Current Issues in Linguistic Theory 353

All Things Morphology

Its independence and its interfaces

EDITED BY
Sedigheh Moradi
Marcia Haag
Janie Rees-Miller
Andrija Petrovic

JOHN BENJAMINS PUBLISHING COMPANY

ALL THINGS MORPHOLOGY

CURRENT ISSUES IN LINGUISTIC THEORY

AMSTERDAM STUDIES IN THE THEORY AND HISTORY
OF LINGUISTIC SCIENCE – Series IV
ISSN 0304-0763

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 353

Sedigheh Moradi, Marcia Haag, Janie Rees-Miller and Andrija Petrovic (eds.)

All Things Morphology
Its independence and its interfaces

ALL THINGS MORPHOLOGY

ITS INDEPENDENCE AND ITS INTERFACES

Edited by

SEDIGHEH MORADI
Stony Brook University

MARCIA HAAG
The University of Oklahoma

JANIE REES-MILLER
Marietta College

ANDRIJA PETROVIC
Stony Brook University

JOHN BENJAMINS PUBLISHING COMPANY
AMSTERDAM & PHILADELPHIA

The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

DOI 10.1075/cilt.353

Cataloging-in-Publication Data available from Library of Congress:
LCCN 2021016873 (PRINT) / 2021016874 (E-BOOK)

ISBN 978 90 272 5963 9 (HB)
ISBN 978 90 272 5974 5 (E-BOOK)

John Benjamins Publishing Company · https://benjamins.com

Table of contents

Part V. Interfaces

CHAPTER 1

All things morphology

An introduction

Marcia Haag[1], Sedigheh Moradi[2], Andrija Petrovic[2] and Janie Rees-Miller[3]
[1]University of Oklahoma / [2]Stony Brook University / [3]Marietta College

It is of course too big a promise that this volume would deliver summaries of important work in all subfields of the vast and ancient discipline of morphology, like a medieval street fair of arcane morphological ideas. This is not an encyclopedia, despite the title. Rather, we have attempted here to provide a view of where the field has been and where it is today within a particular theoretical framework, gathering up new and representative work in morphology by both eminent and emerging scholars, and touching on a very wide range of topics, approaches, and theoretical points of view. The place where the authors of these seemingly disparate articles have a common touchstone is in their focus on a word-based, paradigmatic approach to morphology, which was brought to attention in the modern era by the likes of Matthews (1965), Aronoff (1976, 1994), and Anderson (1992). The chapters of this book show that this perspective accounts for both the independence of morphology and for its interactions with syntax and phonology. One prominent name missing from the list of contributors is Mark Aronoff. This is hardly an accident: this volume was conceived as a way to shine light on the importance of his work by contributors who in virtually all cases have collaborated with or been mentored by Prof. Aronoff. We hope that he will be honored with the research we have produced and see his influence in it.

Keywords: morphology, autonomous morphology, word-based morphology, morphological paradigms

1. Reaching back: Reinterpreting traditional views

Time and again we see the resurgence of ideas from the past as well as the influence, both historically and currently, of ideas from other disciplines. Although the term morphology was only coined in the 19th century to refer to the study of form in any scientific field (Aronoff 1983: 355), grammarians have been conducting morphological analysis for centuries. Indeed, some 25 centuries ago, Panini's

https://doi.org/10.1075/cilt.353.01haa

Sanskrit grammar included approximately 4000 grammatical rules, many of which dealt with morphological processes such as word formation by compounding and suffixation as well as with morphological alternants created through replacement or blocking (Kiparsky 1993). Morphological concerns were also of importance for Roman grammarians with their focus on noun declensions and verb conjugations. Grammarians of classical Arabic pointed to the symmetry and logic of inflectional endings as evidence of the ultimate perfection of the Arabic of the Quran, one more proof that it was literally the word of God (Campbell 2017: 98–99). In the traditional grammar of inflected languages, a table of all the inflected forms of a particular part of speech served as a model for other words of the same declension, showing an "interdependence that facilitates processes of analogical generalization" (Milin & Blevins 2020). Hockett (1954: 210) dubbed such an approach the word and paradigm model.

Far from being forgotten in the mists of time, these early grammarians still have influence in the modern era. The work of Panini, for example, was influential for 19th century linguists working on comparative morphology and provided a model for Bloomfield's grammatical descriptions of various Algonquian languages (Kiparsky 1993: 291). The use of paradigms continued into the 20th century; for example, the standard Hebrew grammar of the 19th-early 20th century (Gesenius, Kautzsch & Cowley 1910 [1966]) devotes a special appendix to paradigms of personal pronouns and verbs. Some half century later, Matthews (1965) resurrected the paradigm as a way of conceiving of morphology.

2. Morphemes and the autonomy of morphology

The term morpheme was coined in 1880 by Jan Baudouin de Courtenay, who defined it in 1895 as:

> … that part of a word which is endowed with psychological autonomy and is for the very same reason not further divisible. It consequently subsumes such concepts as the root (radix), all possible affixes, (suffixes, prefixes), endings which are exponents of syntactic relationships, and the like.
>
> (Baudouin de Courtenay 1895 [1972]: 153; Stankiewicz's translation)

Baudouin de Courtenay's idea of a morpheme is fundamental to so-called canonical morphology, which is – in an ideal world – concatenative, regular and productive, isomorphic in form and content, and compositional such that the overall meaning can be understood from the parts (Corbett 2003; Hippisley & Stump 2016). Alas, as teachers of introductory linguistics find when looking for easy morphology problems aside from the ubiquitous exercises in Turkish, this ideal is far from messy reality.

A somewhat younger contemporary of Baudouin de Courtenay, Ferdinand de Saussure did not utilize this idea of the morpheme but rather saw the word as the basic sign (Anderson 2015; Audring & Masini 2018: 9); the sign itself was the linguistic unit, intimately uniting two parts – the concept or 'signified' (*signifié*) and the 'sound image' or 'signifier' (*signifiant*) (Saussure 1915 [1959]).

Meanwhile, in America in the early 20th century, Franz Boas and his student Edward Sapir sought to document Native American languages, many of which were fast disappearing. Boas established and Sapir elaborated a procedure for describing morphology, labeled by Hockett (1954) as Item and Process. In this approach, a derived linguistic form could be analyzed as the result of a process applied to an underlying simple form, the root. This means of description allowed for some purely phonetic material in the derived form as a marker of the process.

Within American Structuralism, in contrast with the Item and Process approach of Boas and Sapir, the morpheme-based Item and Arrangement description was implicit in Bloomfield's chapters on grammar (Hockett 1954: 213). Bloomfield (1933: 161) defines a morpheme as "a linguistic form which bears no partial phonetic-semantic resemblance to any other form." Assuming that the component parts determine final meaning, Bloomfield stated, "linguistic study must always start from the phonetic form and not from the meaning" and asserted that at least part of the meaning depends upon the arrangement of morphemes (Bloomfield 1933: 162–163).

For much of the 19th and into the 20th centuries, morphology held a position of central importance in linguistic study. However, with the revolution of generative grammar in the mid-20th century, morphology lost its autonomy. Early generative grammar allowed for two main components, namely syntax and phonology, with morphology relegated to a subsidiary role (Aronoff 1983: 356). The renewed importance of morphology in generative linguistics was due in part to Chomsky's *Aspects of the theory of syntax* (1965). In this book Chomsky radically revised the relation between syntactic structures and lexical items. He proposed a view of terminal syntactic positions as collections of syntactically relevant features; as lexical items would be listed with features that characterize them, lexical insertion would then allow the association of a lexical item with a terminal position containing a consistent set of features. These features can be properties like [±Count], [±Animate], but also morphosyntactic properties like Gender (inherent to the lexical entry) or Number (representing a value on a paradigmatic dimension), or morphological properties like Declension Class. Chomsky overtly argues for his reinterpretation of a more traditional, paradigmatic analysis, and against morpheme-based accounts (Anderson (2016) provides an informative review of Chomsky's early views on morphology).

In the same year as *Aspects* was published, Peter Matthews (1965) revived and formalized word-and-paradigm morphology, seeing morphology as a separate component from syntax that was composed of four sets of primitive elements, namely: lexemes, phonological or graphological elements, morphosyntactic properties (such as imperative, singular, accusative, etc.), and morphosyntactic categories (such as mood, voice, case, number, etc.). In the paradigmatic structure, all forms of a lexeme would be paired with the appropriate property and category. In contrast with a morpheme-based approach, as Anderson (2017) explains, word-and-paradigm does not treat words as units of syntax subject to syntactic rules and procedures. Instead, the word-and-paradigm approach "describes these words in terms of their connections to one another in form and content" (2017: 4).

Lest we give the impression that all work in morphology was heading toward a word-based or word-and-paradigm approach, the morpheme-based approach got a new lease on life with the introduction of distributed morphology, very much a syntax-based theory (Halle & Marantz 1993). It might be fair to say that morphemes are essential to the syntax-based theories of morphology, where they are treated as the minimal units of syntactic combination. "Within such a theory, morphemes are subject to a recursive merge operation that builds hierarchical structures of constituents" (Marantz 2015: 160), and morphology is subject to the same constituent structures as syntax. In fact, it is syntax that generates structure through operations like Move and Merge. Only after the syntactic structure is in place, are underspecified vocabulary items inserted. There is no lexicon as such and work that lexicalist theories assign to the lexicon has instead been distributed (Harley & Noyer 1999). In this theory, not only is morphology not autonomous, but it also essentially disappears altogether into syntax.

As a summary of the development of morphological theories, Stump (2001) proposes two pairs of contrasting criteria to distinguish morphological theories. Criteria 1 and 2 are in contrast with each other, and criteria 3 and 4 are in contrast with each other:

1. **Lexical**: (morphosyntactic) content and (phonological) form are listed in a lexicon, and the word-form is assembled from lexical items.
2. **Inferential**: "[T]he associations between a word's morphosyntactic properties and its morphology" are "expressed by rules or formulas" (Stump 2001: i), and forms are inferred by rules from stems. Some stems, for example *sing* and *dance* (with past tense *sang* vs. *danced*), will have different rules from each other for the same notion, in this case PAST.
3. **Realizational**: A word bears a given content property exclusively as a concomitant of a specific formal realization. In other words, the content of the word-form determines the form.

4. **Incremental**: The presence of a given element of content licenses a specific realization but does not depend on it. In other words, the content of the word-form is built up from the sum of its parts.

As Hippisley & Stump (2016: 9) conclude, the contrast between current theories boils down to whether morphology and syntax are a single system or whether they are distinct, each with its own principles. In other words, is morphology autonomous?

3. Autonomous and word-based morphology: Aronovian perspective

If we begin with Aronoff's *Word formation in generative grammar* (1976) as a starting point for modern theories of word-based morphology, two of the first enduring problems that Aronoff treats are allomorphy and the nature of stems. He defines allomorphy rules as those that are separate from phonology but bring about phonological change. Even though the most influential idea from *Word formation in generative grammar* is that of word-based productive derivation, considerable space is devoted to allomorphy that is at least partially productive.

In his discussion of Latin stems, Aronoff (1994) shows that stems are neither phonological nor syntactic units. Aronoff's attention to stems contrasts with the main idea of earlier word-based morphology: some languages (including Latin as a prime example) clearly build words based on more than one stem, the forms and not the meanings of which are the bases for predictable word formation whose derivatives are divorced semantically and even syntactically from the morphosyntactic grounding of the original stem. Equally important is attention to the paradigm as an empirical morphological concept. As Aronoff (2018: 12) states,

> This traditional approach to word formation provided an intuitively satisfying solution to the problem of the morpheme that my work on Latinate roots had uncovered. If derivation is not a matter of combining morphemes but of attaching affixes to naked words, then we don't need all the morpheme components of words to be meaningful and we don't need the internal semantics of words to be compositionally derived from these components. All we need is for naked words to be meaningful. We don't need to worry about morphemes at all, only naked words and what the derivational affixes do with them.

Citing the 1530 work of Palsgrave, Aronoff (2018) notes that this traditional approach allowed him to see word formation as "words formed from words".

This view of morphology as word-based and autonomous brings us to the contents of this book, the chapters of which exemplify the myriad ways in which these basic concepts have been explored in recent linguistic research.

4. Outline of the book

4.1 Paradigms

Part I opens with Farrell Ackerman's "Making sense of morphology: Foxes, hedgehogs and a calculus of infinitesimals" (Chapter 2). The word and paradigm view of morphology is the framework within which Ackerman presents his thoughts influenced by traditional insights concerning the systematic organization of morphological phenomena. The chapter reiterates the importance of "exploring conceptual and methodological insights from other disciplines which successfully address the nature of complex (adaptive) systems." It also argues in favor of "developing old and sometimes forgotten insights from language study and investigating them with new quantitative methodologies and detailed data sets."

In Chapter 3 ("A formal restriction on gender resolution"), Sedigheh Moradi presents another way of using paradigms in conjunction with notions external to the mainstream study of language. She combines a typologically motivated hierarchy of gender with the mathematical notion of monotonicity to probe the upper bound on the attested variation in gender resolution rules. Her chapter proposes that all the patterns of resolved gender agreement are instances of monotonic mappings from the base hierarchy of gender to surface forms.

4.2 Words, stems, and affixes

Several authors have contributed chapters that touch on developments in the concepts of words, stems, and affixes; the chapters in Part II (re)examine the definitions and properties of these notions in natural languages via a veritable cornucopia of methods, including a case study from data collected in fieldwork, computational methods, corpus study, and experimental methods.

Wendy Sandler enlarges on her seminal work in the Al Sayyid Bedouin sign language with her case-study "Signs and words" (Chapter 4), which examines a large number of morphological types of signs. She concludes that the word is a particular and primary way that sign language is organized.

Olivier Bonami and Sacha Beniamine take up the question of how to decide whether some bit of form is part of a stem allomorph or an independent exponent. They use computational methods to examine discontinuous stems as inflectional exponents themselves rather than as types of stem allomorphy in "Leaving the stem by itself" (Chapter 5).

Two chapters examine the visual aspect of stems and words – their orthography – that creates morphological cues to their lexical, grammatical, or structural nature. Caroline Hettwer and Nanna Fuhrhop, in "Stem constancy under the microscope: A systematic language comparison of types and limitations of stem spelling" (Chapter 6), report on a corpus study of stem constancy in Dutch, German, French, and English that examined stem spellings with double consonant letters. They conclude that the strength of stem constancy varies with the language and that the different types of deviations are systematic in their own ways.

Kristian Berg, in "Major lexical categories and graphemic weight" (Chapter 7), shows that the English spelling of stems preserves an important distinction between lexical and function words. This evolved over time in a process of self-organized spelling that produced a minimal three-letter stem for lexical words but two for function words as an aid to reading.

The last two chapters in Part II approach stems and affixes experimentally, providing evidence of how speakers may be representing and processing morphologically complex words. Carlo Semenza in "Word formation in the brain: Data from aphasia and related disorders" (Chapter 8) offers neurological evidence that stems and affixes are represented separately in the brain. He also finds that morphology persists in the absence of the ability to retrieve phonological forms. Headedness is shown to be psychologically real and neurally founded. This chapter draws attention to the importance of interdisciplinary work and sets the ground for further integration of theoretical linguistics, neuropsychology, cognitive neuroscience, neurolinguistics and psycholinguistics, for the research of morphological processes.

Alice Harris and Arthur Samuel also present a study based on experimental methods. They report on several lexical decision experiments in "The suffixing preference: A preliminary report on processing affixes in Georgian" (Chapter 9). Contrary to the hypothesis that argues for "the suffixing preference", Harris and Samuel's Georgian-speaking subjects were able to respond to Georgian words with prefixes more accurately and/or faster than to suffixed words. On the basis of these results, the authors argue that, contrary to the (widely accepted) suffixing preference hypothesis, prefixes may be easier to process than suffixes.

4.3 Competition, inheritance, and defaults

Part III explores some ways in which morphology is non-canonical, even in languages that have been said to be very close to a fully canonical system (such as Ripano, Chapter 12). The first two chapters in this section deal with competition, the existence of which is already non-canonical by itself. The notion of competition derives from the field of ecology, in which it was noted that no two species can coexist in equilibrium if they are competing for the same ecological niche. Aronoff (2019) observes that ecological niche differentiation is similar to linguistic alternants competing for the set of environments in which to occur, leading to the phenomenon of complementary distribution. The elsewhere or default principle (also called Panini's principle) fits into this scheme since once different variants are parceled out into different niches, the default takes care of what is left. The third and final chapter of the section shows an example of exceptional default agreement with non-canonical controllers, in a language in which three types of default coexist.

In Chapter 10, "Feature-based competition: A thousand years of Slavonic possessives," Greville Corbett analyzes the feature Number as a competitor in selecting forms of possessive versus person pronouns. Studying data from the Slavic languages, Corbett focuses on the use of the genitive case versus adjective-like forms in possessive expressions. He finds that the competition results in a variety of outcomes throughout the Slavic language family, and that it can result in the creation of different niches accommodating the competing forms.

Anna Thornton in Chapter 11 turns to Romance languages in her exploration of competition. "Competition in Romance comparatives" describes the competition between synthetic and paraphrastic comparatives from Latin through numerous Romance daughters. She finds several situations: extinction of the suppletive form, blocking of the analytical form when a suppletive one exists, and co-occurrence of both – which tend to settle into niches, similar to the phenomenon in Slavic languages described in Chapter 10. Thornton concludes her chapter with a corpus analysis of Italian, demonstrating the usefulness of this type of investigation for future researchers.

Remaining in the Romance family, Michele Loporcaro and Tania Paciaroni turn to Ripano (Italo-Romance) in Chapter 12. "Multi-layered default in Ripano" is another chapter that treats hierarchies and their resolutions. Loporcaro and Paciaroni isolate three types of default in the language (one morphological, two syntactic), where the normal-case (syntactic) default is maximally specified, while the syntactic exceptional-case and the morphological default have the same realization. In this way, the authors effectively propose an inheritance hierarchy for nominals that also includes sensitivity to syntax. The hierarchy thus accounts for gender assignment in a language that has overt gender dependent on syntactic context – a feature that makes Ripano virtually unique in the existing literature (but see Chapter 14 for context-dependent morphomic exponents).

4.4 Morphomes

Morphomes, as instances of purely morphological form unassociated with meaning, have been invoked in the literature as the strongest support for the autonomy of morphology. Aronoff (1994) conceived of the important concept of the morphome, which has led to both refinement of the concept and its employment in analyzing numerous data problems from languages with widely varying structures. The chapters that belong to Part IV thus focus on the morphome as a core concept, and show data that reinforce not just the importance but indeed the indispensability of morphomes in linguistic theory.

Andrew Spencer in “Morphomes all the way down!” (Chapter 13) addresses the architecture of Paradigm Function Morphology (PFM; Stump 2001), as a prime example of a realizational, lexeme-based model of morphology. Such an understanding of morphology was sketched out in Aronoff (1994), and the important concept of the morphome has motivated the development of PFM and similar theoretical models. Spencer introduces coherence, and therefore simplification, into the most recent version of PFM by claiming that all features in Form paradigms are morphomic, whereas all Content paradigm features are interpretable, and interface with syntax/semantics. This, Spencer argues, holds across-the-board, even for canonical systems, where Form paradigm features appear to be directly related (in a one-to-one correspondence) to non-morphomic Content paradigm features.

Gregory Stump in “Conditional exponence” (Chapter 14) provides data from Breton verb inflection, and shows that there are morphomic exponents whose content is sensitive to context. This is then explained as an effect of rule combination (rule composition, supplementational rule combination). Stump thus examines core concepts of PFM, but by explicating his claim formally he contributes to a wider discussion on morphological theory, which has significance beyond PFM itself.

The three chapters that follow use the morphome to analyze seemingly intractable problems in very different languages. Robert Hoberman straightforwardly accesses it in “My favorite morphome: The Arabic suffix AT” (Chapter 15), wherein he analyzes the Arabic suffix AT as clearly a morphome rather than a “clutch of accidentally homophonous suffixes” with diverse functions.

In Marcia Haag’s chapter “In further pursuit of the adjective: Evidence from the Siouan language Osage” (Chapter 16), she uses morphomes to revisit a famous problem in many Native American languages: the lack of derivational morphology and syntactic distribution that would signal lexical categories.

Finally, Stela Manova and Georgia Knell (“Two suffix combinations in native and non-native English: Novel evidence for morphomic structures,” Chapter 17) devise an experimental study to show that English suffix combinations are stored in the lexicon as morphomes. This is a finding that is in agreement with natural language processing research on how patterns, rules, and semantics can be induced based solely on form relations. The authors also discuss how morphomic knowledge can be used for educational purposes, and for testing language proficiency.

4.5 Interfaces

One of the important characteristics of autonomous morphology is that, despite its autonomy, it still interacts with other parts of language, and Part V offers quite different perspectives on these important connections. Part V opens with two chapters on the interface of morphology and phonology. Stephen Anderson's "A short history of phonology in America: Plus c'est la même chose, plus ça change" (Chapter 18) provides a useful perspective since phonology both conditions morphology and is conditioned by it, and is a critical part of how and why these interesting and unruly forms emerge. The chapter examines what happens to particular frameworks over time – perhaps they are in competition.

Zheng Xu's "Realization Optimality Theory: A constraint-based theory of morphology" (Chapter 19) addresses the interface of morphology and phonology by proposing a unified and coherent account of morphological form via Optimality Theory. Xu argues that Optimality Theory (OT) provides a way of accounting for morphological phenomena with ranked violable realization constraints, markedness constraints, and morphotactic constraints. In this way, Xu applies OT principles to model phenomena like blocking, extended exponence, affix ordering, and syncretism. As OT is originally a phonological framework, Realization OT is especially appropriate for representing the interaction of morphology and phonology.

Not only does morphology interface with phonology and other areas of interest to theoretical linguists, but it also interfaces with sociolinguistics and applied linguistics generally. Not unlike other chapters in this volume, the next two chapters rely heavily on language data; in addition, however, these chapters also examine the sociolinguistic and cultural variables underpinning the linguistic phenomena at hand.

Janie Rees-Miller's "A-prefixing in the ex-slave narratives" (Chapter 20) is a corpus-based examination of data from ex-slave narratives collected in the 1930s and demonstrates that a-prefixing was widely used in African-American speech at the time, and was not limited to white speakers, as has been postulated. The chapter begins with a focus on phonological constraints and the morpho-syntactic patterns in which a-prefixing is found. These constitute the author's form-based analysis; Rees-Miller, however, supplements this with a meaning-based analysis, and shows that the semantic and discourse functions of a-prefixing are crucial to understanding this morphological process.

Dorit Kaufman in "Trajectory of children's verb formation in Hebrew as a heritage language" (Chapter 21) explores on-going competition and innovation in word formation in Israeli immigrant children in the United States who are acquiring both Hebrew and English. With language shift and attrition, morphological interaction and innovative lexical constructions are attested in the language of young native Hebrew speakers in the US, while their Hebrew derivational morphology and verb formation mechanisms prove to be vulnerable and susceptible to the increasing diffusion of English lexical formation.

As for applied linguistics more generally, when we described the chapters on orthography in Part II, we noted that each of those chapters had implications for reading. Here, carrying the theme of orthography forward into the history of reading instruction in the United States, Edwin Battistella offers "A primer for linguists on the reading wars" (Chapter 22), finishing with a description of our current "politically fraught landscape." Thus, we conclude the book having moved from the rarified world of theoretical morphology to its applications in the "real world" of 21st century society.

5. Conclusion

As this brief introduction makes clear, the current field of inquiry in morphology ranges far and wide. It has grown out of an ancient and august tradition. From the time of Panini in the 4th century BCE through the grammarians of classical languages and into the 19th and 20th centuries with the early developments in modern linguistic science, certain recurring notions about morphology and morphological analysis have formed the backdrop of inquiry. From the mid-20th century to the present, work by scholars such as Peter Matthews, Stephen Anderson, Gregory Stump, and particularly Mark Aronoff has brought back the notion of paradigms and viewed morphology as word-based and autonomous. These scholars and others in like vein investigate morphological phenomena from the perspective that sees morphology as both independent of other branches of language but also interfacing with these branches.

The chapters in this book elaborate on these basic themes, from the further exploration of paradigms (Part I) to studies involving words, stems, and affixes (Part II), to examinations of competition, inheritance, and defaults (Part III), to investigations of morphomes (Part IV), to consideration of the various ways that morphology interacts with other parts of the language from phonology to sociolinguistics and applied linguistics (Part V). In all these various areas of research, we are not starting de novo, but working within the context of a long and venerable tradition. Like Isaac Newton and Bernard of Chartres before him, we "stand on the shoulders of giants."

References

Anderson, Stephen R. 1992. *A-morphous morphology*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511586262

Anderson, Stephen R. 2015. The morpheme: Its nature and use. In Matthew Baerman (ed.), *Oxford handbook of inflection*, 11–34. Oxford: Oxford University Press.

Anderson, Stephen R. 2016. The role of morphology in transformational grammar. In Andrew Hippisley & Gregory Stump (eds.), *The Cambridge handbook of morphology*, 588–608. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781139814720.021

Anderson, Stephen R. 2017. Words and paradigms: Peter H. Matthews and the development of morphological theory. *Transactions of the Philological Society* 115(1). 1–13. https://doi.org/10.1111/1467-968X.12090

Aronoff, Mark. 1976. *Word formation in generative grammar*. Cambridge, MA: MIT Press.

Aronoff, Mark. 1983. A decade of morphology and word formation. *Annual Review of Anthropology* 12. 355–375. https://doi.org/10.1146/annurev.an.12.100183.002035

Aronoff, Mark. 1994. *Morphology by itself*. Cambridge, MA: MIT Press.

Aronoff, Mark. 2018. Morphology and words: A memoir. In Olivier Bonami, Gilles Boyé, Georgette Dal, Hélène Giraudo & Fiammetta Namer (eds.), *The lexeme in descriptive and theoretical morphology*, 3–17. Berlin: Language Science Press.

Aronoff, Mark. 2019. Competitors and alternants in linguistic morphology. In Franz Rainer, Francesca Gardini, Wolfgang Dressler, & Hans Christian Luschützky (eds.), *Competition in inflection and word formation*, 39–66. Berlin: Springer. https://doi.org/10.1007/978-3-030-02550-2_2

Audring, Jenny & Francesca Masini. 2018. Introduction: Theory and theories. In Jenny Audring & Francesca Masini (eds.), *The Oxford handbook of morphological theory*, 1–19. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199668984.001.0001

Baudouin de Courtenay, Jan. 1895 [1972]. An attempt at a theory of phonetic alternations. In Edward Stankiewicz (ed.), *A Baudouin de Courtenay anthology*, 144–212. Bloomington, IN: Indiana University Press.

Bloomfield, Leonard. 1933. *Language*. New York: Holt.

Campbell, Lyle. 2017. The history of linguistics. In Mark Aronoff & Janie Rees-Miller (eds.), *The handbook of linguistics*, 2nd edn., 97–117. Hoboken NJ: Wiley Blackwell. https://doi.org/10.1002/9781119072256.ch6

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Corbett, Greville. 2003. Agreement: Canonical instances and the extent of the phenomenon. In Geert E. Booij, Janet DeCesaris, Angela Ralli, and Sergio Scalise (eds.), *Topics in morphology: Selected papers from the Third Mediterranean Morphology Meeting*, 109–128. Barcelona: Universitat Pompeu Fabra.

Gesenius, Heinrich Friedrich Wilhelm, Emil Kautzsch & Arthur Ernest Cowley. 1910 [1966]. *Gesenius' Hebrew grammar*, 2nd English edn. Oxford: Clarendon Press.

Halle, Morris & Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In Morris Halle & Samuel J. Keyser (eds.), *The view from Building 20*, 111–176. Cambridge, MA: MIT Press.

Harley, Heidi & Rolf Noyer. 1999. Distributed morphology. *Glot International* 4(4), April. 3–9.

