (PA), the dialect under investigation here, is a dialect subgroup of Levantine Arabic. In colloquial PA, a distinction is made between urban, rural, and Bedouin spoken dialects. The Jaffan and Gazan dialects under study are to be considered as urban PA dialects.

2.2   Socio-historical, linguistic, and demographic situation
in the Gaza City context

To understand the current linguistic situation in Gaza and to establish what makes Gaza an ideal context for contact studies, this section provides a review of the historical political shifts and their reflections on the demographic composition and

social structure of the Palestinian community. The context in Gaza City, the largest urban centre in the Gaza Strip, is very complex and can be described as a special socio-linguistic situation characterized by intensive contact between different dialects over a period of 70 years.

In 1947, Gaza, like other parts of Palestine, had a stable social structure and population. However, following the Arab-Israeli war of 1948 and the displacement of thousands of Palestinians from their homeland, and following the establishment of the State of Israel in 1948–1949, the Gaza Strip began to witness great change, as a large part of the Palestinians displaced from their homes in what is now the State of Israel arrived in the Gaza Strip as refugees. The arrival of about 200,000 Palestinian refugees from outside the Gaza Strip has had dramatic demographic effects on the composition of the Gazan population. These changes included the establishment of refugee camps by UNRWA where Palestinian refugees settled, the arrival of new family groups, new social relations, and new dialects. Palestinian refugees displaced to the Gaza Strip brought with them their original dialects, both urban and rural, and came into direct contact with the Gazan dialect.

According to the latest figures from the Palestinian Central Bureau of Statistics (PCBS) in 2015, about 1.85 million Palestinians (a population density of 5068 people per km^2) live in the Gaza Strip. About 67%[1] of the population in the Gaza Strip are refugees who were displaced to the Gaza Strip in 1948, with the largest number of refugees coming from Jaffa and Ashkelon and surrounding villages. About 625,000 Palestinians, comprising both Gazans and refugees, live in Gaza City in an area of 56 km^2, with a population density of 11,175 per km^2. Today, Gazans and refugees and their descendants live side by side in the same neighbourhoods and engage in face-to-face interaction in all aspects of life at work, schools, universities, hospitals, markets, etc. Also, at the social level, the two groups are not separated, and social ties between them are increased by marriage.

2.3   Morphology of the Arabic verb

This paper focuses on the expression of the future in the PA verb. The two main Tenses/Aspects are traditionally known as "perfective/past" and "imperfective/present". Finite verbs in PA show agreement inflection for three persons, two genders (in both the second and third persons) and number (singular and plural). Regarding formal expression of agreement, the "perfective/past" employs suffixes only, while

---

1.   http://www.wafainfo.ps/atemplate.aspx?id=2357 (Website of the National Palestinian Information Centre Wafa; http://www.pcbs.gov.ps/site/512/default.aspx?tabID=512&lang=ar&ItemID=1420&mid=3915&wversion=Staging (Website of the Palestinian Central Bureau of Statistics (PCBS)).

the "imperfective/non-past" employs prefixes marking Person and suffixes marking Number and Gender. Following Holes (2004), I use the labels "suffix-stem" (s-stem) and "prefix-stem" (p-stem), respectively,

3. Contact and language change

Broadly speaking, contact linguistics is concerned with the long-term linguistic consequences that result when speakers of different languages or dialects come into contact (Trudgill, 1986). For decades, most attempts to resolve the question of why languages change have looked for internally-motivated explanation. Change is seen as an internal characteristic of language; therefore, explanations can ignore all other factors (see Farrar and Jones, 2002, p. 1–8). However, since the second half of the 20th century, increasing attention has been devoted to the role contact can play in language change, beginning with Weinreich (1953) and continuing with Thomason and Kaufman (1988) and Trudgill (1986), after which the treatment of dialect contact was considered on a par with that of languages in contact (Hickey, 2010, p. 5).

3.1 The code-copying framework and the dialect contact situation in Gaza City

The code-copying framework as proposed by Johanson (1999, 2002, 2008) is a unified model that aims to deal with phenomena such as borrowing, transfer, adaptation, convergence, levelling, koinéization, shift, etc. under one umbrella (Johanson, 2002, p. 285). Code-copying results from the interaction of linguistic codes whereby linguistic elements are copied from one code to another. The main working assumption I adopt from the framework is that language change and the historical development of copies are code-internal, occurring in a specific code. Linguistic changes may be triggered by code-external factors which include contact or extra-linguistic psychological and social factors, in specific socio-political situations. By contrast, other changes occur without external or extra-linguistic motivation, that is, due to purely internal factors (Johanson, 2002, p. 286). Within this framework, internal factors are thought of as inherent tendencies of systems where the structural properties of linguistic elements make them more or less likely to undergo change. In cases where linguistic tendencies seem to give rise to a change, extra-linguistic social or psychological factors are important for the outcomes of the change (Johanson, 2002).

The linguistic elements copied in a contact situation can be units such as segments, morphemes, and words or their phonological, combinational, semantic, and

frequential properties. Copying of entire units is called global copying, by which the whole form and function of a unit is copied. Copying of properties is referred to as selective copying. The code-copying model accounts for both variational patterns and developmental stages of the structures involved (Johanson, 2002, p. 288). Synchronically, the framework examines the complex variational patterns of "donor" and "recipient" languages and dialects and highlights the role of the processes of "adoption", "imposition", and "shift". In the diachronic dimension, the model considers the developmental stages of the linguistic structures along the lines of habitualization and conventionalization that pertain to extra-linguistic developments that may result in complex language change over time (Johanson, 1999).

4. Study design

The procedure followed in this study can be summarized as follows: recruiting participants, collecting, transcribing, and arranging data, and identifying patterns, arranging them as paradigms, identifying the meanings they encode, labelling them, and comparing them across speakers. Once the forms and their basic approximate meaning and place in the system of contrasts were determined, my next step was to refine the material at two levels: comparing data from speakers from different age groups but with the same dialect background, and then comparing speakers in a particular age group with their counterparts in the same age group from the other dialect background.

4.1 Recruiting participants, interview, data collection

To investigate the expression of future time reference in Jaffan and Gazan dialects and whether a change is taking place in them, I compared samples of the natural speech of sixteen speakers of the two dialects. I prepared an interview script of six questions to stimulate participants to use whatever grammatical structures encoding 'future' that might be available to them. For example, participants were asked questions about their expectations for life in the future and things they plan to do in the near or remote future. The participants' answers were audio-recorded, and then transcribed and coded for analysis.

Participants were selected from the two dialect backgrounds under study, Gazan (G) and Jaffan (J) dialect, and from four different age groups: 8–17, 18–39, 40–65, and > 66. Investigating the speech of speakers from different age groups makes use of the apparent-time theoretical construct of language change in which language change is based on the distribution of the linguistic variable across age

groups in a speech community (cf. e.g., Labov, 1994, 2001; Eckert, 1998). If the results reflect variation within different age groups, this can be an indicator of either linguistic change in progress where variation over a broader range of ages is observed, or just an age-graded variation, by which there is a stable variation within a population based on age, and which does not necessarily indicate actual language change. In addition to the sixteen speakers who live in Gaza, the study also included four female speakers from different age groups who still live in Jaffa. These Jaffa Jaffan (JJ) participants serve as a control group against which the speech of the speakers living in Gaza is compared. Interviewing people who still live in Jaffa and have not experienced intensive contact with the Gazan dialect or any other Palestinian dialect is important to address these two questions posed by Heine and Kuteva (2005, p. 22): "What evidence is there for transfer to have taken place? Could that change have taken place without involving language contact?"

5. Findings: Future time reference within and outside the verbal paradigm in PA

This section of the paper presents findings from the field study of the expressions of future time reference in PA. It identifies how the future is expressed in the finite verb system. It also shows where the verb paradigm can be augmented by complex constructions allowing more distinctions. The data indicate a major two-way temporal contrast in the PA verb form between the 'non-past' associated with the prefix-stem and 'past' associated with the stem-stem.

Based on the simple finite verb forms appearing in the data, there appear to be two verb paradigms operating in PA, as shown in Figures 1 and 2.
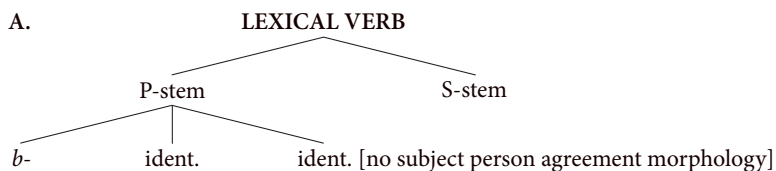
Figure 1. Verbal Paradigm A in PA[2]

---

2. For agreement morphology of the PA verb, see Section 2.3. The abbreviation 'ident' is used to mean that two of the mood forms (imperative and subjunctive) built on the p-stem are identical to it as they are expressed by the absence of any formal change to the p-stem.