Hippesley, Andrew & Gregory Stump. 2016. Introduction. In Andrew Hippesley & Gregory Stump (eds.), *The Cambridge handbook of morphology*, 1–24. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781139814720.001

Hockett, Charles F. 1954. Two models of grammatical description. *Word* 10. 210–231. https://doi.org/10.1080/00437956.1954.11659524

Kiparsky, Paul. 1993. Paninian linguistics. In Ron Asher (ed.), *The Encyclopedia of Language and Linguistics*, 1st edn., Volume 6, 2918–2923. New York: Pergamon.

Marantz, Alec. 2015. Morphology. In Gregory Hickok & Steven L. Small (eds.), *Neurobiology of language*, 153–163. San Diego: Academic Press.

Matthews, Peter H. 1965. Some concepts in word-and-paradigm morphology. *Foundations of Language* 1(4). 268–289.

Milin, Petar & James P. Blevins. 2020. Paradigms in morphology. *Oxford research encyclopedia of linguistics*, https://oxfordre.com/linguistics/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-551. (22 August, 2020) https://doi.org/10.1093/acrefore/9780199384655.013.551

Saussure, Ferdinand de. 1915 [1959]. *Course in general linguistics*. Translated by Wade Baskin. New York: McGraw-Hill.

Stump, Gregory. 2001. *Inflectional morphology: A theory of paradigm structure*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486333

PART I

Paradigms

CHAPTER 2

Making sense of morphology

Foxes, hedgehogs and a calculus of infinitesimals

Farrell Ackerman
UC San Diego

Aronoff (2016) argues for the value of interpreting different approaches to the analysis of morphology as reflecting the sensibilities of foxes and hedgehogs as characterized in Berlin (1997). It is argued that Berlin also provides a way to go beyond these oppositional sensibilities and the morphological theories they develop by exploring a complex systems perspective on morphological organization of the sort adumbrated by certain late 19th century and early 20th century linguists: systems that arise from the dynamic co-activity of surprisingly many factors. The kind of view they envisioned is now quantitatively and computationally practicable and can benefit from research on complex behaviors within the developmental sciences. A view of this sort is motivated by two types of empirical evidence, one concerning implicative paradigm organization in Māori and Baale, the other concerning the dynamic relation between diachrony and synchrony as attested in Mari and Beserman Udmurt.

Keywords: developmental science, implicative organization, morphologization, morphological organization, paradigms, complex systems, variable morphotactics

> The study of the development of living organisms, including humans, has evolved from a field dominated by dichotomous either/or approaches (e.g., either psychogenic explanation or biogenic explanation) to an interdisciplinary approach to the life span that recognizes the scientific value of integrating multiple perspectives – biological, psychological, sociocultural, historical – into a synthetic, holistic, complex, coactional system.
> (Overton & Molenaar 2015: 12)

> Comparative biology provides sophisticated ways to think about commonalities that underlie biological diversity. Bringing order to that diversity is not about identifying universal elements, but about finding order in the patterns of similarity and difference.
> (Griffiths 2011: 328)

https://doi.org/10.1075/cilt.353.02ack

1. Introduction

Aronoff (2016) reflects on Isaiah Berlin's essay *The hedgehog and the fox: An essay on Tolstoy's view of history.* Berlin famously offers a taxonomy of thinkers categorized into hedgehogs (those guided primarily by a single big and synthesizing idea or method) and foxes (those guided primarily by the empirical enchantments of arresting specificities) and he analogizes it to opposing, even warring, sensibilities within linguistics over the past 60 or so years. While he recognizes that certain linguists or some linguists in certain research display properties of both types, he provides a picture of a field essentially split between reduction minded tree-theoretic grammarians and empirical minded puzzlers: he wonders whether a reconciliation is possible. In this brief reflection, I will suggest that Aronoff, in the interdisciplinary company of numerous others, is engaged in an exciting reconciliation of the sensibilities associated with hedgehogs and foxes, which, effectively, answers this basic question. Moreover, the spirit of this answer is consistent with the less remembered, but central, purpose of Berlin's exploration, namely, Tolstoy's views on explanation in history as the dynamic confluence of innumerable factors:

> How would an ideal historical science operate? By using a kind of calculus whereby this 'differential', the infinitesimals – the infinitely small human and non-human actions and events – would be integrated, and in this way the continuum of history would no longer be distorted by being broken up into arbitrary segments.
> (Berlin 1997: 459)

Aronoff himself (2017: 444) explores the seeds of such a synthesis by connecting Darwin's observations about evolutionary change as a way of redressing the modern imbalance between synchrony over diachrony as sources of linguistic explanation:[1]

> A century later, though, it has become clear that an evolutionary account of what Darwin himself was referring to when he wrote that "The survival or preservation of certain favoured words in the struggle for existence is natural selection" (Darwin 1871: 61) can provide just such a framework, allowing historical explanation of language to return to its rich relation with evolution and evolutionary theory.

Beyond the competition among words, there is a need for recalibrating and reconceptualizing the relation between diachrony and synchrony in the explanation of organization for morphological systems. To paraphrase a famous phrase from

1. The explanatory role of diachronic changes for synchronic phenomena has been a recurrent source of investigation in typologically oriented and grammaticalization approaches as typified in Blevins (2004). It is also reflected in such research as that found in Bybee (1985), Harris (2008, 2017), Anderson (2016), among many others.

evolutionary biology, nothing can be understood about synchronic grammars except in the light of diachrony, where diachrony can be understood as the manifold of dynamic interacting factors that obtain both in historical change and ontogenetic mastery of a language. Accordingly, the basic questions are what kind of relations between historical change and synchronic patterns do we posit within the domain of linguistic morphology and how might we understand this in terms of guidance from evolutionary theory. And perhaps most importantly, which among the very different views about (phylogenetic) evolution and (ontogenetic) development should we draw our instruction from? What are likely to be the most illuminating conceptualizations of development over time, and what are the appropriate tools to use when extending such notions to language morphology?

In line with the epigrams above, I will suggest that the notion 'calculus of infinitesmals' permits morphological foxes to formulate their findings within a rich explanatory framework: in fact, the more and the more accurate the comparative data, the more refined this alternative becomes, and the more refined it becomes, the more reliably we can identify insights about kaleidoscopic (patterns of) cross-linguistic morphological organization. I will suggest that emergent systems of morphological organization are the result of the co-action of co-influencing elements in this calculus of infinitesimals (Overton & Molenaar 2015), yielding the sort of patterns of similarities and differences whose analysis permits us to make sense of morphology (Griffiths 2011). The main goal of this short exploration, then, is to argue that in order to make sense of morphology, adapting Keller's (2010) perspective on making sense in biology, we must get rid of the *mirage between diachrony and synchrony*. This type of multicausal dynamic inquiry is associated with linguistic research paralleling research in the developmental sciences,[2] a perspective on the analysis of complex systems which is often unfamiliar among linguists. It is a perspective consistent with recent Item and Pattern[3] models of morphology

2. See Depew & Weber (1994), Bailey (1996), Gottlieb (1997), Lass (1997), Gilbert & Sarkar (2000), Locke (2009), Sultan (2015), Norde & Van de Velde (2016), and Balari, Giullermo & Sultan (2020), among others.

3. More familiarly referred to as Word and Paradigm (see Blevins 2016). I choose the term 'Item and Pattern' for a simple reason: in practice Word and Paradigm approaches assume that any items that participate in reliable patterns of morphological behaviors, sometimes participating in implicative relations, are primary morphological objects. This includes synthetic and periphrastic expressions. Given the problem of identifying cross-linguistic criteria for wordhood, (Dixon and Aikhenvald 2003; Haspelmath 2011a, 2011b; Bickel & Zuniga 2017; and Tallman 2019, among others) I will use the term 'word' and 'morphological item' or simply 'item' interchangeably in the following, where item refers to any type of surface exponence which displays reliable participation in identifiable morphological patterns. This reflects the proposal in Ackerman & Webelhuth (1998) to replace the term 'word' with 'lexical combinatorial item', where these were

(Ackerman, Blevins & Malouf 2009; Bonami 2015; Sims 2015; Blevins 2016; Blevins et al. 2016; Bonami & Strnadová 2019; Elsner et al. 2019). Within this general enterprise, I will suggest that the need to think systemically for morphological explanation warrants the more expansive use of concepts and tools from the developmental sciences to make sense of morphology, but, obviously, this is a much larger task than can be accomplished here. I can only collate the insights of many others with convergent intuitions and provide a programmatic picture for a small set of phenomena.

I begin in § 2 by identifying the systems intuitions of Mikołaj Kruszewski, typifying a broader class of traditional insights concerning the systemic organization of morphological phenomena. In § 3, I illustrate how assumptions about systems of whole word relatedness are applicable in the analysis of novel verb form generation in Māori and nominal number marking distributions in Baale (Surmic). This demonstrates the explanatory value of systemic paradigmatic organization: it illustrates how patterns of relatedness among morphological items can explain the production of unfamiliar forms without positing underlying or surface forms from which other related forms are derived. It also shows how the residual effects of phonological changes yield material for synchronic morphological organization in terms of implicative relations between (classes of) words.[4] In § 4, I focus directly on grammaticalization by examining morphologically liminal nominal constructions in two Uralic languages, Mari (Luutonen 1997) and Beserman Udmurt (Arkhangelskiy & Usacheva 2015). The development of nominal case-marking from postpositions illustrates the dynamic ongoing interaction between diachrony and synchrony in these languages. I conclude in § 5 with some summarizing observations.

2. Early systemic intuitions concerning Item and Pattern models

The basic outlines of a systems-oriented modern Item and Pattern model of morphology were anticipated in the insights of both Hermann Paul and Mikołaj Kruszewski. In particular, Kruszewski (1995) viewed the morphological system as facilitating two fundamental aspects of language knowledge and usage: 'reproduction' was the

objects participating in morphological patterns which combine with items of other types in clausal relations: this permitted one to refer to morphological objects with very different types of exponence, specifically, synthetic or periphrastic and twilight stages in between. The participants in such patterns are identifiable in each language or relevant language subsystem, but they need not correspond formally to units in other languages which exhibit similar network behaviors.

4. See Wurzel (1987) on implicative organization in inflectional systems.

more or less faithful utterance of stored lexical representations, i.e., fully derived and inflected wordforms and their networks of related forms, while production was the utterance of novel wordforms licensed by the analogical inferences intrinsic to networks of related words.

> … every word is connected by twofold bonds: by innumerable ties of similarity with its relatives according to sounds, structure, or meaning and by equally numerous ties of contiguity with its various fellow travellers in every possible kind of phrase. A word is always a member of certain nests or systems of words and at the same time is a member of certain series of words. This explains the ease with which we memorize and recall words. Moreover, these properties of words make it possible for us not to have to resort to straight memorization every time. It is sufficient for us to know words like *idu* ["(I) am walking"], *idës′* ["(you sg.) are walking"], or *vedu* ["(I) am leading"] in order to produce the new word *vedet* ["(he) is leading"], although we may never have heard it before. In the majority of cases we can not say with certainty which words we have learned from other people and which we have produced ourselves; in the majority of cases, as in the above cited examples, parallel forms make it possible to produce only one form, regardless of who is producing it. For this reason W. von Humboldt early on pointed to the perpetual creativity of language. (Kruszewski 1995: 97)

Kruszewski here suggests that the production of a novel inflected form for the Russian verb *vesti* "lead" is guided by knowledge of other forms of *vesti* as well as other inflected forms of the different verb *idti* "to go". This represents, according to him, a clear example of an essential challenge presented to theory for language analysis, namely, the 'perpetual creativity of language'. In this framework, words (following Kruszewski 1995; Bybee 2003) can be usefully conceptualized as bundles of phonological, lexical and morphosyntactic properties. The relevant entities arise as identifiable units on the basis of frequency distributions keyed to speaker usage (Bybee 2003, 2007). Given a usage-based conception of wordhood propounded by Hermann Paul, as discussed in Fertig & Hopper 2015, there is no reason to restrict morphology to the study of synthetic wordforms,[5] hence the broader term morphological item used herein. Such parallelisms in the organization of declensional, conjugational, and derivational relations recall Kruszewski's observation that "language forms a harmonious whole". He illustrates this by distilling rich patterns of implicative relations that suffuse and define the Russian inflectional and derivational morphological system. I quote him at length because of the richness and relevance of his insights.

5. See Robins (1959), Matthews (1991) and Fertig & Hopper (2015), among others.

> The more precisely we define the type of a given category of words, the more precisely we can answer the question as to what other category it presupposes. Thus, for example, the [Russian] NOM. SG. with the complex *-oro-*, the suffix *-a*, and the stress on this suffix (*borodá* ["beard"], *skovorodá* ["frying pan"]) presuppose the NOM. PL. with the suffix *-y* and the stress on the first syllable (*bórody*, *skvórody*…). Likewise, two syllable verbs with stress on the ending *-it′* (*nosít′*…) presuppose three syllable verbs with the ending *-ivat′* and stress on the root vowel (*nášivat′*…); moreover, if a verb of the first category has an (etymological) *o* or *a* in the root, the verb of the second category has a stressed *a*; a verb with a *-d-* presupposes a verb with a *-ž-* (*xodit′* – *xaživat′*); alongside *-t-* we encounter *-č-* (*platit′* ["to pay"] – *vyplačivat′* ["to pay off (debts)"]); alongside a labial, a complex with the same labial and an *-l′-* (*kormit′* ["to feed"] – *vskarmlivat′* ["to rear"]); alongside *-s-*, *-š-* (*nosit′* – *našivat′*), etc. Thus, it goes without saying that every word category is more or less related to and dependent on many word categories. The reason is that, despite all deviations, language forms a harmonious whole. (Kruszewski 1995: 148–149)

In this passage, Kruszewski identifies alternations that support reliable inferences in verbal derivation, involving verbal prefixes and frequentative suffixes, e.g., *ževat′* "to chew" ~ *perežëvyvat′* "to masticate" and nominal inflection, involving nominative singular and nominal plural marking, e.g., *borodá* "beard" ~ *bórody* "beards". These implicative relations parallel those attested for inflection previously. The force of this observation is clear: morphological organization reflects the patterns of relations that emerge when systems of alternations among words are identified. This means that each wordform, connected to other wordforms as part of a system of relations, bears some information with respect to the patterns in which it participates. This leads to a more nuanced notion of the diagnostic value of wordforms than exemplified in the traditional notion 'principal part': principal parts are those forms that are most predictive of other related forms, while if we take morphological organization to be a web of relations among wordforms, we would expect all forms to be predictive to some degree or other.

These earlier insights of Kruszewski and his contemporaries find clarifying expression in more recent explorations of systemic relatedness among words. In particular, Bochner (1993) develops a word-based model of morphology which he calls Lexical Relatedness Morphology (LRM). The guiding intuition behind this proposal is that the lexicon, following Jackendoff (1975), consists of full entry lexical representations related by redundancy rules interpreted as patterns for (sets of) lexical entries. In Bochner's (1993) and Barr's (1994) developments of Lexical Relatedness Morphology, the existence of full lexical representations associated with each other by pattern statements permits the calculation of cost associated with the existence of specific words, either already stored or newly generated. The greater the match or fit between a lexical representation and the pattern it instantiates, the less the cost of this item. In contrast, the more the lexical representation deviates from established patterns in any ways concerning discriminable differences in the constitution of the words, the greater its cost to the grammar.

Making sense of language morphology requires identifying how the organization of internal structure for (complex) words, serves to discriminate one word from another and how resulting systems of discriminably different morphological items cohere into patterns of relatedness between items that participate in morphological patterns: where morphemic composition is only one among many stunning strategies of discriminability.[6] Morphology, accordingly, needs to be conceptualized in terms of the systemic organization of parts and wholes: the internal parts that are constitutive of morphological items as wholes and the morphological items themselves as parts constitutive of systemic patterns of organization viewed as wholes.

The guiding assumption concerning systemic relatedness among items participating in patterns is explored in Hayes (1998: 2), where he proposes to "adopt a rich set of string mappings (incorporating phonology and morphology) that relate members of the paradigm in pairwise fashion. This tentatively is proposed by Bochner (1993); see also the computational references cited by Sproat (1992: 215–216)." He provides the following schematic in order to visualize the nature of the network of relationships between surface morphological items:

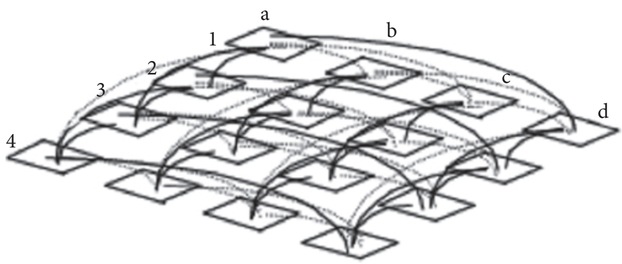


Figure 1. Mappings between surface exponents of morphological items from Hayes (1998: 2)

Following recent Item and Pattern proposals, I assume that understanding morphological organization does not depend upon privileging any one particular encoding strategy (say, morpheme-based agglutinative combinatorics in a language such as Turkish) out of the huge repertoire of attested and possible encoding strategies seen in the languages of the world (say, the coalitions of stem internal changes associated with different morphosyntactic properties sets involving segmental length, tone, vowel height and voice quality in a language such as Dinka). Rather, the task of

6. See Ramscar et al. (2017) and Marzi, Ferro & Pirrelli (2019) for discussion.

morphology is to identify fundamental principles, in the spirit of Paul (1890), of morphological organization which license the unimaginable variations in morphological expression that is manifestly learnable wherever it appears.

3. The mirage of space between diachrony and synchrony: Systemic morphological organization in Māori and Baale

The value of identifying systemic relations between, e.g., synthetic wordforms can be seen by examining a classic problem in the morphophonological literature, namely, accounting for allomorphic variation in Māori verbal forms.[7] In this language, the passive variant is associated with a suffix *-ia*, often preceded by a single consonant which does not appear in the active form of the related verb. The basic pattern is observable in the 12 active-passive pairs in Table 1:

Table 1. Pairs of missing consonants and correlated values of overt consonants

Class	Word type	Active	Passive	Gloss
1	X:**Xia**	maka	makaia	"throw"
2	X:**Xa**	whiu	whiua	"chase"
3	X:**Xtia**	ahwi	ahwitia	"embrace"
4	X:**Xkia**	hopu	hopukia	"catch"
5	X:**Xmia**	inu	inumia	"drink"
6	X:**Xina**	aroha	arohaina	"love"
7	X:**Xna**	tahu	tahuna	"burn"
8	X:**Xngia**	tohu	tohungia	"point out"
9	X:**Xnga**	kai	kainga	"eat"
10	X:**Xria**	mau	mauria	"carry"
11	X:**Xhia**	kimi	kimihia	"seek"
12	X:**Xwhia**	whao	whaowhi	"put into"

There have been essentially two types of analyses proposed to explain these distributions. One is a phonological approach, which posits an abstract underlying stem-final consonant that is covert in the active but surfaces overtly in the passive. The other is morphological and posits as many conjugation classes as there are distinct allomorphs: thus, each -(C)*ia* represents a distinct allomorphic variant. There is an important similarity between these approaches: they are predicated on the assumption that passive forms are based upon active forms, and derived from the latter in some way. One of the perennial problems associated with this direction

7. See Barr (1994: 490&494) on Māori. See also Hale (1968, 1973), Kenstowicz & Kisseberth (1979), Sanders (1990), Blevins (1994), Harlow (2007). This discussion results from collaborative conversations with J. P. Blevins.

of derivation (from active to passive) is its stipulative nature, often relying on questionable notions of markedness where the active is assumed to be less marked than the passive. If markedness is related to frequency of occurrence, then from the perspective of stimuli available to the learner of Māori, it has been observed that they are more likely to encounter the passive variant than the active: "It has long been recognised and remarked on that the passive is used much more frequently in Māori than in … English, and even, in certain types of text, more frequently than the corresponding active" (Harlow 2007: 28).

This suggests that learning the formal relationships between active and passive variants may be assisted by familiarity with the passive, where passives are simply more informative than actives concerning the specification of final consonants. In this view, it is not that a specific consonant in the active is synchronically covert and underlying, but that it has disappeared diachronically and does not have a synchronic role in the active forms at all: this loss doesn't matter synchronically for the generation of unencountered forms, since knowledge of the patterns of surface exponence provide an alternative resource for the generation of forms. Uncertainty reduction in the generation of novel forms of familiar words is ensured by patterned systemic organization, where knowledge of the consonant value in the passive is trivially predictive of the absence of that consonant in the active. Where morphological organization is designed to facilitate relatedness among words, it is expected that there might be other forms in the language to support correct inferences about previously unencountered words than just the active and passive forms, i.e., that the resources of a language would provide other cues to the, e.g., final consonant of verbs, and that this would obviate the need to posit a phantom synchronic consonant in the active forms. In fact, this expectation is evident in Māori, where gerundial nominalizations exhibit the same final consonant as is attested in the passive form of any given lexeme. This is illustrated in Table 2:

Table 2. Correlated allomorphy in passives and nominalizations

Consonant	**Active**	**Passive**	**Nominalization**	**Gloss**
–	mahi	mahia	mahinga	"work"
–	noho	nohoia	nohoanga	"sit on"
k	moto	moto**k**ia	moto**k**anga	"punch"
k	tomo	tomo**k**ia	moto**k**anga	"pass in/out"
m	inu	inu**m**ia	inu**m**anga	"drink"
m	tanu	tanu**m**ia	tanu**m**anga	"bury"
r	mau	mau**r**ia	mau**r**anga	"seize"
r	tau	tau**r**ia	tau**r**anga	"come to rest"
t	mahue	mahue**t**ia	mahue**t**anga	"be left"
t	(pu)puri	puri**t**ia	puri**t**anga	"hold"
h	(yi)tiro	tiro**h**ia	tiro**h**hanga	"look"

As can be seen, access to the surface distributions of existing variants of words can provide clues for the correct production of previously unencountered words. Accordingly, directionality of derivation and the need for abstract representations essential for the derivation of passives from actives can be construed as artifacts of a particular analytical approach. This familiar approach provides what the Gestalt psychologist Wolfgang Köhler referred to as a 'tranquilizing explanation' (Henle 1971), an illusory solution that in presenting the problem as settled forecloses asking the sorts of questions that could disclose less familiar or utterly different types and sources of explanation. The value of imagining alternatives becomes particularly clear in Barr's (1994: 468) neglected word-based analysis to this problem, where the essential insight is that (1) exposure to either the passive or the nominalized form provides information about the active form; (2) knowing the passive or the nominalized forms is enough to predict the unencountered form of either as well as the active form; and (3) Māori possesses a word-based system organized into multiple patterns of lexeme relatedness, differing in frequency and regularity, but providing enough information to reliably generate correct new forms of words. Among other properties such a proposal additionally accounts for how speakers could know the value of the consonants missing in the active forms and the identity of consonants across the passive and nominalized forms. On this type of analysis problems posed to the learner by the evident sparsity of data are obviated by the patterned and implicative morphological organization of words. The basic relations can be visualized as in Figure 2:

$$\begin{bmatrix} \textit{/moto /} \\ \textit{V} \\ \textit{`punch'} \end{bmatrix} \begin{bmatrix} \textit{/motokia /} \\ \textit{V} \\ \textit{`be punched'} \end{bmatrix} \begin{bmatrix} \textit{/motokanga /} \\ \textit{N} \\ \textit{`punching'} \end{bmatrix}$$

$$\begin{bmatrix} \textit{/ X /} \\ \textit{V} \\ \textit{`v'} \end{bmatrix} \begin{bmatrix} \textit{/ Xkia /} \\ \textit{V} \\ \textit{`be v – ed'} \end{bmatrix} \begin{bmatrix} \textit{/ Xkanga /} \\ \textit{N} \\ \textit{`v – ing'} \end{bmatrix}$$

$$\begin{bmatrix} \textit{/ X /} \\ \textit{V} \\ \textit{`v'} \end{bmatrix} \begin{bmatrix} \textit{/ Xmia /} \\ \textit{V} \\ \textit{`be v – ed} \end{bmatrix} \begin{bmatrix} \textit{/Xmanga /} \\ \textit{N} \\ \textit{`v – ing'} \end{bmatrix}$$

$$\begin{bmatrix} \textit{/ X /} \\ \textit{V} \\ \textit{`v'} \end{bmatrix} \begin{bmatrix} \textit{/ Xhia /} \\ \textit{V} \\ \textit{`be v – ed'} \end{bmatrix} \begin{bmatrix} \textit{/ Xhanga /} \\ \textit{N} \\ \textit{`v – ing'} \end{bmatrix}$$

$$\begin{bmatrix} \textit{/ X /} \\ \textit{V} \\ \textit{`v'} \end{bmatrix} \begin{bmatrix} \textit{/ Xtia /} \\ \textit{V} \\ \textit{`be v – ed'} \end{bmatrix} \begin{bmatrix} \textit{/ Xtanga /} \\ \textit{N} \\ \textit{`v – ing'} \end{bmatrix}$$

Figure 2. Schemata for correlated allomorphic variation

The preceding representative array of alternative patterns encodes minimally distinct relations between related words and establishes particular expectations among them. For example, if a speaker knows the consonant value associated with the passive form, then the speaker also knows that this is the same value for the nominalized form of the verbal lexeme. Similarly, knowing the consonantal values of these forms trivially permits formation of the active, which simply does not contain that consonant. Additionally, since the patterns essentially vary with respect to the consonantal value of their allomorphs, knowing one particular pattern encourages expectations about the terms of the relations in the others: the reliability of consonant identity between the passive and nominalized forms creates the expectation that this is likely to be the same for all allomorphic pairs.

A remarkably similar, but less known, example of this type of allomorphic distribution, occurs in the Eastern Nilotic Surmic language Baale, as discussed in Dimmendaal (2000).[8] Baale differs from the more conservative related Surmic language Murle with respect to the loss of word final consonants. This contrast is typified with respect to the cognate forms found in Table 3, where the forms for "chicken" and "ax" appear without the final consonants that appear in their Murle counterparts:[9]

Table 3. Cognates in Baale and Murle (from Dimmedaal 2000: 223)

Baale	Murle	
kówú(t̪)	kówut	"chicken"
mɛɛlɛ́(k)	mɛlɛk	"ax"

Baale, like many of the Nilotic languages, participates in an extraordinarily complex system of number marking.[10] This concerns how (classes of) nominals are marked for singular and plural number. The pattern for non-nominative forms found in Table 4 is relevant for present purposes. While the singulars lack final consonants of different values, their plural congeners contain geminated variants of these missing consonants followed by some tonal value associated with *-a*.

8. Both the Māori and Baale cases are discussed together in less detail in Dimmendaal (2011).

9. Dimmendaal provides the phantom missing consonants in the orthography but indicates that they are null in the accompanying phonetic transcription.

10. See Dimmendaal (2000) for a detailed discussion. It is important to observe that Dimmendaal (1987, 2011) has been developing a dynamic systems perspective on language change for many years.

Table 4. Allomorphic realization of -*Ca* in Baale (from Dimmendaal 2000: 223)

Singular	Plural	Gloss
ŋɛrɛ́(t̪)	ŋɛrɛ́-**t̪-t̪**ǎ	"healer, witch doctor"
agá(t̪)	agá-**t̪-t̪**ǎ	"tongue"
kówú(t̪)	kówú-**t̪-t̪**á	"chicken"
keyí(c)	keyí-**c-c**á	"udder"
álle(c)	álle-**c-c**á	"stool, headrest"
mɛɛlɛ́(k)	mɛɛlɛ́-**k-k**ǎ	"ax"

As with Māori, one might specify the covert final consonant of the singulars and this would establish the value of the absent consonant that is be geminated in the plural. Or, as with Māori, one might explore a systemic explanation by searching for forms other than plural which could specify the value of the missing consonants in the singular. In fact, Baale contains several such surface distributions of forms. Consider the following non-nominative nouns which belong to the geminate allomorph class (from Dimmendaal 2000: 220–221).