B. **LEXICAL VERB**

P-stem   S-stem

b-   *ḥa-*   ident.   ident.[no subject person agreement morphology]
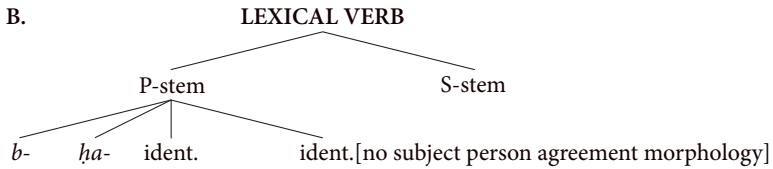
Figure 2. Verbal Paradigm B in PA

Paradigm B is distinct from Paradigm A in that it has one more contrast within the forms building on the prefix-stem: a form prefixing *ḥa-*. Within paradigm structure A, the data show that the non-past tense as marked on the prefix-stem refers to situations which hold at the moment of speech (1) or will hold subsequent to it (2).

(1) *kul-hum   b-yi-zraʕu: ʤanb ʔil-bait   yaʕni fi*
     all-3PL.MASC INDIC-3-plant.NON-PAST-PL.MASC besides the house that   is
     ʔil hadiqa   ta:ʕit ʔil bait   (GJW1)
     in the garden POSS   the house
     'All of them [the countryside residents] grow plants in the house's garden.'

(2) *kula:hi:t-ha lamudit   ʔusbu:ʕain zay riḥleh   wi*
     all-3SG.FEM for period week.DUAL as   trip-FEM and
     b-n-irjaʕ   (GJH3)
     INDIC-1PL-return.NON-PAST
     'The whole thing is a trip for two weeks, and we will return back again.'

In addition to marking future time reference in the simple verb paradigm, Type A, there are many complex constructions in PA that also allow speakers to express future time reference. For example, the volitive quasi-verb *bidd-* 'want' + p-stem of the lexical verb is a complex construction that can express future time reference. Note that while an element of volition is evident in (3), it is absent in (4).

(3) **bidd-i** *a-ṭlaʕ   yoam ʔil-sabt*   (GJH3)
     want-1SG 1SG-go out.NON-PAST day   Saturday
     'I **will** go out on Saturday.'

(4) *miš ʔa:dr-ah a-tṣawar   kaif **bidd-uh***
     NEG able-FEM 1SG-imagine.NON-PAST how want-3SG.MASC
     y-ku:n   ʔil-ʕalam baʕd ʕišri:n sanah   (GK2)
     3SG.MASC-be.NON-PAST the world after twenty year
     'I could not predict how the world **will** be after twenty years.'

The particle *ra:ḥ*,[3] originally from the verb 'go', as will be discussed in Section 6 of the paper, is used by some speakers in a complex construction involving prefix-stem forms to indicate future time reference as in (5) and (6). Note that while the semantic elements of locomotion and agency are evident in (5), they are absent in (6).

(5)   ***ra:ḥ*** *ti-rdʒaʕ-i*   *ʕa-ġazeh*   (JN1)
      FUT. 2-return back.NON-PAST-SG.FEM on-Gaza
      'Will you return back to Gaza?'

(6)   *yoam ʔildʒumʕah ʕina*   ***ra: ḥ*** *y-ku:n*   *ʕawaṣif w maṭar*
      day   the Friday at-1PL FUT. 3SG.MASC.be.NON-PAST storm.PL and rain
      šadi:d   (JF2)
      heavy
      'Next Friday, there **will** be storms and rain [according to the weather forecast].'

For some speakers, the particle *ra:ḥ* has further grammaticalized into a prefix *ḥa-*. Example (7) shows the use of the new prefix.

(7)   *ʔana* **ḥ-a-ḍalni**   *fi-l-baladiyeh*   *zayma ana*   (GJR1)
      I   FUT-1SG-stay.NON-PAST in the municipality as   I
      'I will stay with my work in the municipality as I am (I do not have other plans).'

With the emergence of the prefix *ḥa-* as a future tense prefix in PA, a new member is added to the simple verb paradigm: the Tense system of PA is developing a three-way Tense contrast moving from 'a past'- 'non-past' contrast to a three-way distinction of 'past', 'present', and 'future', as modelled in Paradigm B in Section 5. Data from different speakers as shown in Table 1 below show that this shift in the PA verbal paradigm is still taking place and is not yet complete.

     The next section of this paper will examine whether variation in the expression of future time reference among PA Jaffan, Gaza Jaffan, and Gazan speakers is a manifestation of a wider process of ongoing language change in PA. I also investigate whether these differences interact with social factors such as speaker age, dialect background, and social network. Using data from the 20 speakers of urban PA interviewed in this study, I examine the role of linguistic factors, contact, and extra-linguistic factors in language change, and how they relate to one another.

---

3.   Mitchell and Al Hassan (1994, p. 24) use *ḥa-* as a unitary means of indicating a large set of related forms which includes *ra:ḥ*, *rayyiḥ* 'ACT-PART 'go' and its variants with different number and gender agreements. However, they do not propose how these forms are historically related.

6.   Dialect contact and language change in PA

Section 5 outlined the available morphological and extra morphological means of expression of future time reference of 20 speakers of PA. The rest of this paper will focus on the development of the future particle *ra:ḥ* from the lexical motion verb *ra:ḥ* and its later development into the prefix *ḥa-*. The goal is to investigate whether the observed variation points to ongoing linguistic change in the Gazan and Jaffan dialects spoken in Gaza City. If yes, are these changes to be traced back to independent internal linguistic tendencies and development and/or to dialect contact?

6.1   Development of future markers in PA

In this section, I examine the historical development of the verb *ra:ḥ* into a future marker in PA and the role played by internal linguistic factors, individual speakers, and contact. PA shows great variation in the way future time reference is expressed, as indicated in Section 5. Further variation that correlates with the speaker age and dialect background is also observed. Below is a summary table of the different forms PA speakers in the present study use when describing situations occurring subsequent to the moment of speech. It only intends to show the variation among the speakers and the relative frequencies of the available options for each speaker and makes no claim for statistical significance since it is based on a small dataset.

Table 1.   Forms available for future time reference in urban PA

Dialect	Non-past stem	*Bidd-*	Particle *ra:ḥ*	Prefix *ḥa-*
Jaffa Jaffan Dialect	15	19	31	–
Gaza Jaffa Dialect				
8–17	1	4	–	32
18–39	5	2	10	31
40–65	1	2	15	–
> 66	10	14	13	–
Gaza Dialect				
8–17	–	5	–	27
18–39	1	4	6	25
40–65	2	13	6	25
> 66	9	10	–	–

Based on Table 1, the variation among PA speakers in the means available to them for the expression of future time reference can be described as follows:

1. For Jaffa Jaffan (JJ) speakers, there are no examples of morphological marking of future time reference as distinct from 'non-past'. Extra-morphological[4] forms of expression are available, i.e. the particle *ra:ḥ* and the quasi-verb *bidd-* 'want' + p-stem.
2. For Gaza Jaffan (GJ) speakers, Group 4 (> 66.) is identical to the JJ speakers in expressing future time reference. Only GJ speakers in age groups 1 and 2 use the prefix *ḥa-* with fewer examples of the non-past (*b-* + p-stem), *bidd-*, and *ra:ḥ* + p-stem.
3. For Gazan (G) speakers, only the oldest group uses the 'non-past' (inflected p-stem) when referring to future situations within the verb paradigm, and the quasi-verb *bidd-* outside of it. Otherwise, G speakers use the prefix *ḥa-* to mark future reference within the paradigm, and primarily the quasi-verb *bidd-* outside of it.

Based on comparison of the data, two possible scenarios of development are suggested:

1. Gazan usage served as a model for Jaffans who came into contact with G speakers and copied the Gazan prefix *ḥa-* into their basic code. The copy of *ḥa-* is inserted into the simple verb paradigm where there had previously been no specific distinction of future time reference possible.
2. Jaffan dialect served as a model code for the Gazan dialect. The lexical motion verb *ra:ḥ* had not yet developed into a grammaticalized future marker in the dialect of Gaza by the time of contact. In the Jaffan dialect, *ra:ḥ* was in the midst of a grammaticalization process by which it had become a particle marking 'future'. Through a process of selective grammatical copying (see Section 3.1) G speakers adopted the grammatical function of *ra:ḥ* as a future marker on the model of the Jaffan dialect and inserted it into their own basic code, followed by change in the syntactic behaviour of the verb.

The second possibility is the one I adopt here and base my analysis on. This decision is based on the absence of *ra:ḥ* or *ḥa-* for the expression of future time reference in the samples of the two oldest Gazan speakers (70, 75 y). This finding suggests that *ḥa-* was not part of the Gazan dialect; it is therefore more likely that the Gazan dialect speakers took Jaffan usage as a model at the time of initial contact and copied *ra:ḥ* into their basic code. The development of a future marker in PA spoken in Gaza may be accounted for if we assume that the development occurred in the following stages.

---

4. I assume that *ḥa-* is a prefix and part of the morphology, while ra:ḥ and *bidd-* are particles available in the lexicon, and represented in the syntax.

Stage I: Grammaticalization of the lexical verb ra:ḥ *into a grammatical particle*
The source of the development of future markers *ra:ḥ* and *ḥa-* in urban PA spoken in Gaza today can be traced back to a grammaticalization process that took place in the Jaffan dialect. Hopper and Traugott (2003, p. xv) define grammaticalization as "the change whereby lexical items and constructions come in certain linguistic contexts to serve grammatical functions and, once grammaticalized, continue to develop new grammatical functions". Parallel to universal pathways of development, the lexical motion verb *ra:ḥ* 'go' developed into a future marker in the Jaffan dialect. Bybee et al. (1991, 1994) find that movement verbs are more frequent lexical sources for 'future' than lexical verbs or material of any other type. The source meaning for movement features is that 'the agent is on a path moving toward a goal'. Thus, this path of development for the verb *ra:ḥ* 'go' into a future marker in PA is as follows:

(8)   Movement towards a goal > intention > future

The movement towards a goal and the intentional meaning led to the development of 'go' into future meaning. That is, since the intention to do something is often realized in a period subsequent to the moment of speech, the future meaning is inferred in these structures. According to Bybee (2003), habituation which results from repetition is instrumental in the development of grammatical elements. The semantic force of the expression that is repeated frequently is weakened as it loses specific features of its meaning and is used in more contexts which makes it more subject to grammaticalization (Bybee, 2003, p. 605). The future marker in PA presumably developed from constructions as in (9) where the verb *ra:ḥ* is followed by another verb.

(9)   *ra:ḥ-at*   *ʃaša:n ti-ʕmil*   *ʕumrah*   (JL3)
       go.PAST-3SG.FEM to   3SG.FEM-do.NON-PAST pilgrimage
       'She went to (Mecca) to do pilgrimage.'

The mechanisms of inference, extension, and habituation can account for the rise of future meaning for the verb *ra:ḥ* in PA. First, the hearer infers that the speaker's intention is to be realized in the future. The more frequently the verb *ra:ḥ* is used, the more the inferred future meaning is enhanced in the speaker's mind and the more connection, based on this context-induced reinterpretation, is established between these constructions and the future (Hopper and Traugott, 2003). As this connection is established, the verb *ra:ḥ* is extended to new contexts, as in (10), where no intention is evident.

(10) *yoam ʔil-dʒumʕah ʕina*   *ra:ḥ y-ku:n*   *ʕawaṣif*   (JF2)
       day   the-Friday at-1PL FUT. 3SG.MASC.be.NON-PAST storm.PL
       'Next Friday, there will be storms [according to the weather forecast].'

As the use of *ra:ḥ* is extended to include a future meaning, it has undergone de-categorialization (Heine and Kuteva, 2005). Its status has changed from a content word to a function word, i.e. from a member in an open lexical class of verbs to a functional marker that is a member of a closed-word class and that no longer inflects for agreement.

Stage II: Copying and conventionalization of ra:ḥ *into the Gazan dialect*
I claimed earlier in this section that Gazan speakers copied the property of *ra:ḥ* as a future marker into their basic code on the model of the Jaffan dialect. Elements copied from one code to another may start as momentary and sporadic ephemeral instances of copying, the result of singular individual dynamic acts (Johanson, 1999, p. 47, 2008). However, with more acceptance from other members in the speech community, copies may become used habitually, with various degrees of recurrence in the individual. Copies may also become more or less conventionalized. Conventionalization is the integration with respect to acceptance in the speech community (Johanson, 2002, p. 299). Table 1 above shows different degrees of acceptance of the innovative forms *ra:h* and *ḥa-*. While they are more conventionalized and used habitually by middle-aged and younger speakers, they are not used by older Gaza Jaffan and Gazan speakers.

Stage III: Further grammaticalization of the particle ra:ḥ *into the prefix* ḥa-
In a later more advanced stage of grammaticalization, the future particle *ra:ḥ* has further become a prefix *ḥa-* that attaches to the p-stem of the lexical verb as in (11).

(11) *ʔana ḥ-a-ḍalni   fi-l-baladiyeh   zayma ana*   (GJR1)
      I   FUT-1SG-stay.NON-PAST in-the-municipality as   I
      'I **will** stay with my work in the municipality as I am (I do not have other plans).

As it developed into a prefix *ḥa-*, *ra:ḥ* has undergone loss of phonetic substance and decategorialization. While the lexical verb *ra:ḥ* still retains its full form, the particle *ra:h* was reduced to the prefix *ḥa-*. According to Bybee (2003), such changes are hastened by the increasing frequency of use of grammaticizing constructions as they generalize to a wider range of contexts. As PA speakers got together in larger numbers when thousands of Palestinian refugees arrived Gaza in 1948, this intensive and extensive contact created a context for a wider and more frequent interaction and use of salient linguistic forms with the form *ra:ḥ* probably being one of these forms. Also, in being reduced to an affix, the future particle has further lost its morphosyntactic properties. As noted in Section 5, this has major consequences for the structure of the simple verb paradigm.

6.2   Discussion

The further grammaticalization of the future marker in PA from the particle *ra:ḥ*
into a prefix *ḥa-* supports the idea that once a linguistic element is copied, it auto-
matically undergoes internal development (Johanson, 2002). However, the crucial
question to ask is whether this development of the future marker in PA spoken in
Gaza City could have taken place without contact. This possibility cannot be ruled
out as this development of a future marker from a motion verb is a path attested
in many languages (see Section 6.1). However, we need to consider the following
two observations from the data. First, if we assume that the particle *ra:ḥ* and its
reduced form *ḥa-* in the urban PA spoken in Gaza City are the outcome of a purely
internal development, how can we account for the variation among Gaza Jaffan and
Gazan speakers in different age groups with respect to these two forms? Gaza Jaffan
and Gazan speakers show different degrees of acceptance of *ra:ḥ* and *ḥa-*, which
interacts with speaker age and dialect background. Second, if the development of
the future marker in PA spoken in Gaza from a lexical verb is a purely internal
development process, why has not the Jaffa dialect spoken in Israel undergone
a similar process? The fact that it has not may be a piece of evidence to support
the claim that grammatical change is more likely to occur and that the extent of
grammaticalization tends to be greater in a situation of extensive contact over time
(Heine, 2012). The next section of the paper will examine how variation among PA
speakers in the expression of future time reference interacts with external factors.

6.3   Extra-linguistic factors

Although the development of new markers to express future in PA spoken in Gaza
could be a purely linguistic development, the data from speakers of different age
groups suggest that the developments depend on speaker sociolinguistic charac-
teristics. This section highlights the role of extra-linguistic factors in the adoption
of the linguistic innovations in PA spoken in Gaza City.

6.3.1   *Role of the individual speaker*
The socially dominant but linguistically non-dominant Gazan speakers, based on
structural conceptual similarities, would have become aware of *ra:h* as a future
marker in the Jaffan dialect. Gazan speakers established an equivalence relation
between the element in the model code and their own lexical verb *ra:h* as a suita-
ble target onto which the relevant future semantic property was copied through a
process of selective copying (Johanson, 1999, 2002, 2008). The findings also show
different degrees of acceptance of the innovative forms *ra:h* and *ha-* by speakers

with different characteristics. Though the shift started as a result of individual innovative acts, individual speakers responded with different degrees of acceptance of the new forms. While conventionalized and used by middle-aged and younger speakers to varying degrees, they are not adopted by older speakers. This is evident by the Jaffan Gazan older speakers' having forms identical to the ones in the Jaffan dialect spoken in Jaffa today, that is, mainly the forms *ra:ḥ* and *bidd-* with the absence of the innovative form *ḥa-* from their repertoire. As noted above, *ḥa-* is the main form that is used exclusively by younger speakers of both dialects in Gaza. Copying as performed by PA speakers can be viewed as a goal-driven communication activity (Matras, 2009, p. 4) and as a creative technique used by the speaker to enrich their variety and enhance its functionality (Johanson, 2008). It is probable that with the copying of the particle ra:ḥ, Gazan speakers enriched their dialect by adding a new linguistic form whose only function was to express the future. Though the form *b-+ p-stem* (non-past form of the verb) was originally available to express the meaning of future, this form has other functions to fulfill such as the present, the imperfective, and the progressive meanings. Later, with the development of the prefix *ha-*, the PA verbal system evolved from a two-way to a three-way tense system. This language change of a motion verb becoming a future marker can be viewed as an internal language change that has become the object of external factors of change (Johanson, 2002). It has resulted in increased resemblance of verbal systems of Jaffan and Gazan dialects spoken in Gaza City.