(1)	a.	ɲamɔ́	"razor blade"	b.	ɲamɔ́**t̪t̪á**	"razor blades"
(2)	a.	álle	"stool, headrest"	b.	állé**ccá**	"stools, headrests"
(3)	a.	mɛɛlɛ́	"ax"	b.	mɛɛlɛ́**kká**	"axes"

When these nominals appear with possessive suffixes, the final consonant becomes overt:[11]

(1)	c.	ɲamɔ́-**ð**-a-naandí	"my razor blade"
(2)	c.	álle-**j**-a-naandí	"my stool, headrest"
(3)	c.	mɛɛlɛ́-**g**-a-naandí.	"my axe"

Similarly, the lexeme relevant final consonant is overt in the context of the nominative case suffix as in (4):

(4) *attíjja* *ɲamɔ́-**ð**-ɛ* *ai*
IMPF.be.sharp razor-NOM very
"The razor is very sharp"

In sum, an enormous amount of information crucial for the prediction of unencountered forms is contained in the implicative organization of fully formed words. Thus, the value of surface patterns for fully inflected and derived words may help to explain speaker mastery of words not yet heard or seen. Implicative relations provide sufficient structure to circumvent the need for exposure to comprehensive

11. The overt singleton consonants are voiced between vowels.

sets of stimuli. This accords with the fact that the observed Zipfian distributions of linguistic forms entails that most wordforms will be very rare. The systemic relational organization of words is a way to solve the so-called Zipfian Paradigm Cell Filling Problem discussed in Blevins, Milin & Ramscar (2017), especially in languages with gigantic inventories of form variation distributed across many distinct patterns of relatedness.

This perspective on morphological organization, thus, solves the presumed curse of dimensionality problem concerning learnability, namely that children cannot in real time narrow the set of hypotheses that would facilitate mastery of their target linguistic systems. From the present perspective, factorial explosions concerning the possible combinatorics of elements are constrained by systemic organization itself: progress in developing command of the relations which define a system requires a (gradual) narrowing of the possible options shaped by the trajectories intrinsic to the system: imagined combinatorial explosions and attendant problems arise (largely) outside of a system and, often in neglect of systemic organization, while the system itself necessarily figures as a guiding constraint on language particular options; this is why understanding systemic organization is fundamental to understanding the learnability of morphology. This is no more mysterious, one could argue, than what has been assumed in the development of motor functions as discussed in Thelen & Bates (2003: 381):

> … when the many, heterogeneous elements that produce movements – nerves, muscles, joints, metabolic processes – cooperate together in a task, they cohere together in a way that is more complex than the sum of the parts. Here was a brilliant solution to Bernstein's degrees of freedom problem [in order to solve motor developmental problems [FA]. Rather than the degrees of freedom being a curse, a computational load to be overcome, they become the very source of organization and indeed, of change. Patterns emerge from the complexity of the system and its energetic status. As such, no component or element has priority or privilege, since it is the particular coalition of elements from which coherence arises. Thus, movements need not be 'represented' in the nervous system in all their detail. The details are created in their assembly.

Such parallels suggest the utility for the analyses of (complex) morphological systems to attend to ideas, tools, and models that have been instructive for understanding physical, biological, and cognitive complex systems.

When we look at how the systems in Māori and Baale work, i.e., the establishment of implicative organization that defines declension classes of words, we see that historical sound changes have produced synchronic consequences usefully organized for the prediction of previously unencountered forms.

4. The mirage of space between diachrony and synchrony

A different type of effect between diachrony and synchrony is evident in morphological change where parts of morphological items display a kind of liminal status with respect to standard morphological and syntactic behaviors. As noted by numerous linguists (Lewy 1911; Sapir 1921; Bybee 2007; Luutonen 1997, among others), the internal structure of words in several languages exhibits weak phonological fusion and syntactic independence (Lewy's *Suffixlockerheit* "loose affixation") for their constitutive pieces; this becomes most evident in variable morphotactics, suspended affixation and in the accretion of nominal case-markers from postpositions within the Uralic languages. The diachronic dimension (Schwegler 1990; Harris & Campbell 1995) of multi-word expressions comprehends constructions that trace a synthetic to periphrastic trajectory from strong phonological fusion and syntactic atomicity for morphophonologically integrated synthetic words to total phonological and syntactic independence of constitutive pieces:[12] I provide a sense of this type of phenomenon from Meadow Mari nominal morphosyntactic marking and Beserman Udmurt postposition marking, both of which provide synchronic variation in the structures used to express the same meanings.

The nominals of Meadow(-Eastern) Mari (Permian branch of Uralic), as described in Alhoniemi (1985), Luutonen (1997), and Riese et al. (2012), display 9 case-marking distinctions (CM) as well as a complete paradigm for possessive marking (PNM). The morphotactics of the suffixes associated with case and possession exhibit three distinct linear order patterns. Two of these patterns are exemplified below for the lexeme *muno* "egg" (data from Alhoniemi 1985: 76):

Pattern 1: STEM – PNM – CM, where CM = {genitive, accusative}

(5) *mun-ə̂ž-***ə̂n**
egg-3SG-GEN
"of his egg"

Pattern 2: STEM – CM- PNM, where CM = {inessive, illative, lative, comitative}

(6) *munə̂-***štə̂***-žo*
egg-INESSIVE-3SG
"in his egg"

The third pattern contrast with these two in that there is 'free variation' among the exponents of the relevant morphosyntactic properties, as well as for the expression of plural number:

12. This is likely one of the main reasons that it is problematic to establish cross-linguistically reliable criteria for the identifications of the construct word (see references in footnote 3).

Pattern 3: 'Free' affix order: CM-PNM v. PNM-CM; PL-PNM v. PNM-PL, where CM = {dative, modal} and NUMBER = {pl}

(7) *mun-**ə̂ž-lan*** ~ *munə̂-**lan-že***
egg-DAT-3SG egg-3SG-DAT
"to his/her egg"

(8) *olma-**m-βlak*** ~ *olma-**βlak-em***
apple-1SG-PL apple-PL-1SG
"my apples"

The free variation displayed by these affixes interacts with 'loose affixation' in Mari. Luutonen (1997: 44–52) observes that Mari displays degrees of synthesis among the pieces of its wordforms and phrases. He proposes (1997: 51) a scale of synthesis/analysis (= periphrasis) to characterize the differences between, e.g., Finnish, and Mari. He represents the scale as follows in terms of degrees of analyticity, where $>_{an}$ means 'is more analytic than':

Finnish postposition constructions
$>_{an}$ Mari postposition constructions
$>_{an}$ Mari plural forms
Mari case forms

According to this, for example, Finnish postposition constructions are more analytic (periphrastic) than those of Mari. This is illustrated by the contrast between Finnish in (9) and Mari in (10): (from Luutonen 1997: 48):

(9) a. *poikani-**kin** kanssa*
boy.1SG-**also** with
"also with my boy"
b. *poikani kanssa-kin* — Finnish
boy.1SG with-also
"also with my boy"

(10) a. *joltaš-**at** δene
friend-**also** with
b. *joltaš δen-**at*** — Mari
friend with-also
"also with friend"

These languages differ with respect to whether they tolerate the interposition of pragmatic/emphatic clitics between a nominal complement and its postpositional head. Since it is possible for these clitics to attach to either the noun or postposition in (9) for Finnish, but only to the postposition (10b) in Mari, there appears to be a tighter bond between the noun and postposition in Mari than in Finnish.

Another distinguishing behavioral property between these languages is suspended affixation, whereby an erstwhile bound affix is elided from conjuncts in a conjoined construction, while the affix remaining on one of the conjuncts has scope over all of the conjoined forms. This is exemplified by the contrasts in (11) and (12): (from Luutonen 1997: 45)

(11) a. *naapuri-ø ja ystävä-**lle**
b. *naapuri-**lle** ja ystävä**lle*** Finnish
neighbor-ALL and friend-ALL
"to the neighbor and friend"

(12) a. *pušku.o-ø ða joltaš-**lan***
neighbor and friend-DAT
"to the neighbor and friend"
b. *pušku.o-**lan** ða joltaš-**lan*** Mari
neighbor-DAT and friend-DAT

A third phenomenon examined by Luutonen is the transitional nature of Mari plural suffixes: these are suffixes that participate in 'free variation' as described above. He demonstrates that the Mari plural suffix *šamə̂č* is similar to postpositional constructions with respect to several criteria (Luutonen 1997: 49). Focusing on one, person/number markers from the possessive paradigm exhibit alternative orderings in combination with both postpositions and *šamə̂č*. This is illustrated in (13) and (14).

(13) a. *joltaš-**em** δene* ~ b. *joltaš δene-**m*** Postposition
friend-**1SG** with friend with-**1SG**
"with my friend" "with my friend"

(14) a. *joltaš-**em**- šamə̂č* ~ b. *joltaš- šamə̂č-**em*** Plural marker
friend-1SG-PL friend-PL-1SG
"my friends" "my friends"

Though not emphasized by Luutonen, the possessive interpretation evident in (12b) and (13b) is intriguing: when the PNM appears suffixed to the postposition and plural marker in the (b) forms, it denotes the possessor of the noun. This is the sort of interpretation one would expect if the PNM appeared within the word over which it had scope: it is not expected if the PNM appears on a word which is merely the head of an ordinary syntactic phrase. This is especially intriguing since the identical form of the, e.g., postposition has a pronominal, not a possessive, interpretation when it occurs without a lexical NP complement. Thus in (15) the PNM functions as the incorporated 1st singular pronoun.

(15) (*məəj*) *δene-**m***
1SG-NOM with-**1SG**
"with me"

Finally, the strategy we have seen in (12b) of alternatively marking the noun or postposition with a PNM to indicate nominal possession, is a common one in Mari as attested in the representative contrast between (16a) and (16b): (Reise et al. 2012: 202).

(16) a. *pört-**em*** *renčəlno* ~ b. *pört* *renčəəln-**em***

(16) a.	*pört-**em***	*renčəlno*	~ b. *pört*	*renčəəln-**em***
	house-**1SG**	in front of	house	in front of-**1SG**
	"in front of my house"			

A similar distribution of morphological markers is also attested in Udmurt[13] and is investigated carefully for Beserman Udmurt in Arkhangelskiy & Usacheva (2015). Simplifying for present purposes, they argue that inflected postpositions in this language show many parallels with inflected nouns with respect to alternative distributions for both possessive marking and plural marking. This is demonstrated by variants of the possessive construction, as in (17a) and (17b). The dependent nominal *house* in (17a) is unmarked for its local possessor, while this is registered on the head of the possessive construction *window*: in some sense, the first person singular possessive marker is on the 'wrong' constituent, since it has scope over the dependent, not the head to which it is attached. The head in both instances bears the ILL case marker supplying its clausal syntactic role. In contrast, (17b) illustrates a more typical possessive construction, where the dependent element is locally marked for its possessor and case marked by the genitive as the dependent in a possessive construction: the head *window* bears the 3rd person singular marking reflecting the 3rd person singular status of its dependent (from Arkhangelskiy & Usacheva 2015: 18).

(17) a. *Mə̂nam* *korka* *kośag-a-**m*** *šukk-iśk-i-z* *ǯ́ə̂rəĺə̂.*
I.GEN house window-ILL-POSS.1SG hit-DETR-PST-3SG sparrow
"A sparrow bumped into the window of my house."

b. *Mə̂nam* *korka-**je-len*** *kośag-a-z* *šukk-iśk-i-z*
I.GEN house-POSS.1SG-GEN window-ILL-POSS.3SG hit-DETR-PST-3SG
ǯ́ə̂rəĺə̂.
sparrow
"A sparrow bumped into the window of my house."

Parallel distributional alternatives occur with inflectable postpositions, as illustrated by (18a) and (18b) (adapted from Arkhangelskiy & Usacheva 2015: 118).

(18) a. *So* *tə̂b-i-z* *korka* *dor-**a-m*** *uža-nə̂.*
he go.up-PST-3SG house near-ILL-POSS.1SG work-INF
"He went up to my house to work."

13. See Csucs (1990) and Kel'makov & Hännikäinen (1999).

b. *So tâb-i-z* *korka-**je*** *dor-e* *uža-nâ.*
he go.up-PST-3SG house-POSS.1SG near-ILL work-INF
"He went up to my house to work."

In both (18a) and (18b), the postpositions bear the ILL case marking indicating its clausal syntactic role, but they differ with respect to the location of the possessor marking: in (18a), showing a distribution similar to (17a), this case-marker is suffixed to the postposition, while having scope over the dependent nominal. The possessor marker, in contrast, appears on the dependent complement in (18b), displaying the expected local relation of a possessor to the possessed.[14]

Concerning the observed distributions in the constructions Arkhangelskiy & Usacheva (2015: 119) note:

> According to the opinion of the speakers and our observations, there is indeed no difference in meaning between these pairs of utterances. It seems that the position of the affix is influenced by factors like the position of the dependent on the animacy hierarchy rather than by other factors such as topicality, focus, etc. In phrases with noun heads, the possessives are more often seen on the dependents than on the heads. For postpositions, the situation is the opposite, but possessives on dependents are still quite frequent in speech and are often accepted as perfectly grammatical.

Finally, plural markers in this language exhibit the same distributional options and semantic scope peculiarity as postpositions. This is illustrated by the contrasts in (19a) and (19b), where the plural marker on the postposition in (19a) references the number of its nominal complement.

(19) a. *Pušner bud-e* *anal-t-em* *korka.*
nettle grow-PRS.3SG be.left.behind-CAUS-PTCP.PST house
*bord-**jos**-ân.* *gâne*
near-PL-LOC only
"Nettle grows only beside abandoned houses."

b. *Pušner bud-e* *anal-t-em* *korka-**os***
nettle grow-PRS.3SG be.left.behind-CAUS-PTCP.PST house-PL
bord-ân *gâne*
near-LOC only
"Nettle grows only beside abandoned houses."

The striking scopal properties of the possessive markers suffixed to postpositions in Mari (16b) and Udmurt (18a) and the Udmurt plural markers in (19a) suggest that

14. It should be observed a signal difference between the constructions in (17) and (18) is the presence or absence of the 3rd person singular possessive marker in the (b) examples. The absence of this marker in (18b) indicates that it is not a nominal possessive construction, but a postpositional construction.

the postpositions are transitioning to the status of case-markers: as case-markers of their erstwhile complements their scopal properties are construable as the expected local distributions of a noun and its affixes. We can also speculate that the independent existence of inflectable postpositions in these languages provides a systemic motivation for the development of new case markers in these languages.[15] The grammar systems of the languages provide the ingredients for a deployment of familiar pieces for new purposes. The fact that some syntactic elements enter into new types of combinations suggests that they possess a status distinguishable from ordinary syntactic behavior, demanding some sort of special syntax, if not an analysis in terms of periphrastic morphological items.

5. Concluding observations

Berlin's observation about the need to recognize a 'calculus of infinitesimals' for understanding history recalls early linguistic intuitions about the need for detailed description in order to understand morphological organization as a system guided by the development of co-evolving elements. The utility of a systems perspective for making sense of the synchronic morphological patterns in Māori and Baale as well as for the dynamic interplay between diachrony and synchrony in Mari and Beserman Udmurt, benefits from exploring conceptual and methodological insights from other disciplines within the developmental sciences, i.e., psychology and biology, which successfully address the nature of complex (adaptive) systems. It requires developing old and sometimes forgotten insights from language study and investigating them with new quantitative methodologies and detailed data sets. Exploring the nature of evident consilience between different sciences provides a way to engage the necessary sensibilities of both hedgehogs and foxes. In doing so, we will likely discover how modest our abilities are to identify the relevant sources of explanation if unaided by new concepts and techniques. As observed by Camazine et al. (2001: 26):

> In place of explicitly coding for a pattern by means of a blueprint or recipe, self-organized pattern formation relies on positive feedback, negative feedback, and a dynamic system involving large numbers of actions and interactions … environmental randomness can act as the **'imagination of the system'** [emphasis FA], the raw material from which structures arise.

There is reason to expect that we will be startled and instructed by what we find in the imagination of morphological systems.

---

15. For an expansive exploration of how independent elements in a grammar system can yield novelty see Ackerman & Nikolaeva (2013).

Acknowledgements

I thank the members in the Item and Pattern (Word and Pattern) community for their collaborative and congenial interactions over the past 15 years, especially Rob Malouf, Olivier Bonami, Jim Blevins, and Andrea Sims, although nobody else is to blame for what I have written here than me. Additionally, given the scope of the present contribution, I have necessarily omitted many scholars who have better and earlier articulated many of the issues discussed here: I apologize and hope they see me as simply channeling some of their insights.

References

Ackerman, Farrell & Irina Nikolaeva. 2013. *Descriptive typology and linguistic theory*. Stanford: CSLI.

Ackerman, Farrell & Gert Webelhuth. 1998. *A theory of predicates*. Chicago: University of Chicago Press.

Ackerman, Farrell, James P. Blevins & Robert Malouf. 2009. Parts and wholes. Implicative patterns in inflectional paradigms. In James P. Blevins and Juliette Blevins (eds.), *Analogy in Grammar. Form and Acquisition*, 54–81. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199547548.003.0003

Alhoniemi, Alho. 1985. *Marin kielioppi*. Helsinki: Suomalais-ugrilainen seura.

Anderson, Steven R. 2016. Synchronic vs. diachronic explanation and the nature of the language faculty. *Annual Review of Linguistics* 2. 11–31. https://doi.org/10.1146/annurev-linguistics-011415-040735

Arkhangelskiy, Timofey & Maria Usacheva. 2015. Syntactic and morphosyntactic properties of postpositional phrases in Beserman Udmurt as part-of-speech criteria. *SKY Journal of Linguistics* 28. 103–137.

Aronoff, Mark. 2016. A fox knows many things but a hedgehog one big thing. In Andrew Hippisley & Gregory T. Stump (eds.). *The Cambridge handbook of morphology*, 186–205. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781139814720.008

Aronoff, Mark. 2017. Darwinism tested by the science of language. In Claire C. Bowern, Larry Horn & Raffaella Zanuttini (eds.), *On looking into words (and beyond): Structures, Relations, Analyses*, 443–455. Berlin: Language Science Press.

Bailey, Charles J. 1996. *Essays on time-based linguistic analysis*. Oxford: Oxford University Press.

Balari, Sergio, Lorenzo Giullermo & Sonia Sultan. 2020. Language acquisition and EcoDevo processes: The case of the lexicon-syntax interface. *Biological Theory* 15(3). 148–160. https://doi.org/10.1007/s13752-020-00352-9

Barr, Robin C. 1994. A lexical model of morphological change. Cambridge: Harvard University dissertation.

Berlin, Isaiah. 1997. The hedgehog and the fox: An essay of Tolstay's view of history. In Henry Hardy (ed.), *The proper study of mankind: An anthology of essays*, 436–498. London: Chatto and Windus.

Bickel, Bathasar & Fernando Zúñiga. 2017. The 'word' in polysynthetic languages: Phonological and syntactic challenges. In Michael Fortescu, Marianne Mithun & Nicholas Evans (eds.), *The Oxford handbook of polysynthesis*, 158–187. Oxford: Oxford University Press.

Blevins, James. P. 2016. *Word and paradigm morphology*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199593545.001.0001

Blevins, James, Farrell Ackerman, Robert Malouf & Michael Ramscar. 2016. Morphology as an adaptive discriminative system. In Daniel Sidiqqi & Heidi Harley (eds.), *Morphological metatheory*, 271–302. Amsterdam: John Benjamins. https://doi.org/10.1075/la.229.10ble

Blevins, James, Petar Milin & Michael Ramscar. 2017. The Zipfian paradigm cell filling problem. In Ferenc Kiefer, James P. Blevins & Huba Bartos (eds.), *Perspectives on morphological organization*, 141–158. Leiden: Brill Publications. https://doi.org/10.1163/9789004342934_008

Blevins, Juliette. 1994. A phonological and morphological reanalysis of the Māori passive. *Te Reo* 37. 29–53.

Blevins, Juliette. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486357

Bochner, Harry. 1993. *Simplicity in generative grammar*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110889307

Bonami, Olivier. 2015. Periphrasis as collocation. *Morphology* 25(1). 63–110. https://doi.org/10.1007/s11525-015-9254-3

Bonami, Olivier & Jana Strnadová. 2019. Paradigm structure and predictability in derivational morphology. *Morphology* 29(2). 167–197. https://doi.org/10.1007/s11525-018-9322-6

Bybee, Joan. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.9

Bybee, Joan. 2003. *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, Joan. 2007. *Frequency of use and the organization of language*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195301571.001.0001

Camazine, Scott, Jean-Louis Deneubourg, Nigel R. Franks, James Sneyd, Eric Bonabeau & Guy Theraula. 2001. *Self-organization in biological systems*. Princeton Studies in Complexity. Princeton: Princeton University Press. https://doi.org/10.1515/9780691212920

Csúcs, Sándor. 1990. *Chrestomathia votiacica*. Budapest: Tankönyvkiadó.

Darwin, Charles. 1871. *The descent of man, and selection in relation to sex*. London: John Murray.

Depew, David J. & Bruce Weber. 1994. *Darwinism evolving*. Cambridge, MA: MIT Press.

Dimmendaal, Gerrit J. 1987. Drift and selective mechanisms in morphological changes: The Eastern Nilotic case. In Anna G. Ramat, Onofrio Carruba & Giuliano Bernini (eds.), *7th international conference on historical linguistics*, 193–210. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.48.15dim

Dimmendaal, Gerrit J. 2000. Number marking and noun categorization in Nilo-Saharan languages. *Anthropological Linguistics*, 214–261.

Dimmendaal, Gerrit J. 2011. *Historical linguistics and the comparative study of African languages*. Amsterdam: John Benjamins. https://doi.org/10.1075/z.161

Dixon, Robert M. & Aleksandra Y. Aikhenvald (eds.). 2003. *Word: A cross-linguistic typology*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486241

Elsner, Micha, Andrea D. Sims, Alexander Erdmann, Antonio Hernandez, Evan Jaffe, Lifeng Jin, Martha Booker Johnson, et al. 2019. Modeling morphological learning, typology, and change: What can the neural sequence-to-sequence framework contribute?. *Journal of Language Modelling* 7(1). 53–98. https://doi.org/10.15398/jlm.v7i1.244

Fertig, David & Paul J. Hopper (eds.). 2015. *Hermann Paul's Principles of Language History Revisited: Translations and Reflections*, Volume 51. Berlin: Walter de Gruyter GmbH.

Gilbert, Scott F. & Sahotra Sarkar. 2000. Embracing complexity: Organicism for the 21st century. *Developmental dynamics: An official publication of the American Association of Anatomists* 219(1). 1–9. https://doi.org/10.1002/1097-0177(2000)9999:9999<::AID-DVDY1036>3.0.CO;2-A

Gottlieb, Gilbert. 1997. *Synthesizing nature-nurture: Prenatal roots of instinctive behavior*. London: Psychology Press.

Griffiths, Paul. 2011. Our plastic nature. In Snaiy Gissis, Snait B. Gissis & Eva Jablonka (eds.), *Transformations of Lamarckism: From subtle fluids to molecular*, 319–330. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9780262015141.003.0030

Hale, Kenneth. 1968. Review of Hohepa 1967. *Journal of the Polynesian Society* 77. 83–99.

Hale, Kenneth. 1973. Deep-surface canonical disparities in relation to analysis and change: An Australian example. In Thomas Sebeok (ed.), *Current trends in linguistics* 11. 401–458. The Hague: Mouton.

Harlow, Ray. 2007. *Māori: A linguistic introduction*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511618697

Harris, Alice C. 2008. On the explanation of typologically unusual structures. In Jeff Good (ed.), *Linguistic universals and language change*, 54–76. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199298495.003.0003

Harris, Alice C. 2017. *Multiple exponence*. Oxford: Oxford University Press.

Harris, Alice C. & Lyle Campbell. 1995. *Historical syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511620553

Haspelmath, Martin. 2011a. The gradual coalescence into "words" in grammaticalization. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of grammaticalization*, 342–355. Oxford: Oxford University Press.

Haspelmath, Martin. 2011b. The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica* 45(1). 31–80. https://doi.org/10.1515/flin.2011.002

Hayes, Bruce. 1998. On the richness of paradigms, and the insufficiency of underlying representations in accounting for them. Handout of presentation given at Stanford University.

Henle, Mary. 1971. *The selected papers of Wolgang Köhler*. New York: Liveright.

Jackendoff, Ray. 1975. Morphological and semantic regularities in the lexicon. *Language* 51(3). 639–671. https://doi.org/10.2307/412891

Keller, Evelyn Fox. 2010. *The mirage of a space between nature and nurture*. North Carolina: Duke University Press. https://doi.org/10.1215/9780822392811

Kel'makov, Valentin & Sara Hännikäinen. 1999. *Udmurtin kielioppia ja harjoituksia*. Helsinki: Suomalais-ugrilainen seura.

Kenstowicz, Michael & Charles Kisseberth. 1979. *Generative phonology: Description and theory*. San Diego: Academic Press.

Kruszewski, Mikołaj. 1995. *Writings in general linguistics*. Amsterdam: John Benjamins. https://doi.org/10.1075/acil.11

Lass, Roger. 1997. *Historical linguistics and language change*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511620928

Lewy, Ernst. 1911. *Zur finno-ugrischen Wort und Satzverbindung*. Göttingen: Vandenhoeck.

Locke, John L. 2009. Evolutionary developmental linguistics: Naturalization of the faculty of language. *Language Sciences* 31(1). 33–59. https://doi.org/10.1016/j.langsci.2007.09.008

Luutonen, Juola. 1997. *The variation of morpheme order in Mari declension*. Helsinki: Suomalais-ugrilainen Seura.

Marzi, Claudia, Marcello Ferro & Vito Pirrelli. 2019. A processing-oriented investigation of inflectional complexity. *Frontiers in Communication* 4(48). 1–23. https://doi.org/10.3389/fcomm.2019.00048

Matthews, Peter H. 1991. *Morphology*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139166485

Norde, Muriel & Freek Van de Velde (eds.). 2016. *Exaptation and language change*, Volume 336. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.336

Overton, Willis F. & Peter C. M. Molenaar. 2015. Concepts, theory and method in developmental science. In Willis F. Overton & Peter Molenaar (eds.), *Handbook of child psychology and developmental science; Volume 1: Theory and method*, 9–62. Hoboken: John Wiley & Sons. https://doi.org/10.1002/9781118963418.childpsy101

Paul, Hermann. 1970 [1890]. *Principles of the history of language*. College Park: McGrath Publishing Company. Translated from the 2nd edition by Herbert Augustus Strong.

Ramscar, Michael, Melody Dye, James. P. Blevins & Harald Baayen. 2018. Morphological development. In Amalia Bar-On & Dorit Ravid (eds.), *Handbook of communications disorders: Theoretical, empirical, and applied linguistic perspectives*, 181–202. Berlin: De Gruyter.

Riese, Timothy, Jeremy Bradley, Emma Yakimova & Galina Krylova. 2012. *A comprehensive introduction to the Mari language*. University of Vienna, Department of Finno-Ugric Studies: Electronic Version available at https://www.univie.ac.at/maridict/site-2014/book.php?int=0

Robins, Robert H. 1959. In defense of WP. *Transactions of the Philological Society*. 58(1). 116–144. https://doi.org/10.1111/j.1467-968X.1959.tb00301.x

Sanders, Gerald. 1990. On the analysis and implications of Māori verb alternations. *Lingua* 80(2–3). 149–196. https://doi.org/10.1016/0024-3841(90)90019-H

Sapir, Edward. 1921. *Language: An introduction to the study of speech*. New York: Harcourt, Brace and company.

Schwegler, Armin. 1990. *Analyticity and syntheticity: A diachronic perspective with special reference to Romance languages*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110872927

Sims, Andrea D. 2015. *Inflectional defectiveness*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781107053854

Sproat, R. W., & Wilkes, M. V. 1992. *Morphology and computation*. MIT press.

Sultan, Sonia. 2015. *Organism and environment: Ecological development, niche construction and development*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199587070.001.0001

Tallman, Adam J. 2019. Beyond grammatical and phonological words. *Language and Linguistics Compass* 14. 1–14.

Thelen, Esther & Elizabeth Bates. 2003. Connectionism and dynamic systems: Are they really different? *Developmental Science* 6(4). 378–391. https://doi.org/10.1111/1467-7687.00294

Wurzel, Wolfgang U. 1987. System-dependent morphological naturalness in inflection. In Wolfgang U. Dressler, Wili Mayerthaler, Oswald Panagl & Wolfgang U. Wurzel (eds.), *Leitmotifs in natural morphology*, 59–96. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.10.22wur

CHAPTER 3

A formal restriction on gender resolution

Sedigheh Moradi
Stony Brook University

Resolved gender agreement occurs with coordination structures where combinations of two distinct genders are resolved into a different gender value. I have studied a representative sample of languages to understand why only a few of the logically conceivable patterns of agreement are attested. Starting from an underlying hierarchy of gender, I have constructed a pointwise algebra to represent various gender combinations. The gender hierarchy is directly extracted from the organization of the typological data. Combining abstract algebra and the mathematical notion of monotonicity helps us understand the restricted set of attested patterns to only allow monotonic mappings from the base hierarchy to output forms.

Keywords: gender, monotonicity, morphosyntax, gender resolution rules

1. Introduction

Gender, as Hockett defines it, consists of "classes of nouns reflected in the behavior of associated words" (1958: 231). This association gives rise to an agreement relationship between the controller (the element determining agreement, e.g., subject noun phrase) and the target (the element whose form is determined by agreement) (Corbett 2006: 4). If the controller consists in coordination of two noun phrases, e.g., *brother and sister*, then the choice of the target form may create a conflict between the two competing genders. Resolved agreement is a term used to describe the predicate agreement with a subject made up of coordinated elements. The rules that determine the forms to be used are called resolution rules.

Despite the remarkable diversity in grammatical gender systems, the attested gender resolution patterns constitute but a few of the large number of possibilities. Given k possible genders, there are k ways for any two genders and k^{k2} resolution systems. Assuming that the order of elements in a coordination does not matter, the number of resolution systems equals $k^{k(k+1)/2}$. For languages with 2 genders, this yields $2^{2(2+1)/2} = 2^{6/2} = 2^3 = 8$ possibilities. Assume a language with MSC and FEM

https://doi.org/10.1075/cilt.353.03mor

genders. With these two genders, there are 6 possible resolution rules. (In (1) and (2), Cs stand for conjuncts and the subscripts show their gender.)

(1) *Possible resolution rules in a 2-gender system*
(1) $C_f + C_m \rightarrow C_f$ (2) $C_f + C_m \rightarrow C_m$
(3) $C_f + C_f \rightarrow C_f$ (4) $C_f + C_f \rightarrow C_m$
(5) $C_m + C_m \rightarrow C_m$ (6) $C_m + C_m \rightarrow C_f$

There are 8 ways for grouping these resolution rules for any language. Yet only two patterns are attested (Lang 1 and Lang 5).

(2) *Possible languages for a 2-gender system using the resolutions in (1)*
Lang 1: (1), (3), (5) Lang 2: (1), (3), (6)
Lang 3: (1), (4), (5) Lang 4: (1), (4), (6)
Lang 5: (2), (3), (5) Lang 6: (2), (3), (6)
Lang 7: (2), (4), (5) Lang 8: (2), (4), (6)

In this study, I show that the restricted set of attested patterns can be defined by their monotonic nature. In linguistics, monotonicity is mostly associated with semantics, but it has also been used in other aspects of linguistic processing, reasoning, and grammar (Icard & Moss 2014). Graf (2019) proposes monotonicity as a formal universal of morphosyntax. The general idea is based on two criteria: (i) each morphosyntactic domain comes with a base hierarchy (e.g., person: $1 < 2 < 3$), and (ii) the mappings from a hierarchy to surface forms must be monotonic. This account has already been successful in explaining the typology of adjectival gradation, case syncretism, pronoun syncretism, the Person Case Constraint, the Gender Case Constraint, and tense syncretism (Graf 2019; Moradi 2020, 2019).

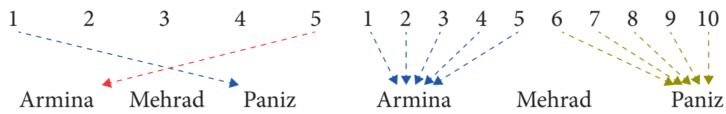
Combining abstract algebra and the notion of monotonicity (as defined in § 2) helps us understand the restricted set of attested patterns. Starting from an underlying hierarchy of gender, we can construct a pointwise algebra to represent various gender combinations. The gender hierarchy is directly extracted from the organization of the typological data. The crucial finding is that even though masculine and feminine genders should be ordered with respect to each other to reflect the ordering of data in individual languages, the general gender hierarchy does not favor one over another. In other words, in a 3-gender system, both $m < n < f$ and $f < n < m$ can keep the system monotonic. One way to look at it is to see gender assignment along a path with two end nodes (masculine and feminine nodes). You can equally use the two nodes to assign gender in a given language. Meanwhile, neuter is negatively defined as neither feminine nor masculine. Those nouns not selected by either end of the hierarchy will be assigned to the middle gender, the neuter.

I start in the next section with a definition of monotonicity and how it can be used for morphosyntactic studies. Based on this general idea, § 3 illustrates the attested patterns of resolved agreement and shows their monotonic nature. I will then conclude the chapter with a short discussion on the implications of this approach.

2. Monotonicity

Monotonicity is a mathematical property that corresponds roughly to the intuitive notion of order preservation. Suppose an ordering relation $\leq$ over a set $\{p, q, r, s, \ldots\}$ such that $p \leq r \leq s$. Then in a monotonic function, one cannot map both p and s to some A without also mapping r to A.

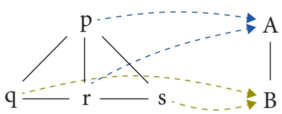
Suppose A is a list of ordered numbers and B is a list of names in alphabetical order. A function f from A to B is monotonic *iff* it preserves the relative order of elements. If f maps 1 to *Paniz* while 5 is mapped to *Armina*, f is not monotonic (this can be seen in crossing branches). However, mapping all the numbers between 1 and 5 to *Armina* and all the numbers from 5 to 10 to *Paniz* still preserves the original order and the function is monotonic.



A linguistically familiar example of linear monotonicity is the ban against crossing branches in autosegmental phonology (Goldsmith 1976). Autosegmental structures are usually presented in tiers, and within each tier segments are linearly ordered. The ban on crossing branches assures that all mappings from tones to segments follow the linear order of the two tiers.

But monotonicity is more general because it is also defined for partial orders. Suppose that $p \leq r \leq s$ as before, and $q \leq p$, but q is unordered with respect to r and s. Then a monotonic mapping could map p and r to A but q and s to B, as shown in (3).

(3) *Monotonic mappings in a partially ordered structure*



Gender resolution rules can form both linear and partial structures depending on the number of gender values that are involved in resolution processes.

3. Gender resolution rules

Conjoined noun phrases make agreement a central problem in languages with grammatical gender. Aside from languages like Yimas, where conjoining noun phrases are prohibited (Corbett 1991: 264), languages use two agreement strategies to determine the realization of gender in the plural: agreement with one conjunct, which is not relevant to this study, and gender resolution rules. Gender resolution rules are quite diverse because they do not always have a unified semantic justification. In French for instance (4), if two nominal heads, one feminine and one masculine, are conjoined, the resolved form is always masculine. Thus the resolution rules in French favor masculine agreement as the default gender. This is different from Icelandic or inanimate coordination in Romanian where neuter and feminine are favored, as respectively shown in (5) and (6).

(4) [le garçon et la fille] sont compétents
[the boy.M and the girl.F] are competent.M.PL
'The boy and the girl are competent.' French

(5) [frægð-ø og fram-i] eru tvíeggj-uð
[fame.F.SG and success.M.SG] are double.edged-N.PL
'Fame and success are double-edged.' Icelandic (Friðjónsson 1991: 90)

(6) [ușa și peretele] ele…
[door.F.the and wall.M.the] they.F.PL…
'The door and the wall, they…' Romanian (Corbett 1991: 288)

Not all languages with gender require gender resolution rules. German and Russian both have three genders that only need agreement in the singular. Agreement in the plural is not problematic since there is only one form for all three genders (neuter). Then there are languages like Luganda and Temne that ban conjunction between human and non-human noun phrases. The preferred construction in such cases is the comitative. In Ojibwa, the same restriction applies to animate and inanimate coordination (Corbett 1991: 264–265). Finally, when the resolutions rules do not apply, agreement is normally with one of the conjuncts, a phenomenon also known as partial agreement, which is not relevant to the current study.

3.1 Patterns of resolution

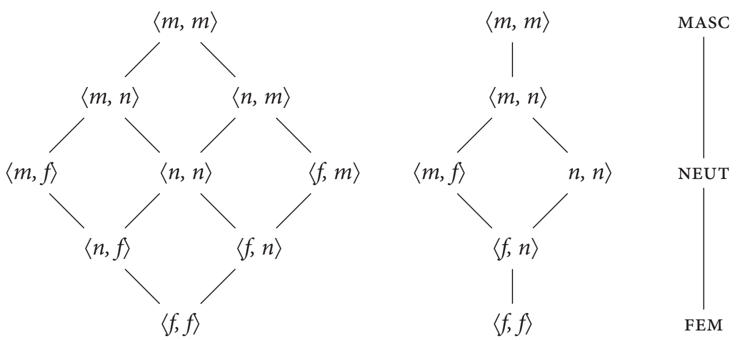
Given k possible genders, there are k ways for combining any two genders and k^{k^2} resolution systems. Assuming that the order of elements in a coordination does not matter, the number of resolution systems reduces to $k^{k(k+1)/2}$.

For languages with 2 genders, this yields $2^{2(2+1)/2} = 2^{6/2} = 2^3 = 8$ possibilities. Yet only two patterns are attested in our sample of seven 2-gender languages (French, Spanish, Latvian, Hindi, Punjabi, Modern Hebrew and Romanian). The same happens in 3-gender languages: out of $3^{3(3+1)/2} = 3^{12/2} = 3^6 = 729$ possibilities only 6 are realized. The space of logical possibilities quickly becomes quite large as more and more genders are added: 8, 729, 1,048, 576 (million), 30, 517, 578, 125 (billion), 21, 936, 950, 640, 377, 856 (quadrillion), etc.

Let us assume an underlying hierarchy of $f < n < m$ and construct a pointwise algebra to represent various gender combinations. At the top of the algebra, $\langle m, m\rangle$ stands for the combination of two masculine genders. At the bottom, $\langle f, f\rangle$ represents the coordination of two feminine noun phrases. All other combinations are ordered between these two nodes. The two edges of the structure expand by combining MASC with NEUT which is the second in the hierarchy. The combination in the center of the next step, $\langle n, n\rangle$, inherits the NEUT gender from the two nodes above it. Then the FEM is introduced to the hierarchy and it combines with the MASC to form the two symmetric nodes in the middle of the structure ($\langle m, f\rangle$ & $\langle f, m\rangle$). Finally there is the combination of the NEUT & FEM genders ($\langle n, f\rangle$ & $\langle f, n\rangle$) right before the final $\langle f, f\rangle$ node. Since in a coordination the order of the coordinated elements does not matter (i.e., $\langle m, n\rangle = \langle n, m\rangle$), we remove all the symmetrically repeated nodes from the previous algebra to arrive at a simplified hierarchy.

Elements of this algebra are then mapped into a hierarchy of plural genders. These mappings, as we will see in the next section, are indeed monotonic for all the languages in our data set.

(7) *The algebra of gender combinations*
(The assumed gender hierarchy is on the right, the simplified gender hierarchy is in the middle, and the hierarchy with repeated nodes is on the left.)



3.2 Typologically attested patterns of resolution

The resolution systems discussed here are primarily based on Corbett's 1991 textbook on gender, which maps out the variation in the gender systems with data from over 200 languages. The four volumes of *Gender Across Languages* (Hellinger & Bussmann 2001; Hellinger & Motschenbacher 2015) that include in-depth discussion of more than 40 languages is our other major source.

Here I present seven representative languages: French, Slovene, Icelandic, Latin, Tamil, Archi and Chibemba. French, Slovene and Icelandic are representative of languages that are argued to have syntactic resolution rules. This means that the classification of gender in these languages is not semantically transparent. Tamil, Archi and Chibemba are examples of semantic resolution, meaning that membership in a gender class is determined by some kind of semantic justification (e.g., animacy, humanness, etc.). And finally Latin is described as a mixed type system where meaning and form are both involved in the patterns of resolution (Corbett 1991: 287).

French

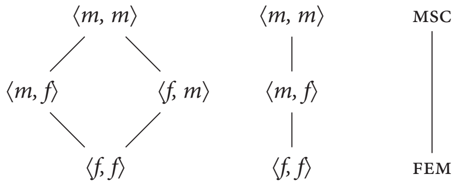
Let us start with the simplest gender system we can consider. In French there are two genders: feminine and masculine. If in a coordination the conjuncts are of the same gender, then that gender will be used as the resolved form. If one conjunct is masculine and another is feminine, then a masculine form is used (see Table 1). Languages like French are quite common, including Spanish, Latvian, Hindi, Italian, Punjabi, Modern Hebrew, etc.

Table 1. Gender resolution in French

	C_{MSC}	C_{FEM}
C_{MSC}	C_{M}	C_{M}
C_{FEM}	C_{M}	C_{F}

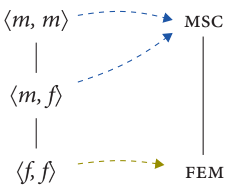
We start by building an algebraic construction based on an underlying hierarchy of gender. Assuming $f < m$ (the hierarchy shown by the right construction in (8)), we construct a pointwise algebra to represent the possible gender combinations. At the top of the algebra, $\langle m, m\rangle$ stands for the combination of two masculine genders. At the bottom, $\langle f, f\rangle$ represents the coordination of two feminine noun phrases. The two other combinations are ordered between these two nodes, $\langle f, m\rangle$, $\langle m, f\rangle$ (shown by the left construction in (8)). These two combinations are the same, so we remove one of them to arrive at a more simplified hierarchy (the middle hierarchy in (8)).

(8) *The gender hierarchy in French*



Elements of this algebra are then mapped into a hierarchy of plural forms which is basically the same as the gender hierarchy. As we can see these mappings are all monotonic because none of the them cross cut each other (9). Even if the hierarchy is flipped over and the FEM is above MSC, the mappings will remain monotonic but in the reverse order. This is partly why there are only 2 attested patterns out of the 8 possibilities for a 2-gender system.

(9) *Monotonic gender mappings in French*



Slovene

Slovene has three numbers and three genders. The predicate agreement forms are given below. In this table, *bil* is the past active participle of the verb 'be' (Corbett 1991: 280). The dual forms will result only if the two conjoined noun phrases are singular. The gender resolution works the same for both dual and plural conjunctions.

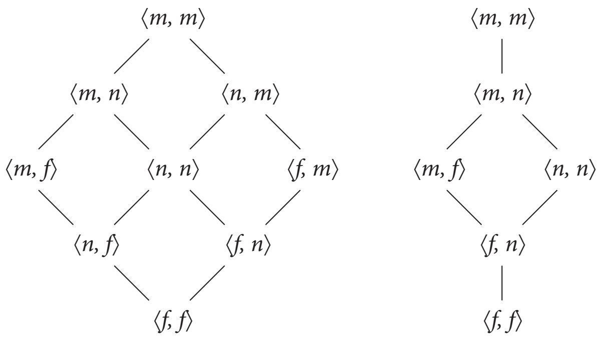
Table 2. Gender values in Slovene

<table>
<tr><th></th><th>SG</th><th>DL</th><th>PL</th><th></th></tr>
<tr><td>MSC</td><td>Ø</td><td>a</td><td>i</td><td>MSC</td></tr>
<tr><td>FEM</td><td>a</td><td rowspan="2">i</td><td>e</td><td>FEM</td></tr>
<tr><td>NEUT</td><td>o</td><td>a</td><td>NEUT</td></tr>
</table>

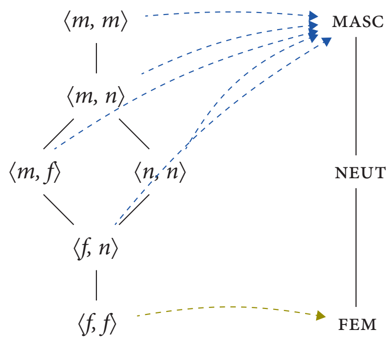
A masculine noun conjoined with a masculine will resolve in masculine. In the same way, a feminine noun conjoined with a feminine will resolve in feminine. All other combinations resolve in a masculine predicate.

In order to explore the hierarchical structure of Slovene, once again, we start from an underlying hierarchy of $f < n < m$ to construct a pointwise algebra and represent the gender combinations. In the simplified structure, all repeated nodes are removed.

(10) *The gender hierarchy in Slovene*



(11) *Monotonic mappings in Slovene*

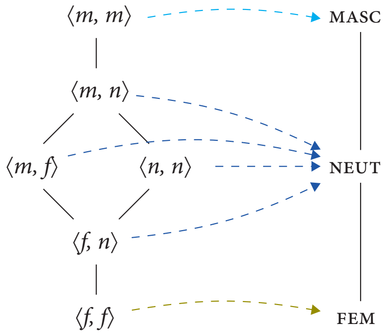


In a sense, the resolved agreement in Slovene (and similar languages like Serbo-Croatian) favors the masculine. Feminine is only used if all conjuncts are feminine, and the neuter is not used at all. Interestingly, we will have the same monotonic mappings if we flip over the structure along with the hierarchy. As long as the neuter is in the middle, all the mappings are indeed monotonic.

Icelandic

Icelandic has three genders: masculine, feminine and neuter. In coordination constructions, if the two conjuncts share the same gender value, the resolved form has that same gender in the plural. In resolution with all other conjoined combinations, neuter plural is used.

(12) *Monotonic mappings in Icelandic*



As was the case with Slovene, the positions of the masculine and feminine values can be reversed in this hierarchy. The important fact is to keep neuter in between those two values and all the mappings will be monotonic.

Latin

There are three genders in this language: masculine, feminine and neuter. Conjuncts of the same gender resolve in a form from the same gender. If conjuncts are of different genders, though, the criterion is purely semantic. Here the resolved form to be used depends on whether the nouns denote persons or not.

(13) *Resolution Rules in Latin*
a. Masculine is used if all conjuncts are masculine;
b. Feminine is used if all conjuncts are feminine;
c. Masculine is used if all conjuncts are human;
d. Otherwise, neuter is used.

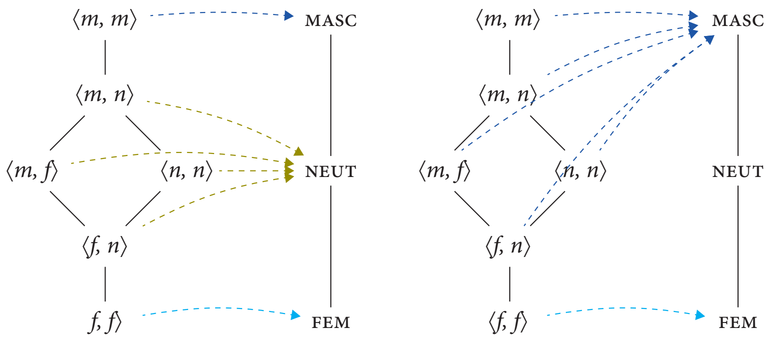
The rules are ordered in this way because the masculine and the feminine genders are not semantically restricted to humans. This means that a human feminine in conjunction with a human masculine resolves in masculine rather than the default neuter.

Table 3. Non-human and human resolution in Latin

Non-human	MSC	FEM	NEUT
MSC	M	N	N
FEM	N	F	N
NEUT	N	N	N

Human	MSC	FEM	NEUT
MSC	M	M	M
FEM	M	F	M
NEUT	M	M	M

(14) *Monotonic mappings in Latin non-human (left) and human (right)*



In order to show these mappings, we divide the rules into two sets of human and non-human rules. Within each sub-system, all the mappings are monotonic.

Tamil

Dravidian languages show examples of semantic resolution. Tamil has three genders: masculine (for nouns denoting male rationals), feminine (for nouns denoting female rationals) and neuter (for non-rationals). The resolved forms, however, result in two forms only: rational and neuter. [1]

Table 4. Gender values in Tamil

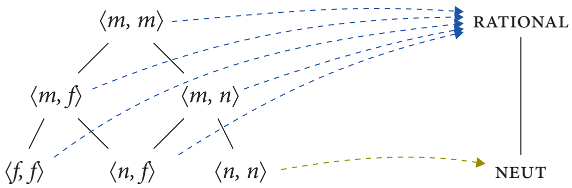
SG	PL
MSC	RATIONAL
FEM	
NEUT	NEUT

If, in a coordination structure, all conjuncts denote rationals, the rational form is used. If all conjuncts denote neuters, the neuter form should be used. The combination of a rational (feminine or masculine) with a neuter is generally avoided. But if ever allowed, the rational form is used.

Over a hierarchy that places rational (including masculine and feminine) over neuter, all the mappings from controller genders to target genders are monotonic.

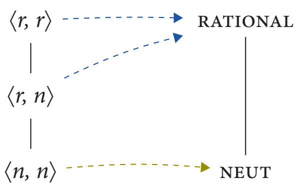
1. The resolution rules in Telugu, another Dravidian language, are the same as Tamil. This happens despite the fact that in Telugu, feminine and neuter are not distinguished in the singular.

(15) *Monotonic mappings in Tamil*



The resolution rules in Tamil are not based on formal gender values but rather follow the two semantic values, RATIONAL and NEUTER. This means that there are only two classes of nouns in the plural. Hence we can reconstruct a hierarchy that only includes those two values in a linear order. The mappings over this hierarchy are all still monotonic.

(16) *Monotonic semantic mappings in Tamil*



Archi

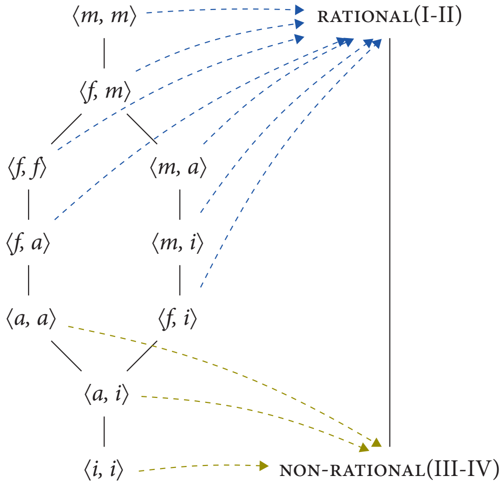
Caucasian languages are also famous for the semantic distinctions they make. Archi is a North-East Caucasian language with a rather intricate gender assignment system (17), but a semantically straightforward resolution system (18).

(17) *Archi gender system*
 I. male humans: God and other spiritual beings
 II. females
 III. most animals and some inanimate nouns
 IV. some animals and most inanimate nouns

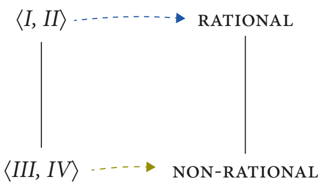
(18) *Resolution rules in Archi*
 a. I/II is used, if there is at least one rational conjunct (R);
 b. Otherwise, III/IV is used (IR).

In (19), *a* refers to animate and *i* refers to inanimate values. For an algebraic structure that includes all the gender values, if the rational gender (including masculine and feminine) resides higher on the hierarchy relative to the non-rational, then all the mappings are monotonic. Even though this account seems to work, gender and animacy are not the defining factors in Archi resolution. If we reduce the structure of conjoined noun phrases to rational and non-rational entities, then a simpler pattern emerges (20).

(19) *Monotonic mappings in Archi*



(20) *Monotonic semantic mappings in Archi*



Chibemba

Let us conclude this section with an example from Bantu languages. The gender resolution system in Chibemba is representative of all Bantu languages. It is a convergent system where the genders in the singular determine genders in the plural.

Table 5. Genders in Chibemba

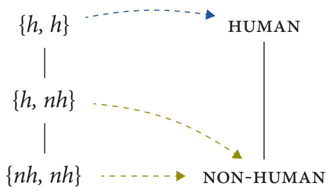
SG	PL
1	2
3	4
5	
14	6
15	
7	8
9	10
11	
12	13

Despite the complicated gender assignment classes in Chibemba, the resolution rules are rather simple. The system is rooted in the human–non-human distinction. Note that in Bantu, '1/2' gender means 'take class 1 agreements when singular and class 2 agreements when plural'.

(21) *Resolution rules in Chibemba*
   a. 1/2 is used if all conjuncts are human (1/2)
   b. 7/8 is used if none of the conjuncts are human (7/8)
   c. 7/8 is used if mixed, although the comitative construction is preferable (7/8).

The resolution rules in Chibemba are semantically motivated regardless of their noun class membership. To show this, we reduce the hierarchy to only include the human and non-human values. The resolved patterns over this hierarchy are monotonic.

(22) *Monotonic mappings in Chibemba*



4. Conclusion

I have shown that even though gender assignment systems in different languages greatly vary, the emerging gender hierarchies are substantially the same and directly motivated by typological data. The distinction made between the syntactic and semantic gender systems boils down to those gender values that are used in the resolved plural forms. In a syntactic system, e.g., French and Slovene, resolution rules are based on formal gender values. These systems mostly include feminine and masculine genders. As we saw, the nature of mappings remains the same as long as the feminine and masculine values reside on the two end nodes of the gender hierarchy. In a semantic system, e.g., Tamil and Archi, resolution rules are based on semantic values (RATIONAL VS NON-RATIONAL), which results in a condensed hierarchy of gender that only includes those two values. Regardless of the required gender hierarchy, in both system types, resolution rules follow monotonic mappings that associate base hierarchies to output forms. Similarly, Latin, as a mixed gender type, combines the two subsystems based on a semantic feature (the property of being human). The findings in this chapter lend further empirical support to the idea that monotonicity is a linguistic universal that extends beyond semantics.

References

Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139166119

Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.

Friðjónsson, Jón. G. 1991. Beygingarsamræmi með samsettu frumlagi. *Íslenskt Mál Og Almenn Málfræði* (1). 79–103.

Goldsmith, John A. 1976. *Autosegmental phonology*. London: MIT Press.

Graf, Thomas. 2019a. Monotonicity as an effective theory of morphosyntactic variation. *Journal of Language Modelling* 7(2). 3–47. https://doi.org/10.15398/jlm.v7i2.211

Hellinger, Marlis & Hadmumod Bussmann (eds). 2001. *Gender across languages: The linguistic representation of women and men*. Amsterdam: John Benjamins. (Volume 1, 2001; Volume 2, 2002; Volume 3, 2003) https://doi.org/10.1075/impact.9

Hellinger, Marlis & Heiko Motschenbacher (eds). 2015. *Gender across languages*. Volume 4. Amsterdam: John Benjamins. https://doi.org/10.1075/impact.36

Hockett, Charles F. 1958. *A course in modern linguistics*. New York: Macmillan. https://doi.org/10.1111/j.1467-1770.1958.tb00870.x

Icard, Thomas F. & Lawrence S. Moss. 2014. Recent progress on monotonicity. *Linguistic Issues in Language Technology* 9(7). 167–194.

Moradi, Sedigheh. 2020. Morphosyntactic patterns follow monotonic mappings. In Dun Deng, Fenrong Liu, Mingming Liu & Dag Westerstahl (eds.), *Monotonicity in logic and language*. Lecture Notes in Computer Science, Volume 12564, 147–165. Berlin: Springer. https://doi.org/10.1007/978-3-662-62843-0_8

Moradi, Sedigheh. 2019. *ABA generalizes to monotonicity. In Maggie Baird & Jonathan Pesetsky (eds), *Proceedings of NELS [North East Linguistic Society] 49*, Volume 2. Amherst, MA: GSLA.