The variational patterns of PA speakers with respect to the morphological variables outlined in Section 6.1 interact with speaker age and regional identity. Older GJ speakers, for example, who identify themselves with the original community in Jaffa, show more linguistic conservatism than younger speakers. As indicated in Section 6.1, the three older GJ speakers in the study express future time reference with a pattern identical to that of Jaffa Jaffans. The data show that the conventionalization of the two codes *ra:ḥ* and *ḥa-* has a wider scope among younger and middle-aged G and GJ speakers; however, the innovative forms are not accepted by older speakers. When asked about how they feel about the kind of dialect they use, older speakers made it clear that they feel proud to speak the same language their parents spoke, "the pure dialect of their original place" as one of the GJ speakers describes it. Older G and GJ speakers are conservative as regards the new innovations, and their speech does not incorporate any of the innovative forms, while middle-aged and younger speakers adopt, spread, and further develop the innovative forms. This finding fits with what Labov (1994, 2001) notes that younger speakers use more innovative forms than older speakers. Eckert (1998, p. 152) also notes that community studies of variation frequently show that increasing age correlates with increasing conservatism in speech.

Johanson (1999, p. 55) notes that first generations of diaspora communities speak codes that are very similar to the regional varieties they spoke in the original communities left behind. Early diaspora varieties are characterized as heterogeneous, whereas later varieties often show less variation. The Gaza context seems to fit within this categorization. The early language situation is heterogeneous and characterized by differences between older Gazan and older GJ speakers, whereas the later language situation can be described as being more levelled, as evident in the speech of younger speakers of both dialects who exhibit the same use patterns of the future.

To conclude, based on the above discussion of the distribution of the future markers *ḥa-* and *ra:ḥ*, I interpret the general age-related pattern in PA spoken in Gaza as representing a generational change in progress. I take young and old people to be representing the contemporary and historical states of the PA spoken in Gaza City, respectively. Moreover, Jaffan and Gazan dialects were different from each other at the early time of contact, which is still reflected in the older speakers' speech. However, later codes are more levelled and homogeneous as shown by data from middle-aged and younger speakers. The age-stratified patterns of Gazan and Gaza Jaffan speakers in Gaza might be suggestive of contact-induced language change.

6.3.2   *Social asymmetry and dominance relations*

Johanson (1999, p. 54) points out that dominance relations and the degree of social asymmetry play a role in the outcome of contact-induced change. He defines dominance relations in terms of social, economic and political relations, and strength where the language of the immigrant community is usually the dominated code. Johanson (1999, p. 54, 2002) states that code-copying processes typically occur in a dominated speech community that is connected to a dominant speech community by strong external bonds. He suggests that the asymmetric dominance relations between the two codes cause "one-directional dynamics" where the dominated variety adopts new linguistic habits from the dominant variety (1999, p. 54). In most cases, the dominated variety is a diasporic variety. Johanson's proposal seems to work well in language contact situations where an asymmetric relationship between a dominant and a dominated language can easily be detected or in a dialect contact situation where one of the dialects in contact serves as a lingua franca. However, in the Gaza context, which includes several mutually intelligible dialects with none of these dialects acting as a lingua franca, it is not easy to assume asymmetric dominance relations that result in unidirectional dynamics between the basic code and the model code. In situations like this, dominance relations need to be defined carefully as it is not necessarily the case that adoption or imposition processes only take place in one direction towards the model or dominant code by the speakers of the dominated code.

As Jaffan PA is a diasporic dialect, it would be plausible to think of Gazan PA as the dominant dialect and of Jaffan as the dominated dialect, and thus to expect the change to be unidirectional i.e. copying from Gazan into Jaffan and other refugee dialects. However, the data collected in this pilot study suggests that copying is performed by speakers of the dominant Gazan dialect. Prestige is also an important factor that needs to be considered here. The Jaffan dialect has always been perceived as the most prestigious Palestinian dialect. It is put in the same category of the Levantine dialects of Jerusalem, Beirut, and Damascus that are labelled as ''prestigious urban dialects'' (Naim, 2006; Rosenhouse, 2007; Al Wer, 2002). Although the socio-political status of Jaffans changed after their arrival in Gaza as refugees, the dialect has maintained its prestige. This view of Jaffa as a prestigious dialect made it acceptable for speakers of the dominant Gazan dialect to copy elements from the dominant Jaffan dialect, for example, copying *ra:h*.

7. Conclusion

This study was designed to meet the following goals: to examine the variation among speakers of PA living in Gaza in the way they express future time reference and whether the variation represents a case of language change in progress, and, if so, to determine the factors involved in these developments. To accomplish these goals, I collected and compared data from 20 PA speakers from different age groups and dialects. I provided a linguistic description of the two verb paradigms operating in PA and examined complex structures that complement the possibilities of the simple verb paradigm to express future time reference in PA. The data show that the PA verbal system is evolving from a past-non-past two-way system to a three-way tense system of past, present, and future.

The developments that move the Jaffan and Gazan dialects towards one another in their marking of future time reference seem to admit a significant role for both internal or linguistic and extra-linguistic factors. With these developments, the verbal systems of Gazan and Jaffan dialects have come to resemble one another more than previously. Though the innovations represent familiar paths of development, they cannot be studied in isolation from certain socio- and psycholinguistic aspects of the speakers involved and the contact setting in which the code interaction occurs. Extra-linguistic factors such as length and intensity of contact, prestige, and speaker characteristics strongly influence the adoption of linguistic innovations in individuals, groups, and generations in Gaza City. Further investigation of data to be collected from a larger sample of PA speakers is needed to determine if the available expressions for future time reference interact with different degrees of remoteness in the future.

Abbreviations

FEM	feminine	NOM	nominative
FUT	future	NON-PAST	non-past
G	Gazan	PA	Palestinian Arabic
GJ	Gaza Jaffan	PAST	past
INDIC	indicative	PL	plural
MASC	masculine	SA	Standard Arabic
MSA	Modern Standard Arabic	1	first person
NEG	negation	3	third person

References

Al-Wer, E. (1991). Phonological variation in the speech of women from three urban areas in Jordan. Doctoral dissertation. University of Essex.

Al-Wer, E. (2002). Jordanian and Palestinian dialects in contact: Vowel raising in Amman. In M. Jones & E. Esch (Eds.), *Language change: The interplay of internal, external and extra-linguistic factors* (pp. 63–79). Berlin: Mouton de Gruyter.

Bybee, J. (2003). Mechanisms in grammaticalization: The role of frequency. In B. Joseph & R. Janda (Eds.), *The handbook of historical linguistics* (pp. 602–623). Oxford: Blackwell.

Bybee, J., Perkins, R. & Pagliuca, W. (1991). Back to the future. In B. Heine & E. Traugott (Eds.), *Approaches to grammaticalization*, Vol. 2 (pp. 17–58). Amsterdam: John Benjamins.

Bybee, J., Perkins, R. & Pagliuca, W. (1994). *The evolution of grammar: Tense, aspect and modality in the languages of the world*. Chicago: University of Chicago Press.

Eckert, P. (1998). Age as a sociolinguistic variable. In F. Coulmas (Ed.), *The handbook of socio-linguistics* (pp. 151–167). Oxford: Blackwell.

Farrar, K. & Jones, M. (Eds.). (2002). *Language change: The interplay of internal, external, and extra-linguistic factors* (pp. 1–16). Berlin: Mouton de Gruyter.

Hickey, R. (2010). Language contact: Reconsideration and reassessment. In R. Hickey (Ed.), *The handbook of language contact* (pp. 1–28). Chichester: Blackwell.

Heine, B. (2012). On polysemy copying and grammaticalization in language contact. In C. Chamoreau & I. Leglise (Eds.), *Dynamics of contact-induced language Change* (pp. 125–166). Berlin: De Gruyter.

Heine, B. & Kuteva, T. (2005). *Language contact and grammatical change*. Cambridge: Cambridge University Press.

Holes, C. (2004). *Modern Arabic: Structures, functions, and varieties*. Washington D.C: Georgetown University Press.

Hopper, P. & Traugott, E. (2003). *Grammaticalization*, 2nd ed. Cambridge: Cambridge University Press.

Horesh, U. (2014). Phonological outcomes of language contact in the Palestinian Arabic dialect of Jaffa. Doctoral dissertation. University of Essex.

Johanson, L. (1999). The dynamics of code-copying in language encounters. In B. Brendemoen, E. Lanza, & E. Ryen (Eds.), *Language encounters across time and space* (pp. 37–62). Oslo: Novus.

Johanson, L. (2002). Contact-induced change in a code-copying framework. In M. Jones & E. Esch (Eds.), *Language change: The interplay of internal, external, and extra-linguistic factors* (pp. 285–313). Berlin: Mouton de Gruyter.

Johanson, L. (2008). Remodeling grammar: copying, conventionalization, grammaticalization. In P. Siemund & N. Kintana (Eds.), *Language contact and contact linguistics* (pp. 61–79). Amsterdam: John Benjamins.

Labov, W. (1994). *Principles of linguistic change, Vol.1: Internal factors*. Oxford: Blackwell.

Labov, W. (2001). *Principles of linguistic change, Vol.2: Social factors*. Oxford: Blackwell.

Matras, Y. (2009). *Language Contact*. Cambridge: Cambridge University Press.

Mitchell, T. & Al-Hassan, S. (1994). *Modality, Mood, and Aspect in Spoken Arabic*. London: Kegan Paul.