PART II

Words, stems, and affixes

CHAPTER 4

Signs and words

Wendy Sandler
University of Haifa

Natural sign languages both constrain and expand our understanding of the nature of words. In each modality, words can be used in isolation, have meanings or grammatical functions, and are identifiable by their prosodic form. Such shared properties constrain the universal definition of the word. The differences, such as pervasive simultaneity and iconicity in sign word structure and word formation, compared to linearity and arbitrariness typical of spoken words, expand our conception of what a word can be. Investigation of an emerging sign language reveals just how fundamental the word is to language, and shows how morphological complexity begins as an abstract pattern. Sign languages introduce a tension between constraining and expanding the definitive properties of words, offering a more nuanced understanding of this most basic unit of linguistic structure.

Keywords: sign language, words, morphology, iconicity, simultaneity, linearity, language emergence, Israeli Sign Language, Al Sayyid Bedouin Sign Language

1. Introduction

Whenever the opportunity arises for deaf people to communicate regularly with one another, a sign language arises. Children who are exposed to sign languages from birth acquire them on a similar timetable to that of hearing children exposed to spoken language (Newport & Meier 1985), and, considering the complexity of the system to be acquired, the length of this timetable is remarkably short. Over the course of 60 years, researchers have convinced the scientific community that sign languages are fully fledged natural languages, with structure and architecture that are similar to those of spoken languages (e.g., Sandler & Lillo-Martin 2006; Pfau, Steinbach & Woll 2012). Yet there are also characteristic differences, attributed to the physical modality of transmission (e.g., Meier, Cormier & Quinto-Pozos 2002), as well as to the relative youth of sign languages (Aronoff, Meir & Sandler 2005). And the differences can be just as important as the similarities for revealing the nature of the word in language.

https://doi.org/10.1075/cilt.353.04san

This chapter begins by demonstrating ways in which words in sign language are similar to words in spoken languages, in § 2. As in spoken languages, the words of sign languages are minimally distinguished by formational primitives and typically conform to an optimal prosodic form; they are subject to morpheme structure constraints; and they can serve as stems for linear, derivational affixation. As in spoken languages, compounding is very productive in sign languages.

Section 3 turns to some of the main differences between words of spoken and signed languages. The building blocks of signs, in some ways comparable to phonological features and categories, are often meaning-bearing; nonlinear inflectional processes motivated by iconicity are common; and morpheme-gesture hybrids (so-called classifier constructions) challenge the overall architecture of linguistic structure, exemplified here by a theatrical monologue. All of these properties expand our notions about the nature of the word, and reveal the profound influence of the transmission system on linguistic structure (spoken or signed).

Sign languages provide novel evidence for the intuition that the word is the most fundamental and essential element in any language. Our work (with Mark Aronoff, Irit Meir, and Carol Padden) on a newly emerging Bedouin sign language in Israel indicates that language begins with the word – before any other structural units emerge. This research, excerpted in § 4, also suggests that other levels of structure that are considered definitive in language – phonology and syntactic complexity – emerge later, and take longer to become systematic, than the word. Evidence from Al-Sayyid Bedouin Sign Language in § 5 suggests further that morphological complexity emerges gradually, observable first as an abstract pattern within the word, rather than as a list of forms and rules. Section 6 is a summary and conclusion.

2. Sign words are similar to spoken words

It took a while for sign languages to be taken seriously by linguists. One reason for this is the apparently transparent link between the form of some signs and their meanings. Here I use the term 'sign' to correspond to the term 'word' (Sandler & Lillo-Martin 2006; Meir 2012). The misconception that sign languages are not real languages and are merely a collection of holistic iconic gestures was shattered by Stokoe's (1960) systematic analysis. Stokoe demonstrated that there is a finite list of formational units – belonging to the categories of handshape, location, and movement – that make up signs in American Sign Language (ASL), and that, if one of the units is substituted for another, this can result in a different sign with a different meaning. By exemplifying minimal pairs, Stokoe showed that sign languages have duality of patterning – a basic design feature of human language (Hockett

1960) – and this changed everything. From there, the field of sign language phonology developed and expanded (see Sandler & Lillo-Martin 2006, Unit III; Sandler 2017; Brentari 2019). This discovery also opened the door to linguistic investigation of all other levels of language, including the word and its internal structure. Figure 1 exemplifies minimal pairs in Israeli Sign Language (ISL), distinguished only by features of handshape, of location, or of movement.

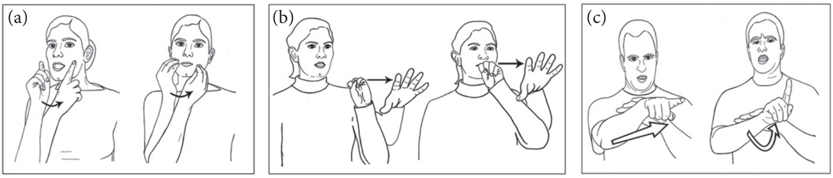


Figure 1. Minimal pairs in ISL. (a) MOTHER and NOON, distinguished by handshape, (b) SEND and TATTLE, distinguished by location, and (c) ESCAPE and BETRAY, distinguished by movement (© 2021 The Author)

The signs MOTHER and NOON in (1a) are both signed on the [head] as the major body area or location, articulating different 'settings' at that location sequentially: informally, at either side of the mouth (Sandler 1989). The contrast in (1a) is in the different handshapes for each sign: index finger extended for MOTHER, vs. all fingers curved and touching the thumb for NOON. In (1b), two different body areas create the minimal contrast in the two signs: [torso] for SEND [head] for TATTLE. In (1c), the body area, the [nondominant hand], and the index-extended handshapes are the same in the pair. The contrast is in the different movement path shapes – [straight] for ESCAPE and [arc] for BETRAY.

2.1 Constraints

In any language, linguistic elements cannot be combined randomly. To take just one example, in Semitic languages, a canonical triconsonantal root can end with two successive identical consonants (separated by vowels), but cannot begin that way (McCarthy 1981), thus Hebrew *Xaʃaʃ* "fear" but *XaXaʃ. Sign languages are also constrained in the ways in which linguistic elements can combine. In sign languages, the hand must be configured by selecting a single group of fingers (Mandel 1981) to characterize the whole monomorphemic sign, although the fingers can change their position, e.g., from closed to open, as in SEND and TATTLE (1b).[1]

1. This constraint does not hold for signs borrowed from fingerspelling a word with letters from the alphabet.

Similarly, well-formed monomorphemic signs can only be signed on or near a single major body area or location (Battison 1978), and the hand typically moves from one finer setting to another with respect to that location, like all of the examples in (1) (see Sandler 1989 for motivation of the categories Location and Setting). Articulation with two different groups of fingers, two distinct successive movements (barring reduplication), or on two different body areas generally signals two different morphemes (Sandler 2012b).

The universal, canonical form of a sign is typically monosyllabic (Coulter 1982; Sandler 1989, 1999; Brentari 1990, 1998; Wilbur 1993), with the hand moving in a path from one location setting to another, i.e., Location-Movement-Location or L-M-L (Sandler 1989, 1999; cf. Liddell & Johnson 1989). This means that a canonical sign has only one movement (as in (1a) or (1c)) or two movements simultaneously (as in (1b), which consists of a path movement from one location to another, together with an opening movement of the fingers). Though morphemes, words, and syllables are often coextensive, they are not isomorphic (Sandler 2008).

A high proportion of the signs in any sign language lexicon are produced with two hands, and the interaction of the hands is strictly constrained within a sign (see Crasborn 2011 for an overview). Battison's (1978) symmetry and dominance conditions on two-handed signs have withstood the tests of time and of generality to other sign languages. Broadly interpreted, the dominance condition indicates that if the nondominant hand is static, it serves as a location for the dominant hand, and is configured in a neutral shape, as in Figure (1c) above. If the nondominant hand moves, the symmetry condition indicates that it must move symmetrically with the dominant hand, as in SHOP, shown in Figure (2a) below.

If the modality-specific property of having two identical articulators, the two hands, were exploited in the contrastive feature inventory within the morpho-phonology, this would make sign words very different from spoken words. However, even though about half of the signs in a sign lexicon are two handed, the constraints on two-handed signs mean that only one hand is formally distinctive in a sign (Sandler 1993). Figure (2b) shows a non-occurring two-handed sign, in which both hands move but are not configured in the same shape, violating the symmetry constraint.

(a) *(b)

Figure 2. The ISL sign (a) SHOP conforms to the symmetry constraint on signs in which both hands move (Battison 1978), while (b) violates the constraint because the hands have different shapes (Meir & Sandler 2008) (© 2021 The Author)

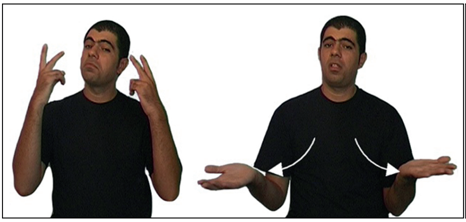
2.2 Linear morphological processes

Morphologically complex forms in spoken languages typically involve linear affixation, either derivational, like *teach+er* in English, or inflectional, like *teach+es*. Sign languages exhibit some linear affixation as well. An example is a productive agentive suffix in ASL that originated as a sign for PERSON, deriving, e.g, TEACH-ER (Figure 3a) from TEACH. ISL also has some derivational affixes, including a negative suffix meaning "not at all". The Examples (3b) and (3c) show that this form is subject to allomorphy: whether the suffix is one- or two-handed depends on the form of the stem (Meir & Sandler 2008).

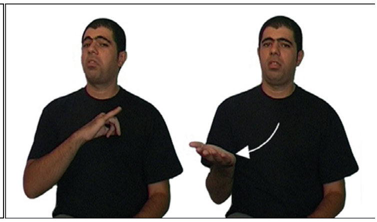
But this sort of linear or concatenative morphology is very limited in sign languages, as is derivation in general, and we attribute the dearth of derivational processes to the relative youth of sign languages, none of which is more than about 300 years old (Aronoff et al. 2005).

One concatenative process found in all sign languages that have been studied is compounding. As in spoken languages, this process is very productive. And also as in spoken languages, compounds can become lexicalized, like the English (exocentric) compound *soap opera,* often resulting in a less transparent semantic contribution by each of its members (Klima & Bellugi 1979; Sandler 1987; Sandler & Lillo-Martin 2006). For example, the lexicalized ASL compound BELIEVE is comprised of the signs THINK and MARRY. The lexicalized form involves regressive assimilation of the handshape and deletion of locations in both members (Liddell & Johnson 1986; Sandler 1987).

(a)



(b)



(c)

Figure 3. (a) An agentive suffix in ASL – TEACH+PERSON: TEACHER; (b) a negative suffix in ISL (with allomorphy) IMPORTANT-NOT-AT-ALL with the two-handed allomorph of the negative suffix; and (c) INTERESTING-NOT-AT-ALL, with the one-handed allomorph of the negative suffix[2] (Meir & Sandler 2008) (© 2021 The Author)

Such forms illuminate properties shared with spoken languages that are sometimes taken for granted, such as the crucial role of concatenation in word-internal phonological alternations. Since sign language words typically have very little linear structure, assimilation is very rare within the word. However, when words are systematically concatenated in compounds, we do find assimilations, in both ISL and ASL. Apparently, these compounds tolerate assimilations and deletions which can mask the forms and meanings of their individual members because, as lexicalized words like the lexicalized compound *pancake* ([pæŋke^{j}k]), their meanings as complex words are conventionalized. The meaning of the first member, [pæn], is not compositionally interpreted, and assimilation of the nasal [n] to the velar [k] is tolerated. The assimilation process in lexicalized compounds in ASL and ISL (Sandler 2012b) contributes to the optimal monosyllabic form of a phonological word (Sandler 1999).

2. Articulation at or near two different major body areas ([head] and [torso] in each case) shows that the forms in (3a and b as well as in 4 below) are bimorphemic.

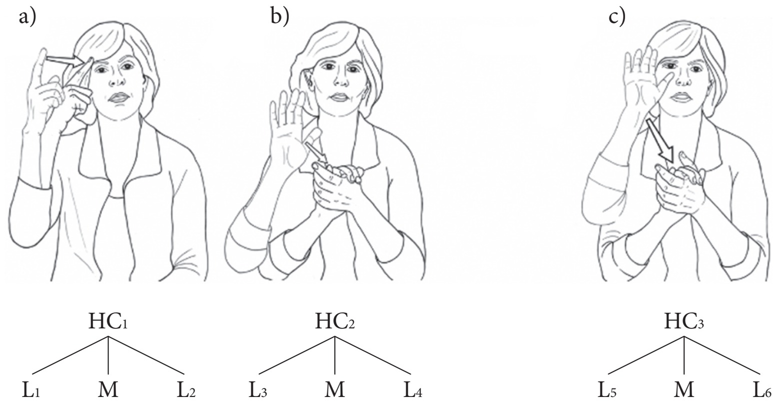


Figure 4. A lexicalized compound in ASL: THINK+MARRY=BELIEVE. The schematic representations show that the second hand configuration (HC_2) assimilates regressively to the beginning of the compound, and that the first locations of both members are deleted, yielding a morphologically complex word that is monosyllabic (© 2021 The Author)

3. Sign words are different from spoken words

From the beginning, sign language researchers were also aware of some of the striking differences between the words of spoken and signed languages. But when the field was new, the compelling goal of research was to probe sign languages for similarities with spoken language, using similar tools of analysis, and to convince the rest of the scientific community that sign languages are real languages in every sense. By and large, we succeeded in this mission, constraining the notion of the word as a universal theoretical construct. But we won't have a comprehensive picture of the human language capacity until we study not only similarities, but also differences in words and word structure between the two natural language modalities. In this way we expand our conception of what a word is.

3.1 Meaningful meaningless parts

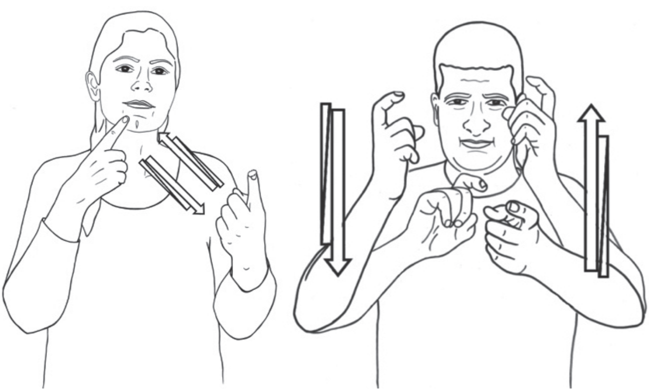
The building blocks of words (phoneme-like units and features) make distinctions and conform to alternations according to form and not meaning, and therefore comprise a phonological system.[3] However, meaning can often be attributed to

3. On sign language phonology, see for example Liddell & Johnson (1989), Sandler (1989), Brentari (1998), Sandler & Lillo-Martin (2006) Unit III, Sandler (2017), Brentari (2019).

these elements through iconicity, as has been noted by researchers of many different individual sign languages (see, e.g, Brennan 1990; Johnston & Schembri 1999 and references there; Taub 2001; Fernald & Napoli 2000).

Let's take the two hands as an example. Section 2 above describes some structural constraints on two-handed signs, based on form alone. But the presence and interaction of two hands in a sign can also be motivated iconically by meaning. In a study of three unrelated sign languages, twice as many concepts were found to be lexically specified for two hands in all three sign languages than would be predicted if the distribution were random (Lepic et al. 2016). The semantic motivation usually involves plurality, broadly conceived. The authors found similar patterns in a young village sign language in the south of Israel, Al-Sayyid Bedouin Sign Language (ABSL, see § 4). For example, the signs that mean NEGOTIATE in ISL and ABSL, two unrelated sign languages, are both two-handed, motivated iconically by the interaction of two sides to a negotiation (Figure 5). The handshapes are different in the two sign languages, but they are also motivated. The handshape in the ISL sign often corresponds to a line of communication, while the handshape in the ABSL sign often corresponds to a person in that language. The repeated, alternating movement (first one hand and then the other, repeatedly), is similar in the two languages, motivated by the give-and-take nature of negotiation.

Despite meaningful internal units, two-handed signs such as these are usually considered monomorphemic, for two reasons. First, the internal units of hand configuration, location and movement are all bound, and, though meaningful, they do not attach to a stem but rather combine with each other to conform to a



(a) (b)

Figure 5. Sublexical iconicity: NEGOTIATE (a) in Israeli Sign Language (ISL) and (b) in Al-Sayyid Bedouin Sign Language (ABSL) (Lepic et al. 2016) (© 2021 The Author)

canonical sign shape; and second, they cannot independently undergo morphological processes. The whole sign, NEGOTIATE, is a verb, and, as such, can take verbal inflections such as temporal aspect (see § 3.2).[4] This means that sign language words are characterized by a kind of 'dual-duality of patterning', in which their sublexical building blocks function both as meaningless and as meaningful elements.

3.2 Iconically motivated, simultaneous morphology

A good deal of the complex morphology of any sign language is iconically motivated semantically, and simultaneous or nonconcatenative phonologically. I will give two examples here: verb agreement and classifier constructions. Both types, while linguistic in commonly understood ways, also have explicitly gestural properties, at once unifying them with and distinguishing them from spoken words.

3.2.1 *Verb agreement*

A particular type of inflectional process, often called verb agreement (Padden 2016; Meir 2002; Lillo-Martin & Meier 2011), is found in many unrelated sign languages. This topic has received different treatments in the literature (see Mathur & Rathmann 2012 for a comprehensive overview), and the examples here are partial and descriptive rather than theoretical. In the agreeing verb system, pointing signs first typically establish particular spatial loci for referents corresponding to subject and object, and the hand, configured for the verb, moves from subject locus to object locus.[5]

While agreement is a very general phenomenon in spoken languages, sign language agreement is different. First of all, while spoken language agreement rules apply across the board to verbs, in sign languages, different verb types take different

4. As with all differences between words in the two modalities, the difference is real but it is one of degree. Spoken languages have a good deal more iconicity within words than once believed, particularly in ideophonic systems (e.g., Dingemanse et al. 2015; Dingemanse & Akima 2017). Nevertheless, unlike the motivated sublexical units in sign languages, ideophones are far from universal among spoken languages, and the amount and type of iconicity in such systems vary a good deal.

5. In so-called 'backwards verbs' like TAKE and INVITE, the hand moves from the object to the subject instead of subject to object. Meir unifies the analysis by showing that in both regular and backwards verbs, the loci correspond to semantic source and goal rather than syntactic subject and object, so that for both agreeing verb types, the hand moves from source to goal. Although direction of movement of the whole hand is determined by the semantic categories source and goal, the syntactic category of 'object' is still relevant for this system, since the direction in which the hand faces is determined by the object (Meir 1998, 2002)

kinds of agreement (Kwok, Berk & Lillo-Martin 2020). Meir (2002) argues that this subset should be characterized as verbs that literally or metaphorically involve transfer, such as GIVE, SEND, SHOW, or TEACH, INFORM. Second, these referential loci are set up in gestural space. Referential loci for present referents are established with respect to the location of the referent, and agreement with non-present referents is governed by arbitrary spatial locus points (Lillo-Martin & Klima 1990; Lillo-Martin & Meier 2011). Some agreement forms for the ISL verb SHOW are illustrated in Figure (6). Figure (7) shows a reciprocal form of the verb LOOK, extracted from the story of Genesis in the sign language play, *It's Not About Ebisu* (see § 3.2.3). Here, in the story of Genesis, the narrator tells us that Adam and Eve (newly created from Adam's rib) look at each other (approvingly).

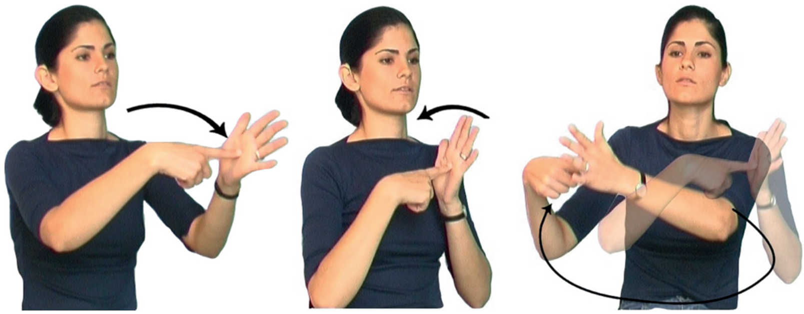


Figure 6. (a) Examples of verb agreement for the ISL verb SHOW (© 2021 The Author)

Figure 7. The ISL agreeing verb LOOK (reciprocal form):
"Adam and Eve look at each other (approvingly)" (© 2021 The Author)

The agreement morphemes are not sequential prefixes or suffixes, but rather they are points in space representing the subject and object, substituted for the first and second location of the citation form of the verb, still conforming to the canonical LML template. The system is somewhat reminiscent of nonconcatenative morphology found for example in Semitic languages (McCarthy 1981; Sandler 1990). But here again the difference is more impressive than the similarity. First, most spoken languages, including Semitic languages, have abundant concatenative morphology, while sign languages are very limited in this regard. Instead, much inflectional morphology such as verb agreement, and temporal aspect inflection described in § 3.2. below, substitute some of the features of the lexical locations and movements, while maintaining the optimal monosyllabic form of the word (Sandler 1999). Second, the agreement elements are points in space rather than concrete and listable morphological entities, and in these ways might be considered gestural rather than linguistic (Liddell 2003).

However, despite universal access to gesture for human communicators, very young sign languages, such as Al-Sayyid Bedouin Sign Language, and early Israeli Sign Language, do not have systematic verb agreement systems (Padden et al. 2010). This leads us to conclude that the agreement system is indeed attributable to linguistic conventionalization, albeit motivated by gestural tendencies (see §3 and §4). The forms agree systematically for linguistic categories of person and number, and, in the case of some East Asian sign languages, for gender (Fischer & Gong 2010). Meir (2002) discovered that the path direction and the direction in which the hands face play independent roles.[6] The direction faced by the hands in ISL agreement is determined by the syntactic linguistic category, 'object' (Meir 2002). So, we may safely consider them to be part of a linguistic system, though one that is different from any in spoken languages (Aronoff et al. 2005; Lillo-Martin & Meier 2011).

3.2.2 *Temporal aspect inflections*

In their seminal volume, *The Signs of Language*, Klima & Bellugi (1979) identified a large number of inflections on verbs and adjectives in ASL that indicate temporal aspects such as Continuative, Iterative, and Habitual. These inflections are produced by changing the rhythmic pattern and path shape of the lexical signs, and usually by reduplication. The resulting forms are iconically motivated, but conventionalized. We have found temporal aspect inflection in ISL as well. Figure (8a) shows the citation form of LEARN in ISL, (b) the durational form, and (c) the continuative form.

---

6. In Meir's analysis, agreeing verbs agree in their path movement with the semantic categories of source and goal, rather than the syntactic categories of subject and object. Facing of the hands, however, is syntactically governed.

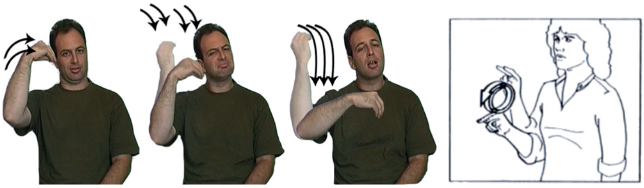


Figure 8. Simultaneous aspectual inflection: (a) the citation form of LEARN in ISL (b) LEARN (Durational) and (c) LEARN (Continuative); (d) shows the Continuative form of the sign LOOK-AT in ASL, following Klima & Bellugi (1979) (© 2021 The Author)

We see the effects of conventionalization within this iconically motivated system by comparing across sign languages. While iterative, durational, continuative, and habitual events all involve reduplication in both ISL and ASL,[7] other features of the template differ from language to language. The citation forms of verbs in both languages typically correspond to the canonical sign language monosyllabic template LML (Location, Movement, Location), each with a default straight path movement. Continuative aspect forms in both languages change the movement of the base sign – but in different ways. The continuative template for ASL imposes a circular movement on the base (see Figure (8d), following Klima & Bellugi 1979), while in ISL, the continuative form lengthens the movement segment of the sign. Both languages impose reduplication for the continuative aspect inflection. The gist is this: iconic and gestural motivation exist in such forms; but function and systematicity define them as linguistic.

3.3 Classifier constructions: Hybrids of lexical and gestural elements

An intriguing phenomenon found in a large number of unrelated established sign languages is the classifier construction (Supalla 1986; Emmorey 2003; Zwitserlood 2012). The literature on sign language classifier constructions contains a range of analyses and taxonomies, and the description here is necessarily simplified. But on one issue, there is consensus in most current treatments: sign language classifier constructions are unlike any structures found in spoken languages (Schembri 2003). This is essentially because classifier constructions are part lexical/morphological and part gestural.

Specifically, handshapes from a finite list constitute lexical morphemes that classify entities according to various properties such as (a) semantic category (e.g., people, vehicles, small animals), (b) size and shape (e.g., flat, round, cylindrical), and (c) how they are handled (e.g., full hand, index and thumb). This finite list of forms representing specific classes of referents implies a morphological system.

But, unlike signed words, which consist of lexically specified locations and movements in addition to handshapes, the only lexically specified part of classifier constructions is the classifier handshape. These shapes combine freely with different locations and movements, and sequences of theses, which are not lexically specified for any classifier. Instead, the locations and movements are analogical to real world motion and location relations, and they are gradient, rather than discrete. That is, the motions and locations occurring with classifier handshapes have properties usually attributed to gesture rather than to language (McNeill 1992; see also Müller 2018).

However, classifier constructions are complex, and they are only fully mastered by deaf children when they reach their teens (Zwitserlood 2012). This suggests that there are aspects of the system that behave in a constrained and conventionalized way, and are thus not purely gestural. But they are not words either, by any familiar definition, for the following reasons: (a) the classifier itself is a bound morpheme that combines with gestural elements; (b) the two hands can represent two independent classifiers in an expression, violating the symmetry and dominance constraints (Aronoff et al. 2003); and (c) a classifier handshape can even persist in the signal without marking temporally signaled prosodic phrase boundaries (Nespor & Sandler 1999) in expressions that correspond to a sequence of propositions (Sandler & Lillo-Martin 2006). It is fair to say that classifier constructions are not fully understood, and yet, they can't be ignored.

3.4 Hybrid forms in creative performance: The Ebisu Sign Language Theatre Laboratory

The hybrid classifier forms are just as ubiquitous in established sign languages as they are challenging for linguistic analysis. Part lexical and part gestural, they are especially common in creative depiction, as in story-telling, poetry, and theatre. Figure 8d above shows a reciprocal form of the sign LOOK-AT, taken from the play *It's Not About Ebisu*, performed by the Ebisu Sign Language Theatre Laboratory.[8] In the group's two productions to date, which combine sign language with physical theatre, the actors make novel, creative use of meaning-structure combinations that lurk beneath the surface in sign languages.

It's Not About Ebisu begins with a narration about a Japanese Shinto god who is deaf, and whose name the group adopted. The narrator skillfully interweaves sign words with classifier constructions and with gestural elements, often produced by

8. The Ebisu Sign Language Theatre Laboratory was created as part of a larger project, called The Grammar of the Body, funded by European Research Council grant 340140. http://gramby.haifa.ac.il/

the hands, or with pantomime or 'constructed action', (Lillo-Martin 2012; Cormier, Smith & Zwets 2013), which can involve the whole body. Both pantomime and gesture are idiosyncratic and can be gradient, while linguistic elements are typically rule governed and discrete. Here I show a sequence from a monologue involving signs, constructed action, and classifier constructions, to highlight both the versatility of sign language, and its dissimilarity from spoken language.

The narrator, actor Golan Zino, explains that a male and female god lived on one of the Japanese islands, and then signs "The woman god became pregnant" with the lexical signs WOMAN, PREGNANT (Figure 9a). She gives birth to a baby god who has no bones, conveyed with pantomime (9b), in which the actor's whole body represents the body of the boneless baby god, Ebisu. The baby is not only boneless, but also deaf. Traumatized, the mother throws the baby into the sea in a basket, which is tossed by the waves. A bit later in the story, a fisherman spots the basket. The narrator uses the small-round-object classifier to represent the eyes of the fisherman watching the baby in the basket. The narrator moves the classifiers along a path, to show that the eyes of the fisherman follow the course of the basket in the water (9c). The signer might have used the more prosaic lexical sign LOOK-AT (a one-handed version of Figure 7) to convey this, but his choice to use a classifier construction is more expressive, incorporating in one image the whole event of visually following the path of the basket, holding the baby Ebisu, in the water, and following the path with his own eyes.[9]

(a) (b) (c)

Figure 9. (a) lexical signs: WOMAN, PREGNANT (b) constructed action: "boneless baby" (c) classifier construction "SMALL-ROUND-OBJECTS-MOVE-IN-ARC": "the fisherman watched (eyed) the basket floating by." (© 2021 The Author)

9. In the Ebisu story, the baby is nursed to health by the fisherman's wife, growing bones and becoming robust and good natured, but remaining deaf. Ebisu is the Shinto god of fishermen and prosperity.

4. Language begins with the word

Al-Sayyid Bedouin Sign Language was born with four deaf siblings in a single household about 75–100 years ago. Today, about 130 deaf people and many hearing people in this village of about 4,000 use ABSL. Over the course of about twelve years, our team – Mark Aronoff, Irit Meir, Carol Padden, and I – visited the Al-Sayyid Bedouin village in the Negev of Israel many times, to learn about a sign language that was emerging de novo in this insular community with a high incidence of deafness (Sandler et al. 2005). We naturally fell into a work pattern in which each of us pursued what we were most interested in and best at, and the balance was perfect. Our perfect balance was cruelly dismantled when our close friend and colleague, Irit, passed away in February 2018. The Al Sayyid experience informed all four of us about the nature of human language, and changed preconceptions that we didn't know we harbored. We continue to learn from this language, in no small part because of insights and analyses by and with Irit.[10]

What have we learned about the word and its internal structure from Al-Sayyid Bedouin Sign Language? We've learned that the word is the most basic unit of communication in a new language, and that it can be distinguished from gesture and pantomime at the outset. We've also learned that a community can tolerate a good deal of variation, both in establishing word-concept relations, and in the internal structure of words for the same concept. While some of the structures described in § 3, such as verb agreement and classifier constructions, were thought to be universal in sign languages, we found neither of these structures in ABSL. Since verb agreement was not found in early stages of ISL either, but eventually developed (Padden et al. 2010), we surmise that this inflection might develop in any sign language, given time. A gradual process of emergence of spatial grammar is also supported by work on Nicaraguan Sign Language (Senghas & Coppola 2001), which arose in a school for deaf children beginning in the late 1970s. But we also understand from the fact that that these systems do not arise immediately in a sign language that they are not merely 'natural' gestures. They draw from a pool of gestural-spatial capabilities, and they conventionalize and systematize given time and interaction.

10. Irit Meir's immeasurable contributions to our research project continue to inspire us. Her original thinking, her loving nature, and her generosity of spirit are profoundly missed.

4.1 Words at the outset

We were fortunate to receive a video of one of the first four deaf siblings born in the Al-Sayyid village, the originators of ABSL. In the video (from 1994), the man, now deceased, was about 65 years old. He was sitting in a tent and telling a story from the history of the Al-Sayyid people about a feud with another Bedouin group. The story was apparently passed on to him by his hearing father, in the home sign system that the family developed. The tale is quite complex, but the linguistic structure is very spare. Most of the man's propositions consist of one or two words, interspersed with some mime-like elements. There are very few utterances that could be compared to complex sentences, no complex morphology, and no prosodic/intonational organizing features found in later stages of the language (see overviews in Aronoff et al. 2008; Sandler 2012; Sandler et al. 2014). Our informal inquiries suggested that an addressee would have to know the story in advance in order to fully understand the narrative. Figure 10 shows a segment of the narrative, with the gloss in the left column and the translation in the right column. Hyphenated forms orthographically represent a single sign glossed as two English words. Words in parentheses in the translation are not in the narrative, but are filled in from context and shared knowledge by the translator, the man's son.

Gloss	Translation
TAKE-OFF	"*(The man from Al-Sayyid)* took off at a gallop.
RUN	Sword and gun *(were at the ready)*.
RUN	*(Someone)* struck with a sword.
SWORD	*(He)* blocked with his gun.
GUN	*(He got)* hit.
STRIKE	*(He)* fired, fired.
GUN **BLOCK**	The horse fell down.
HIT	An eye fell out.
SHOOT	*(He)* waved a cloth *(for reinforcement)*."
SHOOT	
HORSE FALL	
EYE FALL-OUT	
CLOTH WAVE	

Figure 10. Segment of ABSL first generation signer's narrative (Sandler 2012a) (© 2021 The Author)

Many of the sign words glossed in Figure 10 have persisted in the language, such as HORSE, HIT, and RUN. Only the glosses in bold have the pantomime-like form of constructed action; all the other forms are performed by the hands, and fulfill the prosodic conditions of sign words. The difference between whole-body pantomime/constructed action and a manual sign/word and in the story is illustrated in Figure 11.

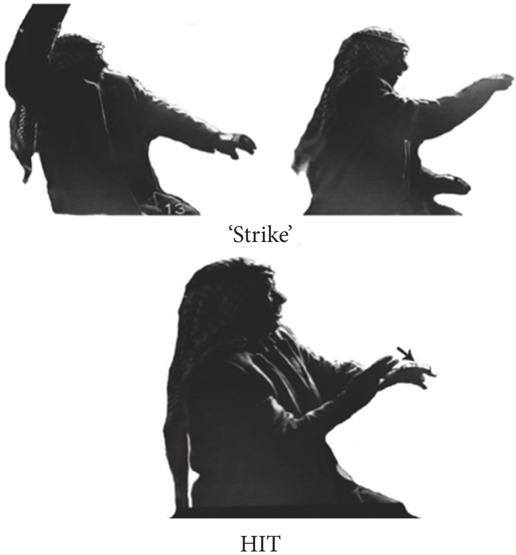


Figure 11. Constructed action vs sign words in a first generation ABSL signer: (a) the constructed action form, “strike”; (b) the sign word, HIT (Sandler 2012a) (© 2021 The Author)

4.2 Lexical variation in the community

We expected there to be broad consensus on lexical labels in this small community, at least for familiar objects. But one of the many surprises we encountered in ABSL is variation in lexical labels across the community. While some forms are certainly conventionalized across the community, we also found variation for everyday items, such as KETTLE or CAT, variants of the latter pictured in Figure (12). Monomorphemic signs conform to the monosyllabic template found in other sign languages canonically: LML (location-movement-location), with a single handshape.

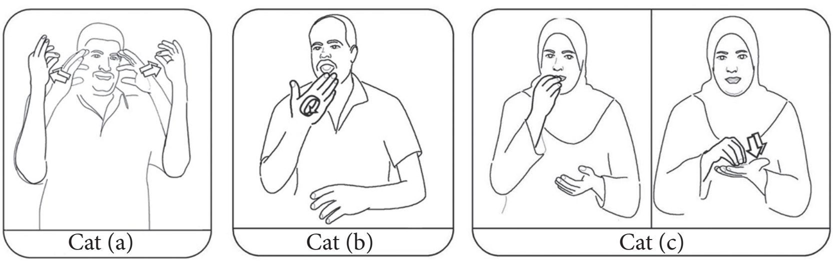


Figure 12. Three different ABSL signs for CAT, each representing a different concept related to cats – *whiskers*, *licking the paw*, and *dainty footprints* (© 2021 The Author)

Our investigation also uncovered a good deal of variation, not only in the different concepts chosen to represent a sign word, as shown for CAT in Figure 12, but also in signs that share the same concept. A clear example is the sign DOG. Ten exemplars of the sign from different signers all represented the barking mouth of the dog as a conventionalized concept, but the sublexical features used to convey this concept varied widely in all parameters, handshape, location, and movement (Israel & Sandler 2011; Sandler et al. 2011).[11] We concluded that a phonological system had not yet crystallized in the village.

5. The birth of abstract productive morphology

Returning to the word and its internal structure, we were surprised that we did not find word formation processes so typical of established sign languages, processes that are iconically motivated and productive, like verb agreement and classifier constructions described in § 3, though we tried to elicit them. However, we did stumble upon a morphological process that we had not encountered in other sign languages, one that defies all expectations. It is iconic but linear rather than simultaneous, and it is productive – but idiosyncratic. It is an impressive example of morphology as an abstract system, morphology 'by itself' (Aronoff 1994).

In a picture-naming task, we found a type of suffixation of a morpheme that specifies size and shape. The suffix uses a small set of generic handshapes to represent long and thin, wide, or flat, reminiscent of a type of classifier handshapes found in other sign languages, yet different in revealing ways. First of all, these elements are sequentially suffixed to nominal forms, and not combined simultaneously with locations and movements. Second, both hands are typically involved: after signing the nominal base, one hand manifests the suffix, and the other hand points to the wrist of the suffixal hand, as shown in Figure 13. The handshape assimilation shown in the closeups in Figure 13 is discussed below.

Many sign languages make use of Size and Shape Specifiers (SASSes) in a variety of ways. However, the SASSes in ABSL have characteristic and systematic qualities. In a comparative study of novel SASS forms in ABSL and ISL, Tkachman & Meir (2018) showed that ABSL forms with SASSes are much more likely to be SASS-final than are ISL forms, and that ABSL forms convey both size and shape, while ISL forms

11. Other sign languages select different concepts for DOG, such as scampering legs and paws in ISL, and a finger snap in ASL. This shows that there is a conventionalized relation between the concept and the sign in ABSL, namely, the barking mouth of a dog, though the 'pronunciation' varies widely across the community.

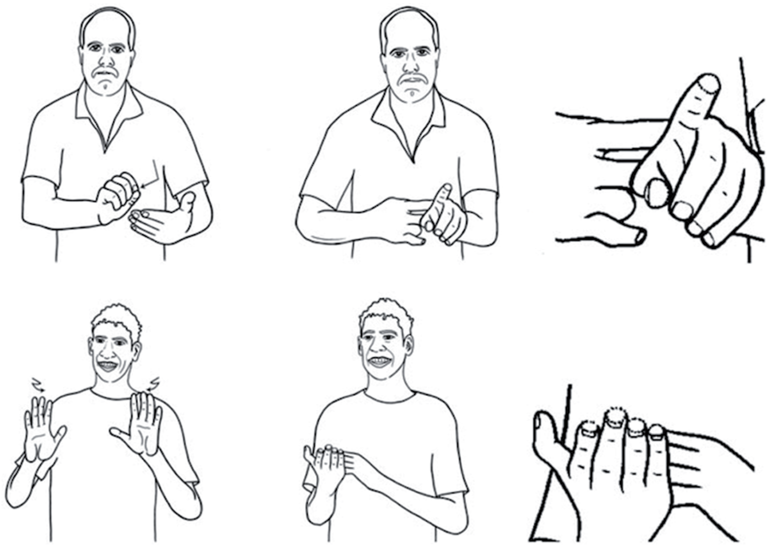


Figure 13. Size and shape specifier suffixes in ABSL, with handshape assimilation: (a) WRITE+ THIN-OBJECT = PEN; (b) TELEVISION+FLAT-OBJECT = REMOTE-CONTROL (Handshape assimilation of the pointing hand to the specifier hand is seen in the closeups.) (© 2021 The Author)

convey only shape. This study confirms our earlier findings that SASS suffixation is a robust, productive, and language-particular pattern in ABSL.

Normally, we think of productive morphology in terms of conventionalized morphemes that combine with a certain class of roots to derive complex forms with predictable meanings, like the NOT-AT-ALL suffix shown in Figure 3c. But, while it is possible that the size and shape suffixes like those shown in Figure 13 are themselves conventionalized as morphemes, they do not necessarily combine with particular words across the Al-Sayyid community. The same object can be labeled with a different base sign and without such a specifier. Nevertheless, the **device** of suffixing a size and shape specifier is widely used.

We found suffixes of this kind modifying various objects, in signers who did not belong to the same family or social group. While word combinations have variable order in elicited compounds of ABSL, the size and shape specifier is a bound morpheme, and nearly always final (Meir et al. 2010; Tkachman & Meir 2018). This suggests that the process of suffixing a size and shape specifier to an object noun is conventionalized, but that the nouns that take the suffix, and the particular size and shape specifier selected, are not conventionalized.

Such variation in this morphological process is found for example in the sign for KETTLE, an everyday household object. In one family, it is DRINK+STOUT-OBJECT, where STOUT-OBJECT is a two-handed size and shape specifier indicating the dimensions of a kettle. Another signer in a different family labeled the same picture of a kettle with a different sign for the object and a different size and shape specifier morpheme: CUP+UPRIGHT-OBJECT.

We've seen in § 2.2 and Figure 4 that assimilation can characterize lexicalized compounds in established sign languages without sacrificing the meaning of the members, since the whole compound word has a conventionalized meaning. On closer inspection of the ABSL size and shape suffixes, we also noticed assimilation. We saw that the shape of the hand that points to the specifier (as in Figures 13a,b) assimilates the handshape of the specifier on the other hand (Sandler et al. 2011). The assimilation attendant with these suffixes is an important indicator of conventionalization, but the question is, what exactly is conventionalized?

Unlike members of conventionalized lexicalized compounds, the complex word formed with the specifier suffix is not lexicalized, that is, not conventionalized. Instead, the shape of the pointing hand can assimilate because its **position,** after a concrete noun, and its deictic **function** are conventionalized. These ABSL forms show us that a word formation process can be conventionalized as a productive strategy, recruited to describe an object, on the fly. This provides a novel kind of evidence from language emergence for the notion of morpheme put forth by Aronoff (1976, 1994), as an abstract element within words, identifiable by its distribution, and its form.

6. Summary and conclusion: What's in a word?

One common way of looking at language is exclusively as a formal system of forms and rules. Another is to see language as an essentially social/cultural phenomenon (e.g., Kirby 2011; Evans & Levinson 2009), and even as a game (Mark Aronoff p.c. July, 2010). It is a game that only humans can play because of our particular cognitive and biological makeup, but it is a game nonetheless. What is appealing about the idea of language as a game is that it frees us of the requirement to impose on language strictly formal constructs that are devoid of human experience. Instead, we play around with others in our community, who are also playing the language game, constantly constraining and expanding the system to meet communicative needs.

In § 2, we've seen some key similarities that emerge between spoken and signed words, supporting certain linguistic universals. In § 3, basic differences were highlighted, showing that each physical modality has a profound effect on molding language, and together revealing aspects of its essential nature. The newly emerging Al-Sayyid Bedouin Sign Language provides concrete empirical evidence in § 4 that the first and most basic element to emerge in language is the word. This young language also supports Mark Aronoff's notion of morphological complexity in words as abstract patterns within an explicitly morphological level of structure (§ 5). Comparing the words of natural spoken and signed languages allows us to explore the consequences of the language game, pinpointing some of the universal constraints on this level of structure, and expanding our understanding of what a word can be.

References

Aronoff, Mark. 1976. *Word formation in generative grammar*. Cambridge, MA: MIT Press.

Aronoff, Mark. 1994. *Morphology by itself, stems and inflectional classes*. Cambridge, MA: MIT Press.

Aronoff, Mark, Irit Meir, Carol Padden & Wendy Sandler. 2003. Classifier constructions and morphology in two sign languages. In Karen Emmorey (ed.). 53–84.

Aronoff, Mark, Irit Meir, Carol Padden, & Wendy Sandler. 2008. The roots of linguistic organization in a new language. *Special Issue of Interaction Studies on Holophrasis, Compositionality and Protolanguage* 9(1). 133–153.

Aronoff, Mark, Irit Meir & Wendy Sandler. 2005. The paradox of sign language morphology. *Language* 81(2). 301–344. https://doi.org/10.1353/lan.2005.0043

Battison, Robbin. 1978. *Lexical borrowing in American Sign Language*. Silver Spring, MD: Linstok Press.

Brennan, Mary. 1990. Word formation in British Sign Language. Stockholm, Sweden: University of Stockholm dissertation.

Brentari, Diane. 1990. Theoretical foundations of American Sign Language phonology. Chicago, IL: University of Chicago dissertation.

Brentari, Diane. 1998. *A prosodic model of sign language phonology*. Cambridge, MA: MIT Press.

Brentari, Diane. 2019. *Sign language phonology (Key topics in phonology)*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316286401

Cormier, Kearsy, Sandra Smith & Martine Zwets. 2013. Framing constructed action in British Sign Language narratives. *Journal of Pragmatics* 55. 119–139. https://doi.org/10.1016/j.pragma.2013.06.002

Coulter, Geoffrey. 1982. On the nature of ASL as a monosyllabic language. Paper presented at the Annual Meeting of the Linguistic Society of America, San Diego, CA, December 27–30.

Crasborn, Onno. 2011. The other hand in sign language phonology. In Marc van Oostendorp, Colin J. Ewen, Elizabeth V. Hume & Keren Rice (eds.), *The Blackwell companion to phonology*, 223–240. Oxford: Wiley-Blackwell. https://doi.org/10.1002/9781444335262.wbctp0010

Dingemanse, Mark & Kimi Akita. 2017. An inverse relation between expressiveness and grammatical integration: On the morphosyntactic typology of ideophones, with special reference to Japanese. *Journal of Linguistics* 53(3). 501–532. https://doi.org/10.1017/S002222671600030X

Dingemanse, Mark, Damián E. Blasi, Gary Lupyan, Morton H. Christiansen & Padraic Monaghan. 2015. Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences* 19(10). 603–615. https://doi.org/10.1016/j.tics.2015.07.013

Emmorey, Karen (ed.). 2003. *Perspectives on classifiers constructions in sign languages*. Mahwah, NJ: Lawrence Erlbaum. https://doi.org/10.4324/9781410607447

Evans, Nick & Steven Levinson. 2009. The myth of language universals. *Behavioral and Brain Sciences* 32. 429–492. https://doi.org/10.1017/S0140525X0999094X

Fernald, Theodoreu B. & Donna Jo Napoli. 2000. Exploitation of morphological possibilities in signed languages: Comparison of American Sign Language with English. *Sign Language & Linguistics* 3(1). 3–58. https://doi.org/10.1075/sll.3.1.03fer

Fischer, Susan & Qunhu Gong. 2010. Variation in East Asian sign languages. In Diane Brentari (ed.), *Sign languages: A Cambridge survey*, 499–518. New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511712203.023

Hockett, Charles D. 1960. The origin of speech. *Scientific American* 203. 89–96. https://doi.org/10.1038/scientificamerican0960-88

Israel, Assaf & Wendy Sandler. 2011. Phonological category resolution: A study of handshapes in younger and older sign languages. In Rachel Channon & Harry van der Hulst (eds.), *Formational units in sign language*, 177–202. Berlin: De Gruyter.

Johnston, Trevor & Adam C. Schembri. 1999. On defining lexeme in a signed language. *Sign Language & Linguistics* 2(2). 115–185. https://doi.org/10.1075/sll.2.2.03joh

Kirby, Simon. 2011. Language as an adaptive system: The role of cultural evolution in the origins of structure. In Maggie Tallerman & Kathleen Gibson (eds.), *Oxford Handbook of Language Evolution*, 589–604. Oxford: Oxford University Press.

Klima, Edward S. & Ursula Bellugi. 1979. *The signs of language*. Cambridge, MA: Harvard University Press.

Kouwenberg, Silvia & Darlene LaCharité. 2004. Echoes of Africa: Reduplication in Caribbean Creole and Niger-Congo languages. *Journal of Pidgin and Creole Languages* 19. 285–332. https://doi.org/10.1075/jpcl.19.2.03kou

Kwok, Lily, Stephanie Berk & Diane Lillo-Martin. 2020. Person vs. locative agreement: Evidence from late learners and language emergence. *Sign Language & Linguistics 23:12, Special Issue in Memory of Irit Meir*, 17–37. https://doi.org/10.1075/sll.00042.kwo

Lepic, Ryan, Carl Börstell, Gal Belsitzman & Wendy Sandler. 2016. Taking meaning in hand. *Sign Language & Linguistics* 19(1). 37–81. https://doi.org/10.1075/sll.19.1.02lep

Liddell, Scott K. 2003. *Grammar, gesture, and meaning in American Sign Language*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511615054

Liddell, Scott. K. & Robert E. Johnson. 1986. American Sign Language compound formation processes, lexicalization, and phonological remnants. *Natural Language & Linguistic Theory* 4(4). 445–513. https://doi.org/10.1007/BF00134470

Liddell, Scott. K. & Robert E. Johnson. 1989. American Sign Language: The phonological base. *Sign Language Studies* 64(1). 195–277. https://doi.org/10.1353/sls.1989.0027

Lillo-Martin, Diane. 2012. Utterance reports and constructed action. In Roland Pfau, Marcus Steinbach & Bencie Woll (eds.). 365–387.

Lillo-Martin, Diane & Edward S. Klima. 1990. Pointing out differences: ASL pronouns in syntactic theory. *Theoretical Issues in Sign Language Research* 1. 191–210.

Lillo-Martin, Diane & Richard P. Meier. 2011. On the linguistic status of 'agreement' in sign languages. *Theoretical Linguistics* 37(3–4). 95–141.

Mandel, Mark. 1981. Phonotactics and morphophonology in American Sign Language. Berkeley, CA: University of California dissertation.

Mathur, Gaurav & Christian Rathmann. 2012. Verb agreement. In Roland Pfau, Marcus Steinbach & Bencie Woll (eds.). 136–157.

McCarthy, John. 1981. A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry* 12(3). 373–418.

McNeill, David. 1992. *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

Meier, Richard P., Kearsy Cormier & David Quinto-Pozos (eds.). 2002. *Modality and structure in signed and spoken languages*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486777

Meir, Irit. 1998. Syntactic-semantic interaction in Israeli Sign Language verbs: The case of backwards verbs. *Sign Language and Linguistics* 1. 3–33. https://doi.org/10.1075/sll.1.1.03mei

Meir, Irit. 2002. A cross-modality perspective on verb agreement. *Natural Language and Linguistic Theory* 20(2). 413–450. https://doi.org/10.1023/A:1015041113514

Meir, Irit. 2012. Word classes and word formation. In Roland Pfau, Marcus Steinbach & Bencie Woll (eds.). 77–111.

Meir, Irit, Mark Aronoff, Wendy Sandler & Carol Padden. 2010. In Sergio Scalise & Irene Vogel (eds.), *Cross-disciplinary issues in compounding*, 301–322. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.311.23mei

Meir, Irit, & Wendy Sandler. 2008. *Language in space: A window onto Israeli Sign Language*. Mahwah, NJ: Lawrence Erlbaum Associates.

Müller, Cornelia. 2018. Gesture and sign: Cataclysmic break or dynamic relations? *Frontiers in Psychology* 9. https://doi.org/10.3389/fpsyg.2018.01651

Newport, Elissa L. & Richard P. Meier. 1985. The acquisition of American Sign Language. In Dan Isaac Slobin (ed), *The crosslinguistic study of language acquisition. Volume 1. The data*, 881–938. Hillsdale, NJ: Lawrence Erlbaum Associates.

Nespor, Marina & Wendy Sandler. 1999. Prosodic phonology in Israeli Sign Language. *Language and Speech* 42(2&3). 143–176. https://doi.org/10.1177/00238309990420020201

Padden, Carol A. (2016. *Interaction of morphology and syntax in American Sign Language*. Routledge Library Editions: Syntax. New York: Routledge. https://doi.org/10.4324/9781315449685

Padden, Carol, Irit Meir, Mark Aronoff & Wendy Sandler. 2010. The grammar of space in two new sign languages. In Diane Brentari (ed.), *Sign languages: A Cambridge survey*, 573–595. New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511712203.026

Pfau, Roland, Marcus Steinbach & Bencie Woll (eds.). 2012. *Sign Language: An international handbook*. Berlin: Walter de Gruyter. https://doi.org/10.1515/9783110261325

Sandler, Wendy. 1987. Assimilation and feature hierarchy ASL. *Chicago Linguistics Society Parasession on Autosegmental Phonology*, 266–278.

Sandler, Wendy. 1989. *Phonological representation of the sign: Linearity and nonlinearity in American Sign Language*. Dordrecht: Foris. https://doi.org/10.1515/9783110250473

Sandler, Wendy. 1990. Temporal aspects in ASL phonology. *Theoretical issues in sign language research. Volume 1. Linguistics*, 7–36. Chicago: University of Chicago Press.

Sandler, Wendy. 1993. Hand in hand: The roles of the nondominant hand in sign language phonology. *The Linguistic Review* 10. 337–390. https://doi.org/10.1515/tlir.1993.10.4.337

Sandler, Wendy. 1999. Cliticization and prosodic words in a sign language. In Tracy Alan Hall & Ursula Kleinhenz (eds.), *Studies on the phonological word*, 223–254. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.174.09san

Sandler, Wendy. 2008. The syllable in sign language: Considering the other natural modality. In Barbara Davis & Krisztina Zajdo (eds.), *The syllable in speech production*, 379–408. New York: Taylor Francis.

Sandler, Wendy. 2012a. Dedicated gestures in the emergence of sign language. *Gesture* 12(3). 265–307. https://doi.org/10.1075/gest.12.3.01san

Sandler, Wendy. 2012b. The phonological organization of sign languages. *Language and Linguistics Compass* 6(3). 162–182. https://doi.org/10.1002/lnc3.326
Sandler, Wendy. 2017. The challenge of sign language phonology. *Annual Review of Linguistics* 3. 43–63. https://doi.org/10.1146/annurev-linguistics-011516-034122
Sandler, Wendy, Mark Aronoff, Irit Meir & Carol Padden. 2011. The gradual emergence of phonological form in a new language. *Natural Language and Linguistic Theory* 29. 503–543. https://doi.org/10.1007/s11049-011-9128-2
Sandler, Wendy, Mark Aronoff, Carol Padden & Irit Meir. 2014. Language emergence: Al-Sayyid Bedouin Sign Language. In Jack Sidnell, Paul Kockelman & Nick J. Enfield (eds.), *The Cambridge handbook of linguistic anthropology*, 250–284. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139342872.012
Sandler, Wendy & Diane Lillo-Martin. 2006. *Sign language and linguistic universals*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139163910
Sandler, Wendy, Irit Meir, Carol Padden & Mark Aronoff. 2005. The emergence of grammar: Systematic structure in a new language. *Proceedings of the National Academy of Sciences* 102(7). 2661–2665. https://doi.org/10.1073/pnas.0405448102
Schembri, Adam. 2003. Rethinking 'classifiers' in signed languages. In Karen Emmorey (ed.). 13–44.
Senghas, Anne & Marie Coppola. 2001. Children creating language: How Nicaraguan Sign Language acquired a spatial grammar. *Psychological Science* 12(4). 323–328. https://doi.org/10.1111/1467-9280.00359
Stokoe, William C. 1960. *Sign language structure*. Silver Spring, MD: Linstock Press.
Supalla, Ted. 1986. The classifier system in American Sign Language. *Noun Classes and Categorization* 7. 181–214. https://doi.org/10.1075/tsl.7.13sup
Taub, Sara. 2001. *Language from the body: Iconicity and metaphor in American Sign Language*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511509629
Tkachman, Oksana & Irit Meir. 2018. Novel compounding and the emergence of structure in two young sign languages. *Glossa* 3(1). 136–140. https://doi.org/10.5334/gjgl.632
Wilbur, Ronnie. B. 1993. Syllables and segments: Hold the movement and move the holds!. In Geoffrey R. Coulter (ed.), *Current issues in ASL phonology*, 135–168. New York: Academic Press. https://doi.org/10.1016/B978-0-12-193270-1.50012-1
Zwitserlood, Inge. 2012. Classifiers. In Roland Pfau, Marcus Steinbach & Bencie Woll (eds.). 158–186.

CHAPTER 5

Leaving the stem by itself

Olivier Bonami and Sacha Beniamine
Université de Paris, LLF, CNRS / University of Surrey, SMG

Stem allomorphy plays a central role in the recent history of morphology, in no small part thanks to a research program initiated by Aronoff (1994). Yet, there is no agreed upon way of deciding whether some bit of form should be considered a proper part of a stem allomorph or an independent exponent. We explore the possibility of just doing away with the notion of stem allomorphy in inflection. We use computational methods to identify within each word a sequence of strings that do not take part in any alternation within that word's paradigm. We then discuss the relationship of such sequences to the classical notion of a stem, and argue that discontinuous stems are both conceptually and empirically more satisfactory.