Naim, S. (2006). Beirut Arabic. In K. Versteegh, M. Eid, A. Elgibali, M. Woidich, & A. Zaborski (Eds.), *Encyclopedia of Arabic Language and Linguistics*, Vol. 1 (pp. 274–286). Leiden: BRILL Publishers.

Rosenhouse, J. (2007). In K. Versteegh, M. Eid, A. Elgibali, M. Woidich & A. Zaborski (Eds.), *Encyclopedia of Arabic Language and Linguistics*, Vol. 1 (pp. 481–493). Leiden: Brill Publishers.

Thomason, S. & Kaufman, T. (1988). *Language contact, creolization and genetic linguistics*. Berkley CA: University of California Press.

Trudgill, P. (1986). *Dialects in contact*. Oxford: Blackwell.

Weinreich, U. (1953). *Languages in Contact: findings and problems*. New York: Linguistic Circle of New York.

Neuters to none

A diachronic perspective on loanword gender in Bosnian/Croatian/Serbian

Mary Ann Walter

Middle East Technical University, North Cyprus Campus

This study documents a diachronic change in the status of neuter noun gender in Bosnian/Croatian/Serbian (BCS). Previous research demonstrates that the neuter class is closed in BCS (Simonović, 2010), and that recent loanwords from western European languages fitting its phonological profile are instead classed as masculine. I show that this is not the case for earlier loanwords from Turkish. New neuter nouns are still accepted, and when changed, are classed as feminine, not masculine. This follows an attested pattern of assigning gender according to lexical distributions. An account of gender assignment utilizing Optimality Theory and incorporating gradiently-ranked constraints captures this pattern in which rankings can shift over time, thus leading to the observed historical changes in rates of neuter gender assignment.

Keywords: grammatical gender, loanwords, borrowing, Bosnian, Croatian, Serbian, Slavic, Balkan, phonology, Optimality, Turkish

1. Introduction

Contemporary Bosnian/Croatian/Serbian (BCS)[1] differs from other South Slavic languages in that the neuter noun class is currently vestigial and closed to new lexical items. In this study, I present data on BCS vocabulary of Turkish origin, dating from an earlier historical period of the language during which loanwords from Turkish constituted a significant number of new BCS lexical items. The treatment of these loanwords with respect to gender assignment reveals a different grammatical

---

1. Although the acronym BCMS (Bosnian/Croatian/Montenegrin/Serbian) is becoming increasingly common, I eschew it here due to the ongoing concurrent use of BCS in many academic sources and the lack of specifically Montenegrin data in this study.

system than contemporary BCS. In the pre-contemporary period during which these Turkish-origin forms were borrowed, new neuter nouns were acceptable, although already dispreferred. Moreover, the choice of 'repair strategy,' if used, also differs between the contemporary and earlier time periods. In contemporary BCS, incoming neuter-like lexical items are categorically classed as masculine (Simonović 2010). However, in the earlier period, neuter-like loanwords are classed as feminine, if they are changed at all.

The observed patterns prompt two questions. The first is how to account for the gender assignment phenomena of earlier BCS. The second is how to account for the transition between the earlier system and the contemporary system of gender assignment.

I propose that the earlier pattern is due to a probability-matching strategy of gender assignment in conjunction with a slightly greater prevalence of neuter nouns in the lexicon of the time. Such behavior has been previously documented for gender assignment in both L2 speech and in loanwords, both within and outside of the South Slavic language family (Walter, 2011, 2017). It can be accounted for in an approach to gender assignment using Optimality Theory and incorporating gradiently-ranked constraints (Rice, 2006).

As for the second question on how the transition occurred, the very gradience of the constraints in the Optimality Theory model allows for their reranking in such a way as to mirror the lexical distribution as it changes. When fewer neuter nouns are used – perhaps due to changing political or social factors, such as decreasing willingness to use words of Turkish origin, which are likely to be of neuter type – the constraint ranking shifts to reflect this decrease in frequency. Thus socially-driven lexical turnover could lead to the observed historical changes in rates of neuter gender assignment.

Section 2 outlines the noun gender system of BCS. Section 3 presents the data on gender assignment and loans of Turkish origin, drawn from three different sources. Section 4 presents comparative data from South Slavic and elsewhere. The formalization proposed for the BCS loanwords appears in Section 5, and discussion, conclusions and remaining questions in Section 6.

2.   Noun gender in BCS

Like its Slavic sisters, BCS has a three-gender system of masculine, feminine and neuter nouns. Masculine nouns show the most phonological variation in stem-final position, and are typically considered the default. They most often end in consonants, but may also end in /o/ or /e/, especially proper nouns. Feminine nouns typically, though not always, end in /-a/ in the nominative singular citation form.

Neuter nouns typically end in /o/ or /e/. These final vowels may be part of the noun stem. However, they also very often occur as part of a derivational suffix, such as abstract /-stvo/ (roughly equivalent to English '-*ity*') or the diminutive suffix /-tʃe/. In such cases, the noun's neuter gender is associated with the derivational suffix, rather than the noun stem itself (which, when used as a free morpheme, may have another gender entirely). Therefore, neuter gender is something a noun 'gets' more than something it inherently 'is.' In other words, neuter gender is quite likely to be derivationally generated rather than (or as well as) an inherent, lexically-specified property of the noun stem.

Neuter is also by far the smallest category. Gender ratios were calculated based on Cahen's (1920) dictionary, chosen because the proportion of neuters in it is presumably higher than today, and its date of publication is also closer in time both to the period of Western European lexical influx and the previous Turkish lexical influx.

According to Cahen, neuters constitute only 12% of all nouns. Masculines and feminines, on the other hand, constitute 45% and 43% of the lexicon respectively, in terms of lexical type frequency. This yields a .28 ratio of neuter-to-feminine – an important number, as shall become apparent.

Further evidence for the vestigial nature of neuter gender in BCS comes from the treatment of recent loanwords from Western European languages. Simonović's (2010) collection of approximately 150 such loanwords reveals that those ending in /-e/ or /-o/ vowels, which thus fit the neuter paradigm, nevertheless are near-categorically assigned to masculine gender instead. He argues that the phonotactically-dispreferred CV transition probability at the end of some (but not all) of the stems results in their wholesale masculine gender assignment in place of neuter (recall that a small number of masculine items also end in /-e/ or /-o/). As a result, neuter nouns now constitute a 'closed class' which does not admit new monomorphemic members either through borrowing or language-internal mechanisms, aside from neuter-assigning derivational suffixes.

3.   BCS loanwords and gender

Due to several centuries of Ottoman rule, Turkish is a major source of borrowed vocabulary in BCS. According to Škaljić (1966), who compiled a dictionary of "Turkisms," the BCS lexicon contains a total of over 6,000 lexical items of Turkish origin. Roughly 5,000 of them remained in active use by BCS speakers in the mid-20th century. Turkish nouns are also highly likely to be /-e/-final, as such nouns are common both in the native lexicon and in the very large numbers of nouns borrowed from Persian and Arabic. (Nouns ending in /-o/, on the other

hand, are unattested until the advent of later European borrowings). Thus, such nouns are perfect candidates for being assigned neuter gender in BCS. The remainder of this section examines the fate of such nouns in three separate loan mini-corpora.

3.1   Corpus 1

Ban and Matovac (2012) collect a number of Turkish loanwords which were brought into Croatian. Those falling into canonical feminine (/a/-final) and neuter (/e/-final) templates are tallied below in Table 1.[2]

Table 1. Distribution of gender (~final vowel)
in BCS words of Turkish origin (Ban & Matovac, 2012)

	Turkish	BCS
/a/-final	13	31
/e/-final	12	1

The tallies in the 'Turkish' column represent Turkish forms which (1) end in either /a/ or /e/ in Turkish, and (2) have loan counterparts in BCS (which may or may not retain the same final vowel). Their BCS loan counterparts would be expected, *a priori*, to be transparently assigned feminine or neuter gender, respectively, in accordance with their final vowel.

The tallies in the 'BCS' column represent BCS forms which (1) end in either /a/ or /e/ in BCS (which may or may not correspond faithfully to the final phoneme in the Turkish source form), and (2) originate from Turkish.

Note that the resulting number of forms in each column is close but not equivalent. This is due to a small number of additional changes to final phonemes in some forms, which are detailed below.

The results are compatible with Simonović's observations regarding the closed nature of the neuter class. As shown in Table 1, Turkish words are roughly equally likely to appear with final /-a/ (therefore, feminine-like in BCS) or with final /-e/) (therefore, neuter-like in BCS). There are 12 of one type, 13 of the other. However, in BCS, this roughly even distribution has disappeared. Of the 13 neuter-like /-e/-final Turkish words, only one appears as such in BCS. Notably, it is not the case that such /-e/-final, neuter-type words are simply not borrowed. The other 12

---

2.   The gender of these borrowings was verified in Škaljić (1966) and conforms to its expected value according to final phoneme (feminine for /a/, neuter for /e/) in all cases for both this corpus and the two succeeding ones.

lexical items of this group do persist as loans in the BCS lexicon. However, their forms have changed phonologically, in ways that entail an associated change in grammatical gender.