Keywords: stems, allomorphy, exponence, alignment, European Portuguese, English, French

1. Introduction

The notion of stem allomorphy plays a central role in the recent history of morphology, in no small part thanks to a research program initiated by Aronoff (1994) that embraces the systematic existence of organized sets of stem allomorphs as a way of making sense of patterns of morphological behavior. Yet, as Spencer (2012) notes, a fundamental problem for this approach is that there is no agreed upon way of deciding whether some bit of form should be considered a proper part of a stem allomorph (and hence not a morphological unit) or an independent exponent.

In this chapter we explore the possibility of just doing away with the notion of stem allomorphy in the context of inflection. We use computational methods from item-and-pattern morphology to identify within each word a (possibly discontinuous) sequence of strings that are inflectionally inert, in the sense that they do not take part in any morphophonological alternation within that word's paradigm. Such discontinuous sequences can be identified through simple methods and are

https://doi.org/10.1075/cilt.353.05bon

indisputably devoid of morphosyntactic import of any kind. We argue that these 'discontinuous stems' are a useful addition to the morphologist's toolbox, and question whether classical collections of stem allomorphs still have a useful role to play. Specifically, we define an algorithmic procedure for inferring both discontinuous stems and sets of continuous stem allomorphs from raw paradigms, and assess their usefulness in capturing the implicative structure of the French, English and European Portuguese conjugation systems.

2. The quest for stem allomorphy

One of the basic questions facing morphology is what we call the 'Inflected Word Recognition Problem' (IWRP):

(1) What allows speakers to draw inferences from a word's form to its content?

The IWRP is modeled on Ackerman, Blevins & Malouf's (2009) 'Paradigm Cell Filling Problem' (2). Like the PCFP, it is a question about the structure of the morphological system (what allows the inferences to be drawn) rather than a question about learning and processing (which exact inferences are drawn, and how).

(2) What licenses reliable inferences about the inflected […] surface forms of a lexical item? (Ackerman et al., 2009: 54)

A speaker faces the IWRP each time they are exposed to an inflected word for the first time, a common situation (Bonami & Beniamine 2016). The shape of the solution is obvious: when seeing an unknown wordform *w*, speakers are able to identify phonological aspects of *w* that they have previously encountered associated with some content, and consider the possibility that these convey the same content in this new instance. The question faced by morphologists is, what are the relevant aspects of the phonology of *w*. Importantly, the IWRP and the PCFP raise different issues for morphology. Although a classical, constructive approach to morphology typically addresses both issues by reference to a single segmentation, it may well be that aspects of forms that are crucial for inferring other forms are distinct from those aspects that are crucial for inferring content.

In this chapter we focus on a subproblem of the IWRP, namely identifying that part of a wordform that conveys the word's lexical meaning, as opposed to morphosyntactic and morphosemantic content. In Matthews's (1972) terms, this amounts to separating out a word's (inflectional) 'stem' from the 'exponents' combining with that stem. In the simplest systems, this is an easy task. As Corbett (2007) highlights, canonically, each lexeme is equipped with a unique stem that conveys solely lexical information, while each paradigm cell is characterized by a combination of exponents that conveys only inflectional information; the unique stem is hence the relevant part of the word. This is illustrated in the small slice of the European Portuguese conjugation system presented in Table 1: stems are the longest substrings that are constant across rows, exponents are the longest substrings that are constant across columns.

Table 1. Indicative present of three European Portuguese first conjugation verbs

	1SG	2SG	3SG	1PL	2PL	3PL
FICAR	ˈfik*u*	ˈfik*ɐʃ*	ˈfik*ɐ*	fiˈk*ɐmuʃ*	fiˈk*aiʃ*	ˈfik*ɐ̃ũ*
ENTRAR	ˈẽtɾ*u*	ˈẽtɾ*ɐʃ*	ˈẽtɾ*ɐ*	ẽˈtɾ*ɐmuʃ*	ẽˈtɾ*aiʃ*	ˈẽtɾ*ɐ̃ũ*
TENTAR	ˈtẽt*u*	ˈtẽt*ɐʃ*	ˈtẽt*ɐ*	tẽˈt*ɐmuʃ*	tẽˈt*aiʃ*	ˈtẽt*ɐ̃ũ*

Inflectional systems are rarely that simple though. First, consideration of the other conjugation classes of European Portuguese reveal the existence of substrings that are constant neither across cells nor across lexemes, highlighted by underlining in Table 2. The theoretical status of these theme vowels is a constant source of hesitation for the Romance morphologist: should they be considered to be part of the stem? If so, we have pervasive stem allomorphy, and stems express more than just lexical meaning: for instance, the use of the consonant-final allomorph signals that we are dealing with a 1SG form. Should they be considered to be stem-external exponents? If so, then some material outside of the stem contributes to expressing lexical identity, as theme vowels narrow down the set of candidate solutions for the identification of which lexeme was used. We will not be proposing a solution to this puzzle, but we argue that the puzzlement is caused by the assumption that words can be segmented into a part expressing lexical identity only and a part expressing inflectional content only. Postulating stem allomorphs amounts to saving exponents from lexical value at the cost of imbuing stems with exponential value. We see no principled reason why this or the opposite solution should be preferable.[1]

1. Walther & Sagot (2011); Walther (2013) address the issue by comparing the description length of full implemented descriptions of a system relying on one or the other hypothesis, within a given formal framework for morphological description. This is in essence an operational implementation of the early generative grammar notion of an evaluation metric for alternative theories, and in that sense is a remarkable effort to model the heuristics that morphologists use to decide on a segmentation. However Walther & Sagot's empirical results show that variation in length among descriptions are small enough that the costs of a suboptimal choice are low, both for a speaker and for an analyst.

Table 2. Indicative present of three European Portuguese fully regular verbs

	1SG	2SG	3SG	1PL	2PL	3PL
FICAR	ˈfik*u*	ˈfikɐʃ	ˈfikɐ	fiˈkɐ*muʃ*	fiˈkaiʃ	ˈfikɐ̃ũ
VIVER	ˈviv*u*	ˈvivəʃ	ˈvivə	viˈve*muʃ*	viˈveiʃ	ˈvivɐ̃ĩ
IMPRIMIR	ĩpˈɾim*u*	ĩpˈɾiməʃ	ĩpˈɾimə	ĩpɾiˈmi*muʃ*	ĩpɾiˈmiʃ	ĩpˈɾimɐ̃ĩ

A second complication is apparent when looking at the data in Table 3. European Portuguese exhibits stress-conditioned vowel alternations that affect what we call the pre-thematic vowel. While these alternations are almost categorically predictable from regular phonology, they do have an impact on the IWRP. Consider CHEGAR: within its paradigm, having a /ə/ in prethematic position is a partial indication of the fact that the form is 1PL or 2PL. Hence, while this /ə/ does contribute to the expression of lexical identity, it also has some exponential import; on the face of it, deciding whether it should be considered as part of the stem (which then entails that CHEGAR has multiple stem allomorphs beginning in /ʃeg/ or /ʃəg/) or a separate unit should be just as contentious as deciding whether theme vowels are part of the stem.

Table 3. Stress-conditioned vowel alternations in European Portuguese verbs

	1SG	2SG	3SG	1PL	2PL	3PL
CHEGAR	ˈʃeg*u*	ˈʃegɐʃ	ˈʃegɐ	ʃəˈgɐ*muʃ*	ʃəˈgaiʃ	ˈʃegɐ̃ũ
COMEÇAR	kuˈmɛs*u*	kuˈmɛsɐʃ	kuˈmɛsɐ	kuməˈsɐ*muʃ*	kuməˈsaiʃ	kuˈmɛsɐ̃ũ
LIBERTAR	liˈbɛɾt*u*	liˈbɛɾtɐʃ	liˈbɛɾtɐ	libəɾˈtɐ*muʃ*	libəɾˈtaiʃ	liˈbɛɾtɐ̃ũ
PAGAR	ˈpag*u*	ˈpagɐʃ	ˈpagɐ	pɐˈgɐ*muʃ*	pɐˈgaiʃ	ˈpagɐ̃ũ
CHAMAR	ˈʃɐm*u*	ˈʃɐmɐʃ	ˈʃɐmɐ	ʃɐˈmɐ*muʃ*	ʃɐˈmaiʃ	ˈʃɐmɐ̃ũ
RETOMAR	rəˈtom*u*	rəˈtomɐʃ	rəˈtomɐ	rətuˈmɐ*muʃ*	rətuˈmaiʃ	rəˈtom*ɐ̃ũ*
JOGAR	ˈʒɔg*u*	ˈʒɔgɐʃ	ˈʒɔgɐ	ʒuˈgɐ*muʃ*	ʒuˈgaiʃ	ˈʒɔgɐ̃ũ
MUDAR	ˈmud*u*	ˈmudɐʃ	ˈmudɐ	muˈdɐ*muʃ*	muˈdaiʃ	ˈmudɐ̃ũ

Interestingly, discussions of this or similar data with morphologists over the years suggest to us that many would rather treat theme vowels as stem-external but prethematic vowels and similar segments as part of a stem alternant. We submit that such preferences are due to the use of the two heuristic principles in (3).

(3) a. Stem alternants should not be multiplied.
b. Stems should be continuous substrings of words.

Of course neither of these principles can be absolute: stem suppletion cannot be accomodated without relaxing (3a), and infixation is an immediate violation of (3b). But they still have some role to play in less extreme cases such as those found in the European Portuguese data under examination. More specifically, since theme vowels are no threat to (3b) nothing precludes one from adhering to (3a) and having shorter stems. In the case of prethematic vowels, however, adhering to (3a) would entail postulating discontinuous stems, hence it is more tempting to be guided by (3b) and have longer stems.

This example suggests that these principles play an important role in shaping the segmentations that are typically taken to be pre-theoretical decisions not worthy of detailed discussion, but that ultimately constrain our perception of the nature and extent of typological variation across inflection systems. In the absence of a principled way of arbitrating which of the two principles is more important, we run the risk of theorizing on shaky grounds.

From these observations we conclude with Spencer (2012) that there is no agreed-upon method for identifying which part of an inflected word is a stem, and that the heuristics used by morphologists in that area are neither systematic nor principled enough. This is unsatisfactory, given the prominent place taken by stem allomorphy in morphological theorizing in the wake of Aronoff (1994) – see, among many others, Maiden (1992); Brown (1998); Cameron-Faulkner & Carstairs-McCarthy (2000); Pirrelli & Battista (2000); Stump (2001); Bonami & Boyé (2002); Blevins (2003); Boyé & Cabredo Hofherr (2006); Montermini & Bonami (2013); Stump & Finkel (2013), and the papers collected in Bonami (2012).

One possible reaction to this situation is to forego segmentation entirely, and stop worrying about stems. Recent literature has highlighted how purely word-based methods can efficiently be deployed to address morphological problems, most prominently the PCFP (see among many others Ackerman et al. 2009; Ackerman & Malouf 2013; Blevins 2016; Bonami & Beniamine 2016; Sims & Parker 2016). If stem identification proves problematic, perhaps we should just dismiss the very notion of a stem.

Such a reaction isn't entirely satisfactory, however, at this point in the history of the field. First, since Carstairs-McCarthy's seminal work (Carstairs 1987; Carstairs-McCarthy 1994), stem allomorphy has been taken by many to be governed by constraints distinct from those governing the distribution of exponents. In fact, as Blevins (2016) notes, much of the literature about morphomic stem distributions is implicitly or explicitly dedicated to addressing predictability in paradigms.[2] Hence it is important to establish whether stem allomorphs have a dedicated role to play in addressing the PCFP. Second, we take the PCFP and the IWRP to be two

2. See also Stump & Finkel (2013) on exponent-based vs. stem-based implicative relations. Note though that Bonami & Boyé (2014) explicitly argue on the basis of a detailed explicit comparison of stem-based vs. word-based explorations of French conjugation that segmentation decisions for stems actually get in the way of an understanding of implicative structure.

complementary crucial questions that morphological theory should address; and while the PCFP might not require dealing with subword material, such a position is, at first sight, harder to defend for the IWRP, for which subword structure seems crucially relevant.

In the remainder of this chapter we attempt to further our understanding of these issues by exploring the consequences of the principles in (3). We examine two extreme ways of prioritizing the two principles. The first option is to take (3a) as absolute, and tolerate no stem allomorphy. As a consequence, only substrings that occur in all forms of a lexeme can be considered as being part of this stem, and stems will more often than not be discontinuous sequences of substrings of words. This we call the 'unique discontinuous stem hypothesis'. The second option is to take (3b) as absolute, and tolerate no discontinuity. As a consequence, stem allomorphy will be pervasive, as any variation in shape that is encapsulated within stem material has to also be stem material. This we call the 'continuous stem sets hypothesis'.

Although they only constitute two extremes of a vast space of possible ways of applying the principles in (3b), these two options have the advantage that they are simple enough to be implemented in full and tested on realistically-sized dataset. In § 3, we outline our implementation of the two hypotheses. In § 4 we examine how useful the two hypotheses turn out to be in terms of capturing implicative structure. Section 5 draws some general conclusions.

3. Automatic inference of stems

3.1 Alignment

In this section we present an algorithmic method for inferring both single discontinuous stems and sets of continuous stem allomorphs from raw paradigms of surface forms. In order to find stem-like material in forms, we align together all the forms of each paradigm using a heuristic algorithm for multiple alignments. The goal of such an algorithm is illustrated in Table 4. Each cell of the table contains a single phoneme, each row represents a surface form of the paradigm, and each column represents matched material across forms.

The optimal pairwise alignment of two forms can be computed easily given a scoring scheme (Needleman & Wunsch 1970). This is the basis for systems which generate alternation patterns, such as Albright & Hayes's (2006) Minimum Generalization Learner and Beniamine's (2018) Qumin. The generalization to multiple sequences is much more complex, especially for large numbers of sequences. Fortunately, this problem was researched in depth in evolutionary biology (see

Table 4. Aligned surface forms of LIBERTAR

Paradigm cell	Aligned form									
PRS.IND.1SG	l	i	b	ɛ	ɾ	t	–	–	u	–
PRS.IND.2SG	l	i	b	ɛ	ɾ	t	ɐ	–	–	ʃ
PRS.IND.3SG	l	i	b	ɛ	ɾ	t	ɐ	–	–	–
PRS.IND.1PL	l	i	b	ə	ɾ	t	ɐ	m	u	ʃ
PRS.IND.2PL	l	i	b	ə	ɾ	t	a	–	i	ʃ
PRS.IND.3PL	l	i	b	ɛ	ɾ	t	ɐ̃	–	ũ	–

Durbin 1998) for the purpose of aligning DNA and protein sequences. Several heuristic methods thus exist, which can find good solutions in reasonable time, often by performing repeated pairwise alignments. Some of these methods were adapted to sequences of phonemes by List (2014) in order to compare potential cognate words across languages. Our implementation was written specifically for the alignment of inflectional paradigm. The algorithm we rely on is called 'progressive alignments' (Feng & Doolittle 1987), and proceeds in three steps:

1. Find all pairwise alignments, obtaining scores for each pair of form.
2. Using the scores, perform hierarchical clustering. We obtain a binary guide tree which relates sequences.
3. Align sequences pairwise along the guide tree, first joining two sequences together, then joining sequences to alignments, or alignments to alignments.

Our adaptation to inflectional paradigms has two important properties, building on Beniamine (2018). First, our scoring scheme uses phonological similarity (Frisch 1997). This measure relies on natural classes and captures satisfactory morphological alternations in pairwise alignment (Albright & Hayes 2006; Beniamine 2018). Rather than minimizing edit distances, our algorithm maximizes similarity scores. Since they are to be expected in inflection, insertions (alignments to gaps) are free, while substitution costs are proportional to phonological similarity. The consequences of this strategy can be seen in Table 4, where columns contain either entirely identical or similar phonological segments. Second, there is often more than one optimal alignment at each intermediate pairwise step. Most algorithms deal with this ambiguity by selecting a solution randomly. Instead, we keep up to 30 competing alignments at all time. We then choose among these hypotheses based on the generality of the resulting segmentation.

3.2 Unique discontinuous stems

From an alignment table, we can easily identify the sequence of substrings which are constant across the paradigm: they appear in columns of identical segments (shaded in Table 4). This sequence can be discontinuous, and may be of different length for different lexemes. We call the phonemes in these columns ‘inflectionally inert’.

Sequences of inert segments that are always adjacent across the paradigm are collected into strings. The unique discontinuous stem is the sequence of these strings. The sequence is continuous if no cell in the paradigm has exponential material occurring between inert segments. It is fragmented if and only if at least one cell splits the stem in at least two parts. In our example, the unique discontinuous /lib-ɾ-/ is fragmented into two substrings /lib/ and /ɾ/ due to the vocalic alternation between /ə/ and /ɛ/.

It is worth noting that unique discontinuous stems go against much of our muscle memory on the representation of morphophonology. Suppletion is not captured at all. If traditional suppletive stems have nothing in common (think, e.g., of Latin *fero* and *tuli*) they will be inferred to have an empty common unique stem – a reasonable conclusion. If however they by chance have a common substring (e.g., Latin *tuli* and *latum*), that will be considered to be the unique common substring (here ⟨tu⟩). Hence the length of the unique stem is not an indicator of suppletion in the traditional sense. Another potentially surprising aspect is that all phenomena of nonperipheral inflection lead to the same outcome of discontinuities in the unique stem, irrespective of whether they are caused by, e.g., infixation, vowel alternations, stress shift, or root and pattern morphology. While this may seem counterintuitive, we do not think it is problematic: the causes of discontinuity may be variegated, they still are all discontinuities, which raise the same challenges for speakers trying to draw inferences from a wordform.

Table 5 gives some statistics on the prevalence of fragmented stems in English, French, and European Portuguese.[3] Lexemes with zero stem fragment are cases of suppletion, although, as discussed above, some cases of suppletion will result in one or more fragments. Unsurprisingly, fragmented stems are attested but infrequent in both English (due to forms such as *ring, rang*) and French (e.g., *mener* /məne/, *mène* /mɛn/), but much more prevalent in European Portuguese, where they make up one third of the lexicon; the vast majority of these are due to the stress-conditioned vowel alternations documented above.

3. Data is derived from CELEX for English (Baayen, Piepenbrock & Gulikers 1995), *Flexique* for French (Bonami, Caron & Plancq 2014), and the Coimbra pronuciation dictionary for European Portuguese (Veiga, Candeias & Perdigão 2012).

Table 5. Stem fragmentation in three systems: Number of fragments in a discontinuous stem

	0	1	2	3
English	2	5890	172	0
French	4	5136	109	0
Portuguese	2	1255	738	1

3.3 Sets of continuous stems

We can now build on discontinuous stems to deduce a set of continuous stems. The idea here is very simple: for each word, we count as its continuous stem the shortest sequence that encompasses all substrings of the lexeme's discontinuous stem. The stem set associated with a lexeme is then the set of continuous stems of its inflected forms. Table 6 illustrates.

Table 6. Inference of continuous stem set

Cell	Form	Disc. stem	Cont. stem
1SG	libɛɾtu	⟨lib, ɾt⟩	libɛɾt
2SG	libɛɾtɐʃ	⟨lib, ɾt⟩	libɛɾt
3SG	libɛɾtɐ	⟨lib, ɾt⟩	libɛɾt
1PL	libəɾtɐmuʃ	⟨lib, ɾt⟩	libəɾt
2PL	libəɾtaiʃ	⟨lib, ɾt⟩	libəɾt
3PL	libɛɾtɐ̃ũ	⟨lib, ɾt⟩	libɛɾt
Continuous stem set		{libɛɾt, libəɾt}	

Note that this method uniformly arbitrates for shorter and less numerous stems – notably, theme vowels are not taken to be parts of stems. Other than that, we submit that it matches closely the intuition outlined above and shared by many linguists that the postulation of stem alternants is justified by alternations occurring internal to what is otherwise lexical material.

Table 7 documents the prevalence of stem allomorphy in the three systems. The number of stems for French is much lower than assumed in Bonami & Boyé (2002) and related literature. This is due to the fact that Bonami & Boyé (2002) does not implement Principle (3a), and integrates theme vowels and other alternating material occurring marginally with respect to inert material into stem allomorphs. The classification matches closely more traditional accounts such as Swiggers & van den Eynde (1987). Again, Portuguese stands out compared to the other two languages, thanks to the prevalence of vowel alternations.

Table 7. Stem allomorphy in three systems: Number of distinct stem allomorphs per lexeme

	0	**1**	**2**	**3**	**4**
English	2	5909	129	24	0
French	4	5136	108	1	0
Portuguese	2	1255	630	102	7

4. How useful are continuous stem allomorphs?

We now turn to the question we asked at the end of § 2: how helpful are the two notions of stem under investigation in capturing the inferential structure of paradigms, as characterized by the IWRP and the PCFP?

In terms of the IWRP, the answer is quite simple. Sets of continuous stems are by definition less useful than a unique discontinuous stem: the unique discontinuous stem identifies *exactly* that part of the word that has no exponential value, while stem allomorphs blur the distinction between exponential and nonexponential material. In this connection it is worth reflecting on the paradigm of a verb such as LIBERTAR. In a form such as 3SG /libɛɾtɐ/, the /ɛ/ has partial exponential value: it indicates that we are dealing either with a SG or a 3PL form. In this it is not different from the /ɐ/, which also gives partial information that we are dealing with a 2SG, 3SG or 1PL form. But both the /ɛ/ and the /a/ also give partial lexical information, as both narrow down the class of lexemes that we might be dealing with. This illustrates how continuous stem allomorphs are unhelpful with respect to the IWRP: insisting on contiguity stops us from seeing that /ɛ/ makes a contribution different from that of the neighbouring segments but similar to that of more peripheral segments.

While sets of stem allomorphs are counterproductive when addressing the IWRP, maybe they are helpful for the PCFP. We thus turn to the relationship of the two notions of stem to implicative structure.

A first way of addressing this is to consider the implicative structure of the system of exponents. Let us define holistically the exponential part of a wordform as whatever remains when the stem has been removed. Table 8 contrasts what the exponents look like in the present indicative of LIBERTAR depending on whether one works from the unique discontinuous stem or from the set of continuous stems.

Consider now the problem of predicting the exponent in one cell from the exponent in another cell. As the table should make clear, predictability is affected in both directions by one's view on stems. On the one hand, exponents are harder to predict under the unique stem hypothesis, since they contain more material – in this example, prethematic vowels, which may be nontrivial to predict. On the

Table 8. Two notions of exponence exemplified

	1SG	2SG	3SG	1PL	2PL	3PL
Full words	li'bɛɾtu	li'bɛɾtɐʃ	li'bɛɾtɐ	libəɾ'tɐmuʃ	libəɾ'taiʃ	li'bɛɾtɐ̃ũ
Unique disc. stem				⟨lib, ɾt⟩		
Exponents	_ɛ_u	_ɛ_ɐʃ	_ɛ_ɐ	_ə_ɐmuʃ	_ə_aiʃ	_ɛ_ɐ̃ũ
Continuous stems	libɛɾt	libɛɾt	libɛɾt	libəɾt	libəɾt	libɛɾt
Exponents	_u	_ɐʃ	_ɐ	_ɐmuʃ	aiʃ	_ɐ̃ũ

other hand, exponents are better predictors under this view, since they are larger and hence convey more information. In the case at hand, exponents based on the discontinuous stem include the prethematic vowel in the predictor form, which is partially predictive of its counterpart on the predicted form (Bonami & Luís 2014).

Given these two observations, and depending on how correlated central and peripheral exponence are in a particular system, one might expect unique discontinuous stems to lead to easier, harder, or equally difficult prediction than those based on sets of continuous stems. To establish this, we adapt the methodology of Ackerman & Malouf (2013): tables of exponents per lexeme and paradigm cell are used to estimate probability distributions of exponents by paradigm cell, taking into account the type frequency of exponents. We then compute the conditional entropy of choosing an exponent in a predicted cell from the choice of exponent for a predictor cell, and the average conditional entropy across ordered pairs of distinct cells. Table 9 reports the average results for English, French, and European Portuguese conjugation.

Table 9. Average conditional entropy of exponents under two conceptions of stems

	Continuous	Discontinuous
English	0.9917	0.9655
French	0.7109	0.7091
Portuguese	0.5782	0.6016

For all three languages we observe a variation of less than 0.03 bits, leading to the conclusion that there is no difference to speak of between the two strategies in terms of how much implicative structure they capture. For French and English, where stems are overwhelmingly continuous, this is unsurprising. For Portuguese however we might have expected a larger difference. Figure 1 allows us to explore this in more detail. This figure reports, for a distillation of the paradigm (Stump & Finkel 2013), the difference between conditional entropy for exponents based on continuous and discontinuous stems. A positive value means that the cell in row is more predictive of the cell in column if we reason with continuous stems than if

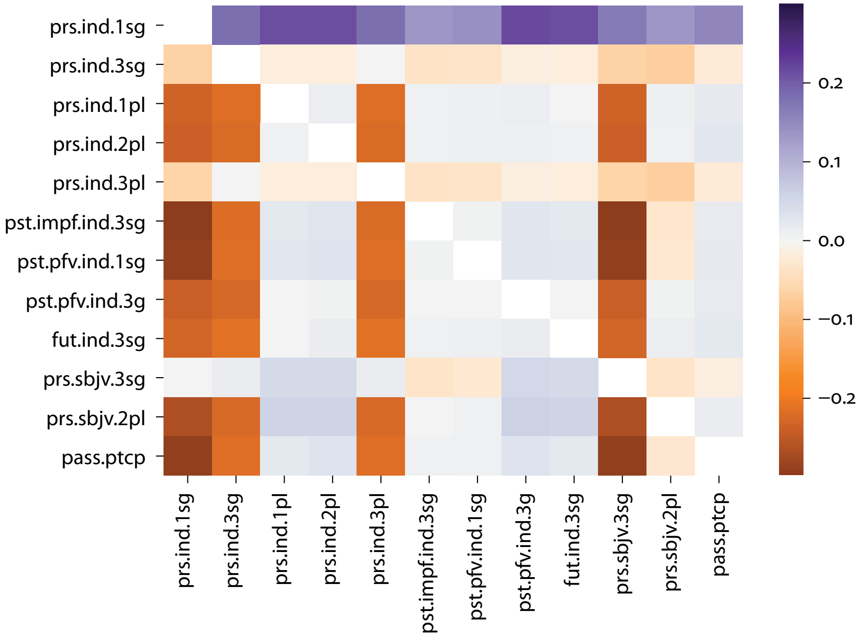


Figure 1. Heatmap of differences between conditional entropy of exponents based on continuous and dicontinuous stems, for a distillation of European Portuguese verbal paradigms

we reason with discontinuous stems. We can see that there are sizeable differences in predictability in both directions, notably located in the first two rows and two columns. However, the picture is almost symmetrical along the diagonal: where cell c is a better predictor of cell c' under continuous stems, cell c' tends to be a better predictor of cell c under discontinuous stems. This is easily explained by the distribution of prethematic vowels in Portuguese (Bonami & Luís, 2014). Negative values correspond to cases where the predicted cell has a stressed prethematic vowel that is hard to predict from a predictor cell with an unstressed prethematic vowel; positive values correspond to the opposite situation, where prediction of unstressed vowels is easy, but knowledge of the prethematic vowel helps predict the rest of the inflectional behavior.

Of course we would need to examine more languages to arrive at a general conclusion. However, based on the evidence examined in this chapter, although there are local predictive advantages to one or the other strategy for particular pairs of predictor and predictee, these balance each other on average, so that the overall predictive value of exponents defined from continuous or discontinuous stems is equivalent.