This is demonstrated both by the numbers in Table 1 and in the specific examples in Table 2 below. The /-e/-final neuter-type Turkish source forms are near-categorically converted to /a/-final feminine BCS forms. Thus, the total of /-a/-final feminine BCS output forms (31 items) is roughly equivalent to the sum of both the originally /-a/-final forms (in Turkish) and the originally /-e/-final forms (in Turkish; 12 + 13 = 25).

There are only three exceptions to this 'feminization' process – one final /e/ remained as such (Croatian *dugme* 'button,' from Turkish *düğme)*, one such vowel was deleted entirely (Croatian *kesten* 'chestnut,' from Turkish *kestane*), with associated masculine gender assignment, and one final-vowel change went in the opposite direction (Croatian *gajde* 'bagpipe,' from Turkish *gayda*).

Table 2. Neuter-to-feminine Croatian nouns*

Turkish	Croatian	Gloss
baruthane	*barutana*	powder magazine
çizme	*čizma*	boot
cezve	*džezva*	Turkish coffeepot
kahve	*kava*	coffee
kahvehane	*kavana*	coffee shop
kule	*kula*	tower
yelpaze	*lepeza*	fan brush
lüle	*lula*	pipe
bahçe	*bašča*	garden

* Examples are given in the standard orthography of each language.

A smaller number of additional feminine nouns are also generated through other means. Affixation of /a/ to consonant-final forms occurs in 4 cases (*çorap > čarapa* 'sock,' *papuç > papuča* 'slipper,' *tambur>tamburica*, *ciğer > džigerica* 'liver'). There is also a separate regular change of final high vowels to /-ija/ which applies in five additional cases.

Forms like these are the cause of the numerical disparity between column totals in Table 1. That is, the total number of /a/ and /e/-final forms in BCS (right-hand column) is greater than the total number of /a/ and /e/-final source forms in Turkish (left-hand column), due to additional changes to the final phoneme like those in the preceding two paragraphs.

Notably, changes in the other direction – final /a/ to final /e/, that is, feminine-like to neuter-like – do not occur, with a single exception (*gajde* 'bagpipe,' Turkish

gayda). Nor are substitutions of these vowels regular phonological changes in any way elsewhere in the word, as also demonstrated by the examples in Table 2. Words such as *džezva* 'Turkish coffeepot' make clear that the vowel change of /e/ to /a/ is in no way regular. It is only the final, gender-associated vowel which changes, never a vowel internal to the stem.

In sum, the data from Turkish loans illustrate a significantly different picture than the one apparent in Simonović's discussion of more recent loans from western European languages. The earlier Turkish loans, unlike the later western European loans, may still (rarely) be assigned neuter gender if they fit the appropriate phonological template. If not, the Turkish loans are assigned feminine gender, rather than the masculine gender assigned to western loans. And if such a change is made, the phonological form of the final vowel changes as well, unlike the western-origin forms.

3.2   Corpus 2

Friedman's (2003) independent collection of Serbo-Croatian loanwords from Turkish illustrates a somewhat different pattern. Forms falling into canonical feminine and neuter templates are tallied below in Table 3.

Table 3. Distribution of gender (~final vowel) in BCS words of Turkish origin (Friedman, 2003)

	Turkish	BCS
/a/-final	17	37
/e/-final	17	8

Recall that the 'Turkish' column represents Turkish /a/ and /e/-final words with BCS loan counterparts (which may no longer display the same final phoneme), while the BCS column represents words of Turkish origin which are /a/ and /e/-final in BCS (but not necessarily in Turkish).

Due to an array of cases in which BCS speakers apparently generated vowel-final forms from originally differing forms, in ways explored below, the column totals are not equivalent.

Once again, we observe that the number of /-e/-final (~neuter) and /-a/-final (~feminine) nouns are precisely balanced in the original Turkish source vocabulary. Yet the number of /e/-final (~neuter-like) nouns is cut in half when borrowed from Turkish to BCS, from 17 to 8 items. This represents a significant decrease, but is not quasi-categorical as in Croatian. On the other hand, there is a drastic increase in the group of /a/-final (~feminine-type) nouns, which essentially double in number. As

in Croatian, the output of borrowing yields a number of /-a/-final ~feminine forms which is roughly equivalent to the sum of the original /-a/-final and /-e/-final forms as they appear in Turkish ($17 + 17 = 34$).

Again, these changes are primarily due to final /e/-to-/a/ vowel shifts ($n = 12$), as shown in Table 4 below. Additional /a/-final forms originate from words with Turkish final high vowels, which regularly get final /a/ ($n = 9$), and a handful of consonant-final forms with added /a/ ($n = 3$; *kınnap > kanapa* 'string,' *sucuk > sudžuka* 'sausage,' *çorap > čarapa* 'sock, stocking').

Table 4. Neuter-to-feminine nouns

Turkish	Serbo-Croatian	Gloss
bahçe	*bašča*	garden
çevre	*čevra*	frame; (gold)-embroidered
gazete	*gazeta*	newspaper
hasta(h)ane	*hastahana*	hospital
kese	*kesa*	bag, purse
kepçe	*kevčija*	large mixing spoon
kopçe	*kopča*	button
lüle	*lula*	pipe
meze	*meza*	snack
şemsiye	*šemsija*	umbrella
tencere	*tendžera*	frying pan
testere	*testera*	saw (tool)

In 2 cases, final /e/ is also deleted (*kestane > kesten* 'chestnut,' *pencere > pendžer* 'window'). There are only 3 cases in which final /e/ is retained (*leke > lece* 'stain,' *çene > čene* 'chin,' *şişe > šiše* 'bottle'). However, unlike in Croatian, a handful of new /e/-final forms are also generated. Two are added to consonant final forms (*merdiven > merdevine* 'stairs,' *keman > cemane* 'violin'). Three additional cases arise from a change of /a/ to /e/ (*toka > toke* 'buckle,' *parça > parče* 'piece,' *paça > pače* 'calf or sheepfoot stew').

As a result of these changes, the ratio of /-e/-final to /-a/ final nouns changes from an even distribution (in Turkish, with 17 nouns in each category) to a highly asymmetrical distribution in BCS (8 versus 37 items). The resulting ratio of neuter-to-feminine nouns in the BCS loanwords is .22. This is less than but very close to the ratio in the lexicon as a whole (.28, per Cahen, 1920) – and, of course, a much closer match than the ratio without such changes (1.0).

Binomial distribution tests reveal that after the changes, the distribution of neuter versus feminine gender within the loanword group is not significantly different

from the distribution in the lexicon as a whole as instantiated in Cahen (1920; $p = .12$). On the other hand, if the distribution had remained as it was in the Turkish source forms, it would be significantly different from the BCS lexicon ($p = .0002$).

Thus, the changes in final vowel/gender status among these loanwords have the result of bringing the loanwords in line with the pre-existing gender ratio of the BCS lexicon.

3.3   Corpus 3

The largest corpus is drawn from the work of Sjoberg (1972) on Turkish loanwords in the work of Mostar-born, ethnically-Serb writer Aleksa Šantić. It is also the closest to the time of borrowing and likely to predate many of the borrowings discussed by Simonović, as Šantić's writings date from 1887–1921. Forms falling into canonical feminine and neuter templates are tallied below in Table 5.

Table 5. Distribution of gender (~final vowel) in BCS words of Turkish origin (Sjoberg, 1972)

	Turkish	BCS
/a/-final	34	96
/e/-final	42	25

As in Corpus 2, the number of /e/-final~neuter forms drops drastically in the transition from the Turkish source form to the BCS output form, by approximately half (from 42 to 25 forms; again, not near-categorically as in Croatian Corpus 1). The number of /a/-final nouns, on the other hand, once again increases drastically. Consistent with the previous two datasets, the number of /-a/-final feminine-type nouns nearly triples (from 34 forms in Turkish to 96 forms in BCS). The resulting neuter-to-feminine ratio is again a close approximation of the overall lexicon, as seen above (.26 here, .22 above, .28 overall based on Cahen). Without these changes, neuter-to-feminine ratio would instead be 1.24, with neuters actually substantially outnumbering feminine forms. Binomial distribution tests confirm that the distribution of the Turkish source forms is significantly different from that of the BCS lexicon ($p < .001$), whereas the distribution of the forms as borrowed is not ($p = .08$).

As expected, the single largest contributing factor is again the shift from final /e/ to /a/ ($n = 24$). A substantial group remain /e/-final but also have /a/-final variants ($n = 13$). The opposite change, /-a/ to /-e/, occurs only twice (*parça > parče* 'piece,' *gonca > gondže* 'flowerbud,' optionally in *toka* 'buckle'). Other additional /a/-final forms originate from high-vowel-final words ($n = 17$). The remainder are from consonant-final forms, with a handful of examples of the reverse.

3.4   Summary

The closing of the neuter class was apparently faster for Croatian than for other varieties within BCS, according to the data of Corpus 1. This may be an artefact of the smaller size of the corpus, or of the shorter length of time during which areas of Croatia were dominated by Ottoman rule, or possibly due to more resistance to loanwords in Croatian in general, as claimed by Mønnesland (1973). Alternatively, Croatian experienced greater proximity to and contact with Italo-Romance varieties (such as Venetian and Dalmatian), particularly in the coastal cities (Ragusa/ Dubrovnik, Spalato/Split, Zara/Zadar, *et cetera*). Bilingualism and/or contact with these varieties with their two-gender systems, versus other Balkan languages with three genders, may have played a role.

For the other varieties within BCS, however, the neuter noun class was clearly still open at the time of borrowing from Turkish. In Corpora 2 and 3, /e/-final nouns were faithfully borrowed as such, and with neuter gender, half or more of the time. For those nouns which were not borrowed as neuter, the new gender was feminine, not masculine, *contra* the situation for later loans. The result of these changes is a close match with the preexisting lexical distribution of gender in BCS.

The resulting ratios of neuter-to-feminine nouns in the loanword mini-corpora are summarized in Table 6 below.

Table 6. Ratio of neuter-to-feminine nouns across corpora

	Language-internal	Corpus 1	Corpus 2	Corpus 3
Turkish		.92	1.00	1.24
BCS	.28	0	.22	.26

As Table 6 demonstrates, nouns ending in /-e/ and /-a/ appear in roughly equal numbers in Turkish, such that the ratio of one group to the other clusters around 1. This is far from the case in BCS, where neuters appear in significantly smaller numbers, such that the ratio of one to the other is well below 1. This is true for the language as a whole (column 1) as well as for the individual corpora of loanwords. Of the two mini-corpora which have not eliminated neuters entirely (see discussion in the preceding paragraph), the neuter-to-feminine ratio is statistically indistinguishable from the ratio given in column 1 for the language as a whole. As discussed earlier in this section, this is a statistically reliable and robust result. As this section has also shown, this precise quantitative match is generated primarily by reclassification of a subset of neuter-type words from Turkish into the feminine class.

A brief examination of the example wordlists given above is sufficient to demonstrate that there is no semantic commonality among the words which change

from neuter-type to feminine-type and nothing which distinguishes them from the non-changing words. The semantic fields include places, objects, food, and body parts.

Similarly, phonological factors such as stress placement, syllable count, or preceding phoneme are found to have no predictive value with respect to gender assignment for these items. Turkish is typically described as a final-stress language (or, alternatively, stressless). Because all of the source forms are consistent in this respect, this means that there are no stress-related phonetic or phonological properties which could determine which words remain /-e/-final and which become /-a/ final in BCS.

As for syllable count, the wordlists given above also illustrate that the source forms range from bisyllabic to quadrisyllabic, with no apparent correlation with whether or not the final vowel (and associated gender) changes.

The consonant preceding the final, gender-associated vowel also ranges across articulatory places and manners with no identifiable pattern.

The following section will review the status of implicit knowledge of lexical statistics in human language and its relevance here, including a series of examples from both within and outside the Slavic language family which appear to exemplify a similar pattern of noun classification. Section 5 then develops a formal theoretical account of this pattern as applied to BCS.

4.   Comparison across time and space

In this section, comparative data from the South Slavic neighbors of BCS will be introduced, as well as previous work on probability-matching and gender distribution in loanwords in other language families. The overall importance and relevance of speakers' implicit knowledge of frequency with respect to abstract grammatical categories, both in general and with respect to the phenomenon under consideration in this study, is discussed in the concluding part of this section.

4.1   South Slavic

Bulgarian, the South Slavic sister language of BCS, features an essentially identical system of grammatical gender. Feminine gender is associated with a final /-a/ vowel, neuter gender with a final /-e/ or final /-o/ vowel, and masculine gender with consonants, or other vowels. Because Bulgarian, unlike BCS, is not a case-inflecting language, these final vowels appear even more stably, rather than surfacing only in the citation nominative singular form. According to Manova and Dressler (2001),

grammatical gender in Bulgarian is determined entirely by the phonological form of the final vowel, rather than by semantic or other considerations.

Previous work on Bulgarian illustrates a very similar pattern for its Turkish loanwords to that observed in BCS (Walter, 2017). Table 7 summarizes the Bulgarian data, in which /a/-final Turkish source forms remain as such (and are assigned feminine gender) but /e/-final Turkish source forms are changed in a substantial minority of cases, such that they are borrowed as /a/-final, feminine-gender forms.

Table 7. Distribution of gender (~final vowel) in Bulgarian words of Turkish origin

	Turkish	Bulgarian
/-a/-final	70	86
/-e/-final	61	43

The resulting Bulgarian loans *with* vowel changes yield a neuter-to-feminine ratio of 43/86 = .50, extremely close to the overall lexical ratio of .48. The original forms as they appear in Turkish, on the other hand, would yield a ratio of 61/70 = .87. Once again, changes result in a close approximation to the pre-existing lexical gender ratio. As with BCS, these shifts result in a significantly distinct distribution becoming statistically indistinguishable from the pre-existing one.

Macedonian illustrates a similar pattern, as shown in Table 8 (Walter, 2015).

Table 8. Distribution of gender (~final vowel) in Macedonian words of Turkish origin

	Turkish	Macedonian
/a/-final	25	52
/e/-final	32	25

The resulting Macedonian gender ratio (.48) is significantly closer to the dictionary distribution of .35 than the importation of unmodified forms would generate.

Alternative explanations for the shift to feminine gender based on phonological and/or semantic criteria are exhaustively explored and dismissed by Walter (2015, 2017). It appears that there is no other motivating factor than the probability-matching, lexical distribution one identified here. While the prevalence of neuter gender differs slightly among these closely-related languages, the tendency of speakers to match the degree of that prevalence in adapting loanwords seems to be consistent.

4.2   Outside Slavic

An array of evidence from non-Slavic languages demonstrates the cross-linguistic frequency of such behavior (Walter, 2011). It is striking that such typologically differing languages display an identical pattern of probability matching.

Spanish loanwords of Arabic origin also involve an otherwise unexpected final /e/–/a/ alternation. While the default epenthetic vowel of Spanish is /e/, /a/ is sometimes used instead word-finally, resulting in a form which matches the /a/final feminine template. This exceptional type of epenthesis occurs in precisely the proportions that generate a match with pre-existing lexical gender ratios. Portuguese replicates this pattern – notably, however, with different specific noun forms than Spanish. This fact implies that properties of the borrowed nouns do not determine the borrowed forms' final vowel and gender. Rather, the overall quantitative distribution of gender in the borrowing language is the determining factor.

Loans going in the other direction – from Romance to (Moroccan) Arabic – illustrate similar behavior. Not just one change in this case, but a conspiracy of phonological processes apply to these loans, including final consonant deletion, epenthesis of final /a/, simplification of nasalized vowels, and changing final vowel quality to /a/. The ultimate result is an exact match of loanword grammatical gender proportions with pre-existing Arabic lexica (both 31% feminine). Without such changes, the loans would be only 19% feminine.

Finally, adult acquisition of Arabic language plurals presents a comparable pattern, in which noun roots are assigned to plural classes by advanced learners in the correct proportions, though often incorrectly for individual items.

Such behavior demonstrates that speakers are not only aware of distributional patterns in their lexicon, but they make use of their knowledge in generating morphological forms.

4.3   Summary

The phenomena under consideration illustrate one of the most astounding discoveries of 20th century psychology and linguistics – that is, the degree and specificity of humans' implicit knowledge of frequency with respect to abstract and non-abstract linguistic categories. A substantial body of research demonstrates that even in early infancy, listeners pay close attention to the statistical distribution of categories such as individual speech sounds (utilizing distributional properties to construct phonemic categories), phoneme transition probabilities, and syllable transition probabilities (reviewed in Jusczyk, 1997). Later in life, listeners remain sensitive to lexical frequency measures both in the native language (Trueswell, 1996) and in language learning (Ellis, 2002), with concomitant effects on linguistic behavior (choice of syntactic structure *et cetera*). Such sensitivity to statistical patterns in linguistic data

is well-attested and known to be particularly influential in the linguistic behavior of adult speakers (Hudson-Kam & Newport, 2005, 2009; Wonnacott & Newport, 2005). Specifically, speakers learn information about the relative probabilities of grammatical variables. Incoming lexical items are then treated in such a way as to maintain the pre-existing probability distribution of these variables.

Such behavior is not limited to the morphological domain, but has also been observed with respect to phonemic feature specifications, as in the assignment of voicing to final consonants in final-devoicing languages. Rather than consistently choosing a transparent, voiceless consonant when presented with a finally-devoiced form, speakers select underlying voicing a substantial minority of the time: 24% for Dutch speakers (Ernestus & Baayen, 2003), 38% for Turkish speakers (Becker *et al.*, 2011). The percentage of voiced final consonants varies according to the place of articulation of that consonant, and for each language, the percentages roughly reflect the relative probability of different voiced final consonants in the lexicon.

With these psycholinguistic facts in mind, it is perhaps not surprising that the gender assignment pattern observed for BCS recurs in its closely related neighbors, and that very similar effects are visible in a variety of typologically distinct other languages when it comes to the distribution of nominal grammatical categories newly entering via loanword borrowing. The apparent relevance of distributional properties of the lexicon in loanword adaptation is not a quirk of BCS, but is well-attested both among its neighbors and among other unrelated languages, and in other language domains.

5. Formalization

Rice (2006) develops a theory of optimal gender assignment employing language-specific gender assignment constraints. Although formulated for German, the constraints can be adapted for BCS as follows:

*-E→M, F: A noun ending in /e/ (or /o/) is assigned neither masculine nor feminine gender

*-A→M, N: A noun ending in /a/ is assigned neither masculine nor neuter gender

*-C→F, N: A noun ending in a consonant (or vowel other than /a/, /e/, or /o/) is assigned neither feminine nor neuter gender

Table 9 exemplifies these constraints as applied to a hypothetical /e/-final native BCS word. The constraint titled "FAITH" refers to a catch-all category of faithfulness (IDENT) constraints which militate against, for example, changing the quality of the final vowel.

Table 9. Typical neuter gender assignment

/atabe/	*-E→M, F	*-A→M, N	*-C→F, N	FAITH	*N	*F	*M
a. → atabe N					*		
b. atabe F	*!					*	
c. ataba F			*!			*	
d. ataba N		*!		*	*		

For this item, neuter gender is assigned as normal according to the gender-related constraints defined above. The markedness constraints against specific genders do not come into play, as they are all ranked below the gender constraints as well as the constraint(s) against changes to final vowels.

A hypothetical /a/-final word, on the other hand, is exemplified in Table 10. In this case as well, the output form unproblematically matches the expected gender according to phonological form, and is faithful to the underlying final vowel.

Table 10. Typical feminine gender assignment

/ataba/	*-E→M, F	*-A→M, N	*-C→F, N	FAITH	*N	*F	*M
a. atabe N				*!	*		
b. atabe F	*!				*	*	
c. → ataba F						*	
d. ataba N		*!			*		

Gradient constraint ranking for the gender markedness constraints, following Boersma and Hayes (2001), entails the assumption that the constraints (*N, *F, *M) are initially more highly ranked. They are adjusted downwards over the course of the L1 acquisition process in response to their lexical type frequency. This entails that the markedness constraint(s) referring to the rarer gender(s) may occasionally remain ranked above the other gender constraints, thereby enforcing an output form of another gender (and in this case, a different final vowel). Such a case is exemplified in Table 11 below.

Table 11. Exceptional feminine gender assignment

/atabe/	*N	*-E→M, F	*-A→M, N	*-C→F, N	FAITH	*F	*M
a. atabe N	*!						
b. atabe F		*!				*	
c. → ataba F					*		*
d. ataba N	*!		*		*		

It is precisely such cases, I argue, which generate the feminized loanword outputs seen in BCS (and its neighbors). Because the markedness constraint rankings are

stochastic and tightly linked with the prevalence of each gender in the lexicon, the observed match between lexical distribution and the number of 'feminized' loanwords is an entirely predictable outcome.

I assume that the winning outputs are thereafter consistently selected via something like the USELISTED constraint(s) proposed by Zuraw (2000), in order for them to continue in usage for both the loan-borrowing individual and, eventually, the speech community.

The long-term persistence of highly-ranked *N would lead to the disappearance of new neuters, as in BCS, and possibly eventually all neuters. Such a shifting ranking might be instantiated by social factors such as increasing dispreference for Turkish loanwords due to the emergence of nationalism and independence movements in the Balkans over the course of the 19th century. Such movements quite typically include a stance on reducing or eliminating loanwords in general, and of the former imperial language in particular, and the region of ex-Yugoslavia was no different in this regard (Kamusella, 2012).[3] Precisely how these factors interact with formal grammatical mechanisms remains to be elucidated.

6.   Conclusions

I have argued that the adaptation of Turkish loanwords in an earlier stage of BCS is influenced by speakers' sensitivity to what the distribution of gender in the lexicon 'should' be. This is instantiated in the grammar in a model which stochastically reflects the proportions of grammatical variables in the lexicon.

These proportions can change over time, due both to normal lexical turnover and social change. In the case of BCS, it is plausible that emerging social dispreference for words of Turkish origin led to a lessening of input /e/-final forms overall, which fed into the decline of neuter as a class.

References

Ban, L. & Matovac, D. (2012). On Turkish loanwords in Croatian language. *Trakya Üniversitesi Edebiyat Fakültesi Dergisi*, 2:4, 155–166.

Becker, M., Ketrez, N. & Nevins, A. (2011). The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language*, 87:1, 84–125.

Boersma, P. & Hayes, B. (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry*, 32, 45–86.

---

3.   Anecdotally, such an attitude was also expressed at a presentation of some of this data in Bulgaria. One member of the audience, a literary scholar, commented in the discussion period that the changes under discussion obviously took place because otherwise the items sounded "too Turkish."

Cahen, L. (1920). *Serbian-English and English-Serbian pocket dictionary*. London: Kegan Paul.

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, 24, 143–188.

Ernestus, M. & Baayen, H. (2003). Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language*, 79, 5–38.

Friedman, V. (2003). *Turkish in Macedonia and beyond: Studies in contact, typology and other phenomena in the Balkans and the Caucasus*. Turcologica 52. Wiesbaden: Otto Harrassowitz Verlag.

Hudson Kam, C. L. & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, 1:2, 151–195.

Hudson Kam, C. L. & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, 59:1, 30–66.

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Kamusella, T. (2012). *The politics of language and nationalism in modern central Europe*. Basingstoke, UK: Palgrave.

Manova, S. & Dressler, W. U. (2001). Gender and declensional class in Bulgarian. *Wiener Linguistische Gazette*, 67, 45–81.

Mønnesland, S. (1973). Loanwords in the two variants of Serbo-Croatian. *Scando-Slavica*, 19:1, 197–205.

Rice, C. (2006). Optimizing gender. *Lingua*, 116, 1394–1417.

Simonović, M. (2010). Lexeme formation, morphological parsing and constituency in loanword adaptation: A case study from Slovenian and Serbo-Croatian. Poster presented at the 14th International Morphology Meeting, Budapest, Hungary.

Sjoberg, M. J. (1972). Turkish loanwords in the language of Aleksa Šantić. PhD dissertation, The Ohio State University.

Škaljić, A. (1966). *Turcizmi u srpskohrvatskom jeziki* [Turkisms in the Serbo-Croatian language]. Sarajevo: Svjetlost Izdavačko Preduzeće.

Trueswell, J. C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language*, 35, 566–585.

Walter, M. A. (2011). Probability matching in Arabic and Romance morphology. In E. Broselow & H. Ouali (Eds.), *Perspectives on Arabic Linguistics*. Amsterdam: John Benjamins.

Walter, M. A. (2015). Probability matching and Macedonian gender morphology: The case of Turkish loanword adaptation. Presentation at the 11th Conference on Formal Description of Slavic Languages. Potsdam, Germany.

Walter, M. A. (2017). Frequency distributions as faithfulness targets: Or, why Bulgarian feminized Turkish nouns. In Y. Oseki, M. Esipova, & S. Harves (Eds.), *Proceedings of the 24th Meeting of Formal Approaches to Slavic Linguistics*. Ann Arbor, MI: Michigan Slavic Publications.

Wonnacott, E. & Newport, E. L. (2005). Novelty and regularization: The effect of novel instances on rule formation. In A. Brugos, M. R. Clark-Cotton, & S. Ha (Eds.), *BUCLD 29: Proceedings of the 29th Annual Boston University Conference on Language Development*. Somerville, Mass.: Cascadilla Press.

Zuraw, K. (2000). Patterned exceptions in phonology. PhD dissertation, UCLA.

Index of Subjects

Index of Languages & language families

The collected articles in this volume address an array of cutting-edge issues in the field of historical linguistics, including new theoretical approaches and innovative methodologies for studying language through a diachronic lens. The articles focus on the following themes: I. Case & Argument Structure, II. Alignment & Diathesis, III. Patterns, Paradigms, & Restructuring, IV. Grammaticalization & Construction Grammar, V. Corpus Linguistics & Morphosyntax, VI. Languages in Contact. Papers reflect a wide range of perspectives, and focus on issues and data from an array of languages and language families, from new analyses of case and argument structure in Ancient Greek to phonological evidence for language contact in Vietnamese, from patterns of convergence in Neo-Aramaic to the development of the ergative in Basque. The volume contributes substantially to the debate surrounding core issues of language change: the role of the individual speaker, the nature of paths of grammaticalization, the role of contact, the interface of diachrony and synchrony, and many other issues. It should be useful to any reader hoping to gain insight into the nature of language change.