We now turn to a different question on the predictive value of stems. The continuous stem set hypothesis induces a segmentation of words into two subparts that each have their own nontrivial implicative structure. The question then arises how the implicative structure of these stems compares to that of whole, unsegmented words. The presumption is that predicting stem allomorph from stem allomorph should be easier than predicting word from word: first, by focusing on the stem, we are abstracting away from the hard problem of predicting exponent variation across inflection classes, which has been the central focus of attention of the literature on predictability since Carstairs (1987) and Wurzel (1989). Second, stems are expected to give rise to little variation overall across the paradigm, which should make them easy to predict.

To assess whether this presumption is warranted, we computed 'implicative entropy' (Bonami & Beniamine 2016; Beniamine 2018) for paradigms of whole words and paradigms of continuous stems. Implicative entropy is the conditional entropy of the alternation pattern linking two forms given the phonological shape of the predictor form. Unlike conditional entropy computed from exponents, implicative entropy does not assume prior knowledge of a segmentation of the forms under consideration. Hence it is an adequate way of assessing the predictive power of whole words.

Table 10. Average implicative entropy for stem allomorphs and for whole words

	Continuous stem allomorphs	**Words**
French	0.0196	0.1844
English	0.0415	0.1739
Portuguese	0.1483	0.1670

Table 10 reports average implicative entropy across all pairs of cells for the three languages under consideration. In French and English, we do get the expected results that continuous stem allomorphs are more predictive of each other than words are. This was to be expected, given the low prevalence of stem allomorphy in these two languages. In Portuguese, however, it is barely harder to predict words from words than stems from stems. This state of affairs has a clearly identifiable cause. The main two sources of unpredictability in European Portuguese conjugation are theme vowel alternations and prethematic vowel alternations (Bonami & Luís 2014). While these two phenomena are mostly orthogonal and complementary, the statistical distribution of prethematic and theme vowels is not entirely independent. Thus, knowledge of one vowel is informative of the other, and it is easier to predict both at the same time from joint knowledge of the two. As a result, segmentation into continuous stems and exponents is unhelpful as it segregates two pieces of information that are more usefully brought together.

In this section we have established that a divide-and-conquer strategy, where stems and exponents are first cleanly separated, does not lead to better performance in addressing either the IWRP or the PCFP. We conclude that stem allomorphs play no major role in capturing the predictive structure of paradigms.

5. Conclusion

Although segmentation of words into subword units plays a central role in morphology, there is a lack of attention in the literature to the motivation of segmentation choices and the consequences of these choices for later theorizing. In this chapter we addressed only part of this general issue by concentrating on stems. We identified two simple and coherent views on the nature of stems that are relatively easy to operationalize: unique discontinuous stems and sets of continuous stems.

We then presented an algorithmic method to infer both kinds of stems from raw paradigmatic data, and examined on that basis the usefulness of both conceptions in addressing what we take to be two central questions for morphology: predictability of forms from forms (the PCFP) and predictability of content from form (the IWRP). We concluded that the two concepts were of equally little use to address the PCFP, and that unique discontinuous stems were inherently more useful to addressing the IWRP. This leaves us questioning the usefulness of the concept of stem allomorphy for morphological theory.

There are different reasons to take this provocative conclusion with a bit of skepticism. First, the conclusion is dependent on particular choices when operationalizing segmentation principles: the devil is in the detail, and a different operationalization might lead to a different conclusion. For instance, note that our algorithm does not rely on any notion of optimization of the size of the lexicon, which typically plays a role in segmentation decisions. This is a conscious choice, motivated both by computational considerations and by literature disputing the usefulness of lexical optimization (see among many others Jackendoff 1975; Bochner 1993; Hay & Baayen 2005; Blevins 2006). However, it would be useful to see whether different operationalizations of segmentation principles, or the adoption of other principles, lead to different results.

Second, our reasoning here only applies to strictly synchronic aspects of morphology. Language change leads to situations where a unique continuous stem has multiple descendants that are related but not identical in form. As a way of describing that situation, 'stem allomorphy' is an essential concept in the description of morphological change. But importantly, such a concept of stem allomorphy is defined purely in terms of diachronic correspondences, not on anything that can be observed directly at a synchronic stage. Our conclusions hence lead to the question of whether a concept that is useful in the description of change is also useful in the description of the synchronic system.

Finally, some may find that we do not go far enough. Our point of departure was to address the IWRP, and more specifically to ask which aspects of a word's form provide information about lexical identity, as opposed to inflectional information. We went on to identify a unique discontinuous stem negatively, as the sequence that has no exponential value. But this cannot be the full story. As we noted before, in any system with multiple inflection classes, there will be material in a word that is partially indicative of lexical identity and partially indicative of morphosyntactic import. A full answer to the IWRP should not ignore this, and should ask instead, about each phonological segment in a word, what it contributes to narrowing down the content of the word. The unique discontinuous stem hypothesis addresses only the easier part of that question – but at least it addresses it.

Acknowledgements

This paper stems from ideas which were presented at a workshop on New approaches to the typology of inflection systems, organized by the now defunct CNRS *Fédération Typologie et Universaux* in Paris, France in November 2018; as well as an invited talk funded by Stony Brook University at the 4th American International Morphology Meeting, in Stony Brook, NY, in May 2019. We thank the organizers and organizing institutions of both events for the opportunity to present our work, as well as the audiences for fruitful discussion. Interactions with Farrell Ackerman, Mark Aronoff, Matthew Carroll, Grev Corbett, Berthold Crysmann, Sebastian Fedden, Jean-Pierre Koenig, and Adina Williams were particularly valuable. We also thank Matías Guzmán Naranjo for numerous engaging conversations on the topic of paradigm alignments, and Ana R. Luís for teaching us most of what we know about European Portuguese conjugation.

We would not even be able to ask the questions we address in this chapter without Mark Aronoff's influence, both in the form of writings which have shaped our thinking on these issues since the very beginning of our careers, and of friendly dialogue over the years. We can only hope that the present chapter will stimulate more of both.

Funding

This work was partially funded by a public grant overseen by the French National Research Agency (ANR) as part of the "Investissements d'Avenir" program (reference: ANR-10-LABX-0083).

References

Ackerman, Farrell, James P. Blevins & Robert Malouf. 2009. Parts and wholes: Implicative patterns in inflectional paradigms. In James P. Blevins & Juliette Blevins (eds.), *Analogy in grammar*, 54–82. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199547548.003.0003

Ackerman, Farrell & Robert Malouf. 2013. Morphological organization: The low conditional entropy conjecture. *Language* 89. 429–464. https://doi.org/10.1353/lan.2013.0054

Albright, Adam & Bruce Hayes. 2006. Modeling productivity with the gradual learning algorithm: The problem of accidentally exceptionless generalizations. In Gisbert Fanselow, Caroline Féry, Matthias Schlesewsky & Ralf Vogel (eds.), *Gradience in grammar: Generative perspectives*, 185–204. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199274796.003.0010

Aronoff, Mark. 1994. *Morphology by itself*. Cambridge: MIT Press.

Baayen, R. Harald, Richard Piepenbrock & Leon Gulikers. 1995. *Celex*. Philadelphia: Linguistic Data Consortium.

Beniamine, Sacha. 2018. Typologie quantitative des systèmes de classes flexionnelles. Paris: Université Paris Diderot dissertation.

Blevins, James. 2003. Stems and paradigms. *Language* 79. 737–767. https://doi.org/10.1353/lan.2003.0206

Blevins, James P. 2006. Word based morphology. *Journal of Linguistics* 42. 531–573. https://doi.org/10.1017/S0022226706004191

Blevins, James P. 2016. *Word and paradigm morphology*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199593545.001.0001

Bochner, Harry. 1993. *Simplicity in generative morphology*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110889307

Bonami, Olivier. 2012. Stems in inflection and lexeme formation. *Word Structure* 5(1). 1–6. https://doi.org/10.3366/word.2012.0016

Bonami, Olivier & Sacha Beniamine. 2016. Joint predictiveness in inflectional paradigms. *Word Structure* 9(2). 156–182. https://doi.org/10.3366/word.2016.0092

Bonami, Olivier & Gilles Boyé. 2002. Suppletion and stem dependency in inflectional morphology. In Franck Van Eynde, Lars Hellan & Dorothee Beerman (eds.), *The proceedings of the HPSG '01 conference*, 51–70. Stanford: CSLI Publications.

Bonami, Olivier & Gilles Boyé. 2014. De formes en thèmes. In Florence Villoing, Sarah Leroy & Sophie David (eds.), *Foisonnements morphologiques. etudes en hommage à Françoise Kerleroux*, 17–45. Presses Universitaires de Paris Ouest.

Bonami, Olivier, Gauthier Caron & Clément Plancq. 2014. Construction d'un lexique flexionnel phonétisé libre du français. In Franck Neveu, Peter Blumenthal, Linda Hriba, Annette Gerstenberg, Judith Meinschaefer & Sophie Prévost (eds.), *Actes du quatrième congrès mondial de linguistique française*, 2583–2596. https://doi.org/10.1051/shsconf/20140801223

Bonami, Olivier & Ana R. Luís. 2014. Sur la morphologie implicative dans la conjugaison du portugais: une étude quantitative. In Jean-Léonard Léonard (ed.), *Morphologie flexionnelle et dialectologie romane. typologie(s) et modélisation(s)*. 111–151. Leuven: Peeters.

Boyé, Gilles & Patricia Cabredo Hofherr. 2006. The structure of allomorphy in spanish verbal inflection. In *Cuadernos de lingüística*, vol. 13, 9–24. Instituto Universitario Ortega y Gasset.

Brown, Dunstan. 1998. Stem lndexing and morphonological selection in the Russian verb: A network morphology account. In Ray Fabri, Albert Ortmann & Teresa Parodi (eds.), *Models of inflection*, 196–224. Niemeyer. https://doi.org/10.1515/9783110919745.196

Cameron-Faulkner, Thea & Andrew Carstairs-McCarthy. 2000. Stem alternants as morphological signata: Evidence from blur avoidance in Polish nouns. *Natural Language and Linguistic Theory* 18. 813–835. https://doi.org/10.1023/A:1006496821412

Carstairs, Andrew. 1987. *Allomorphy in inflection*. London: Croom Helm.

Carstairs-McCarthy, Andrew. 1994. Inflection classes, gender, and the principle of contrast. *Language* 70. 737–788. https://doi.org/10.2307/416326

Corbett, Greville G. 2007. Canonical typology, suppletion and possible words. *Language* 83.8–42. https://doi.org/10.1353/lan.2007.0006

Durbin, Richard. 1998. *Biological sequence analysis: Probabilistic models of proteins and nucleic acids*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511790492

Feng, Da-Fei & Russell F. Doolittle. 1987. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *Journal of Molecular Evolution* 25(4). 351–360. https://doi.org/10.1007/BF02603120

Frisch, Stefan. 1997. Similarity and frequency in phonology. Evanston, IL: Northwestern University dissertation.

Hay, Jennifer B. & R. Harald Baayen. 2005. Shifting paradigms: gradient structure in morphology. *TRENDS in Cognitive Science* 9. 342–348. https://doi.org/10.1016/j.tics.2005.04.002

Jackendoff, Ray. 1975. Morphological and semantic regularities in the lexicon. *Language* 51. 639–671. https://doi.org/10.2307/412891

List, Johann-Mattis. 2014. *Sequence comparison in historical linguistics*. Düsseldorf: Düsseldorf University Press.

Maiden, Martin. 1992. Irregularity as a determinant of morphological change. *Journal of Linguistics* 28. 285–312. https://doi.org/10.1017/S0022226700015231

Matthews, Peter H. 1972. *Inflectional morphology. A theoretical study based on aspects of Latin verb conjugation*. Cambridge: Cambridge University Press.

Montermini, Fabio & Olivier Bonami. 2013. Stem spaces and predictibility in verbal inflection. *Lingue e Linguaggio* 12. 171–190.

Needleman, Saul B. & Christian D. Wunsch. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology* 48(3). 443–453. https://doi.org/10.1016/0022-2836(70)90057-4

Pirrelli, Vito & Marco Battista. 2000. The paradigmatic dimension of stem allomorphy in Italian verb inflection. *Rivista di Linguistica* 12.

Sims, Andrea D. & Jeff Parker. 2016. How inflection class systems work: On the informativity of implicative structure. *Word Structure* 9. 215–239. https://doi.org/10.3366/word.2016.0094

Spencer, Andrew. 2012. Identifying stems. *Word Structure* 5. 88–108. https://doi.org/10.3366/word.2012.0021

Stump, Gregory T. 2001. *Inflectional morphology. A theory of paradigm structure*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486333

Stump, Gregory T. & Raphael Finkel. 2013. *Morphological typology: From word to paradigm*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139248860

Swiggers, Pierre & Karel van den Eynde. 1987. La morphologie du verbe français. *ITL Review of Applied Linguistics* 77–78. 151–251. https://doi.org/10.1075/itl.77-78.04swi

Veiga, Arlindo Oliveira da, Sara Candeias & Fernando Perdigão. 2012. Generating a pronunciation dictionary for European Portuguese using a joint sequence model with embedded stress assignment. *Journal of the Brazilian Computer Society* 19(2). 127–134. https://doi.org/10.1007/s13173-012-0088-0
Walther, Géraldine. 2013. De la canonicité en morphologie: Perspective empirique, théorique et computationnelle. Paris: Université Paris Diderot dissertation.
Walther, Géraldine & Benoît Sagot. 2011. Modélisation et implémentation de phénomènes flexionnels noncanoniques. *Traitement Automatique des Langues* 52(2). 91–122.
Wurzel, Wolfgang Ulrich. 1989. *Inflectional morphology and naturalness*. Dordrecht: Kluwer.

CHAPTER 6

Stem constancy under the microscope

A systematic language comparison of types and limitations of stem spelling

Caroline Hettwer and Nanna Fuhrhop
Carl von Ossietzky Universität Oldenburg

Writing systems show variation in stem spellings, for example with double consonant letters. In German, the double consonant is always preserved (e.g., *rennen – rennt*, "to run – runs"), while in Dutch it is not (*rennen – rent*). In English <nn> is normally not preserved (*running – run*), though in French it varies: *bonne – bon* ("well" feminine – masculine), *donner – donne* ("to give – (I) give"). There are different regularities for double consonant spellings in all four languages; thus, stem constancy varies in strength depending on the language. In order to develop a typology of writing systems, language-specific types and limitations of stem constancy need to be described in greater detail. This chapter will do this by examining the above-mentioned four languages. It will compare the strength and degree of systematicity of stem constancy, as well as issues of frequency. Mark Aronoff was always interested in spelling systems and he is a morphologist – so this chapter is dedicated to Mark.

Keywords: graphemics, visible morphology, stem spellings, stem constancy, German, Dutch, English, French

1. Introduction

Writing systems provide a variety of information that goes beyond mere pronunciation. Morphological cues can be obtained, for example, from stem constancy and affix constancy. In addition, there are so-called morphological patterns, i.e., sequences of letters that occur in certain inflectional contexts or in a certain group of words and thus indicate a certain morphological structure. In the form used for the past tense, three out of the four modal verbs in English, for example, share the letter sequence <ould>: *could, would, should*. The letter sequence <ould> is therefore to be seen as a morphological spelling that marks a word as a modal verb (cf. Fuhrhop 2018: 599; Berg et al. 2014: 293–294). Such morphological spellings

https://doi.org/10.1075/cilt.353.06het

can be reading aids, providing the reader with lexical, grammatical or structural information. Aronoff, Berg & Heyer (2016) investigate the English writing system with regard to morphological cues that can support reading. They make the point that the consistent spelling of a suffix and the consistent mapping of one spelling onto one suffix are beneficial for the reading process, making it easier to recognize the suffix and decompose a word into stem and suffix. This principle can also be applied to stem spelling; if a word stem is constantly spelled in all inflectional and word formation processes, the stem can be recognized more easily and the decomposition of a complex wordform is facilitated.

In this chapter, we will take a closer look at stem constancy and examine how constant stems are spelled and where, how often, and in which way the spelling deviates (systematically) from stem constancy. Our aim is to further establish comparative graphematics, and to this end, we will use a language comparative approach to systematically compare stem spelling in German, English, Dutch and French. This allows for a better understanding of the characteristics and peculiarities of stem constancy in the individual languages, which may not be evident in monolinguistic studies. Typical spellings that explicitly show stem constancy in German include the doubling of consonants as in *Mann* ("man") that refers to the spelling *Männer* ("men") or the umlaut <ä> in *Männer* instead of <e> (which is also phonologically possible) to refer to <a> in *Mann*. However, this principle is not as pronounced in other languages as it is in German – stem constancy is ranked differently and works differently in different languages. In English, for example, most syllable joints are not transferred when occurring in word final position in related words: *runner – run,* but *kisser – kiss* (cf. Cummings 1988: 76). Catach (2003: 272) refers to different kinds of double consonant spelling in French, e.g., with derivation there is no deviation and some double consonants have special phoneme-grapheme correspondences like <cc> – /ks/. Although it has special correspondences, it also shows morphology within inflection like *bon – bonne,*[1] *chien – chienne*.

Our aim here is to capture and compare different types, possibilities and limitations of stem constancy in German, English, Dutch and French and develop a system of parameters for making stem spelling cross-linguistically comparable in a meaningful way. Working with abstract metalinguistic parameters allows a meaningful cross-linguistic comparison of the phenomenon, which works differently in each of the respective languages. This chapter is structured as follows: first, we introduce the theoretical background and our approach to cross-linguistic comparison and parameter development. Next, we take a look at stem constancy in the four languages and examine possible comparative parameters. We carry out a

1. The form *bon* is pronounced with a nasalized vowel, the form *bonne* with an alveolar nasal at the end.

corpus study to show how relevant the principle of stem constancy is in the respective languages, what kind of (systematic) deviations from stem constancy there are and how often a reader of these languages is confronted with stem constancy. The chapter ends with a summary and a preview of future research questions in the field of comparative/typological graphematics.

2. Stem constancy

We will now take a look at the characteristics and manifestations of stem constancy in German, English, French and Dutch. Each comparative aspect discussed will then be compared quantitatively using data from different morphologically annotated newspaper corpora. For German, the corpus Tagged T2 of Cosmas II is used (IDS, 2014). English is captured using sub-corpora of the British National Corpus (BNC, 2007). For stem constancy in Dutch, appropriate sub-corpora of OpenSoNaR (Oostdijk et al. 2013) are used, and for French, the corpus of the daily newspaper *Le Monde* (CLARIN.SI 2017) is used. In each case, the 1,000 most frequent wordforms are manually analyzed in order to show which spellings a reader frequently encounters. In addition, only nouns, adjectives and verbs are collected in order to exclude function words. Only the spellings of the word stems are then compared without subsequent affixes. Due to the information available in these newspaper corpora, we focus on the letter sequences – the congruencies of stem spellings, types and frequencies of deviations, not on phonological adequacy.

2.1 Means of stem constancy

The first thing we look at is what means are used to preserve a particular spelling. Ideally, the spelling of a word stem is kept identical regardless of any change in pronunciation. Examples for this can be found in all of the languages we investigated (see Table 1).

Table 1. Examples for identical stem spellings in German, English, Dutch and French

	German	English	Dutch	French
[+ identical stem spelling]	füllst – füllen, gewann – gewannen, floh – flohen, gab – gaben, Fuß – Füße, Mann – Männer, Hund – Hunde	autumn – autumnal, kiss – kisses, sign – signal, medic – medical – medicine – medicinal	dag – dagen, hond – honden, genieten – geniet	plomb – plombier, enfant – enfantin, fin – fine, chant – chanter, champ – champêtre

The spellings in Table 1 cannot be derived phonographically alone; they are morphologically motivated and thus are signs of visible morphology. Identical stem spellings in German make morphology visible by the presence of double consonant spellings that are not a syllable joint, post vocal <h>, word final voiced plosives (<b>, <d>, <g>) and word final <ß>. Keeping the spelling of the stem constant is a strong feature of German (cf. Eisenberg 2016; Dürscheid 2016; Neef 2013). In French, strongly alternating forms in spoken languages are often kept together visually. Phonetic allomorphism is thus encountered with an approach to graphic morpheme constancy (cf. Meisenburg 1996). For instance, the silent grapheme <b> in French *plomb* is preserved because it is pronounced in *plombier*. The stem constancy in French is even more noticeable in *fin* – *fine* ("fine" masculine – feminine), where the stem is written in the same way despite the different pronunciation [fɛ̃] and [fin]. Phonographically possible alternatives for *fin* would be *fen, *fain and *fein (cf. Catach 2003). It is now important to quantitatively compare how often the stem is kept identical in the respective languages and how often the stem is changed in order to find out how strong the principle of stem constancy is in a language. Thus, [± identical stem spelling] can be one parameter to find out how strong stem constancy is in a language.

Comparing the stems of the words with those of the lemmas, it is evident that – independent of the phonographic level – the sequence of letters of the word stem and the lemma stem is most often identical in French. The most deviations from identical stem spelling occur in Dutch (see Table 2).

Table 2. Percentage of identical stem spellings within the most frequent 1,000 wordforms in French, German, English and Dutch

	French	German	English	Dutch
Percentage of stem identical types	92.0%	87.4%	85.0%	78.6%
Percentage of stem identical token	81.6%	71.4%	70.6%	61.3%

An attenuated type of stem constancy is to keep the stem as constant as possible by replacing single letters with optically similar letters, which is for example systematically the case in French. Another attenuated type is to double or simplify letters, as is systematically done in Dutch. We will look at both types. The complex phoneme-grapheme correspondences in French make it possible to resort to equally possible – albeit less preferred – correspondences that are graphically similar (cf. Meisenburg 1996: 200):

(1) [trase] <tracer> "to draw" [trasɔ̃] <traçons>
[divɛrʒe] <diverger> "to differ" [divɛrʒɔ̃] <divergeons>

The correspondences of <c> and <g> differ depending on the following vowel (see Fuhrhop 2020, for a detailed discussion on visual verbal morphology in the Romance languages). In the first case <ss> would be the preferred correspondence, but <ç> is also possible and is chosen because of its graphic similarity. In the second case, <j> would be the preferred correspondence instead of <ge> but <ge> is selected because of its graphical resemblance. The <ä>-spelling in German is also part of the stem constancy; however, this is not a case of identical stem spelling but of optically similar stem spelling. In *Männer*, <ä>is preferred over the phonographically possible alternative <e> (*Menner) to keep the stem as similar as possible. These cases show that languages try to find ways to keep the stem as constant as possible, using less preferred grapheme correspondences which optically resemble the stem's letters. The second form of the attenuated stem constancy is the doubling or simplification of consonants or vowels. In these cases, a single letter is doubled, or one of two adjacent identical letters is removed (e.g., *maan – manen*, "moon – moons").

Keeping the stem identical would of course be a stronger form of stem constancy, but adding or removing one of two identical letters is still a minor change in the stem spelling compared to changing letters, as for example in French *vif – vive* ("brisk" masculine – feminine) or English <y>-replacement or <e>-deletion (cf. Berg 2013, for these graphemic alternations in English). Thus, in cases where no identical stem spelling can be preserved, there is still the possibility of an attenuated form of stem constancy, i.e., on the one hand the use of graphically similar letters (e.g., *Mann – Männer,* not *Menner, "man – men"), and on the other hand the doubling of letters (e.g., *manen – maan,* not *man, "moons – moon") or the simplification of doubled letters (e.g., *mannen – man,* not *mann, "men – man"). We can summarize these forms with the parameter [± graphical resemblance of stems], which helps us to work out the strength of stem constancy in different languages. Table 3 gives some examples on attenuated stem constancy.

Stem constancy is stronger when the stem is written identically in all forms; it is somewhat attenuated if it deviates slightly from an identical spelling. The general strength of stem constancy in a language can be determined with the help of the parameters [± identical stem spelling] and [± graphical resemblance of stems], but in order to be able to analyze it comprehensively, the degree of the deviations and the limitations of stem constancy must also be determined. We will develop parameters for describing these in the next sections.

Table 3. Examples of graphically resembling and graphically non-resembling stems in German, English, Dutch and French

	[+ **graphical resemblance of stems**]		[– **graphical resemblance of stems**]
	Optically similar letters	**Doubling or simplification of letters**	**Changing letters**
German	Mann – Männer, tragen – trägt, alt – älter	Haar – Härchen	lügen – log, schlafen – schlief
English		runner – run, dig – digger	lady – ladies, foot – feet, man – men, think – thought
Dutch		dromen – droom, rollen – rol, vraag – vragen, man – mannen	blijven – blijf, lezen – lees
French	tracer – traçons, annoncer – annonçait	bon – bonne, gentil – gentille, appeler – appelle	cinq – cinquante, vaincre – vainquons, vif – vive

2.2 Domain-dependent limitations of stem constancy

In this section, we will examine deviations from stem constancy in inflectional, derivational and compounding structures. We will link the deviation from stem constancy with the boundary strength between the word stem and the inflection or derivation suffix or the following compositional element. As Berg (2013) pointed out, there is a low boundary strength for inflectional products and a relatively high boundary strength for compounds. Derivational products are somewhere in between. Deviations in stem constancy can be caused by inflectional processes and processes of word formation. It can be morpheme-dependent whether the constant spelling of the stem is preserved or whether it is altered. An example is <y> in *copy*, which is changed to <i(e)> before the plural suffix {-s} and before the derivational suffix {-er}, but is preserved with other derivational suffixes {-ist},[2] {-able} and in compounds: *copies, copier, copyist, copyable, copybook*. Fuhrhop et al. (2011) point out that <y> marks word boundaries: the stronger the boundary, the more likely <y> is preserved; the weaker the boundary, the more likely it is to change to <i(e)>. Deviations from stem constancy can thus be indicators of the boundary strength of a morphologically complex word. Such morpheme- or domain-dependent deviations from stem constancy can be found in all languages (to varying degrees) (see Table 4). In English, <y>-replacement and <e>-deletion are among the major morphographic alternations (cf., e.g., Berg 2013; Huddleston & Pullum 2002: 1575ff.). The

2. Especially the derivational forms vary, but these are at least possible spellings.

Table 4. Examples of domain-dependent stem preservations and stem alternations in German, English, Dutch and French

	German	English	Dutch	French
[– stem preservation in inflection]	Saal – Säle	copy – copies, wife – wives, love – loving, shake – shaking	leven – leeft, reizen – reist, schrijven – schrijft, lezen – leest	vaincre – vainquons
[+ stem preservation in inflection]	stellen – stellt, Hund – Hunde, drehen – dreht	boy – boys	melden – meldt	provoquer – provoquons, intriguant – intriguons
[– stem preservation in derivation]	Haar – Härchen	copier, lovable, shakable, happiness	café – cafeetje, taxi – taxietje	provoquer – provocant, intriguant – intrigant
[+ stem preservation in derivation]	Kind – kindlich	loveable, boyish, wifeless	beïnvloedbaar, avondlijk, snijdbaar	
[– stem preservation in compounding]	Kirsche – Kirschbaum		reizen – reistijd, lezen – leesbril	
[+ stem preservation in compounding]	Messstand	shakeout, ladybug, loveletter, copybook, boyfriend	eindpunt, tijdbom, schuldvraag, goedkoop	

replacement of <y> to <i(e)> and the deletion of <e> are morpheme-sensitive and therefore serve to indicate the strength of the morpheme boundary. In inflectional contexts, the spelling of the stem is altered; in compounds or compound-like structures, the spelling of the stem is preserved to indicate a stronger boundary (cf. Berg et al. 2014). For derivational suffixes, the stem spelling is either constantly preserved or altered, depending upon the suffix. In Dutch, there are systematic stem alternations in all domains, but these are not primarily suffix or domain-dependent. Stem alternations in Dutch depend on whether the last syllable of the stem is opened or closed due to inflectional or word formation processes. To keep the pronunciation constant, vowels and consonants are systematically doubled or simplified (cf. De Schryver & Neijt 2012: 117f.). The final devoicing in Dutch is made graphically visible with the <z>/<s>-change and the <v>/<f>-change. To keep the vowel quantity the same in all forms, <é> is doubled to <ee> in *cafeetje* and <i> changes to <ie> in *taxietje*. In French, the complex phoneme-grapheme correspondences make it possible to preserve the stem constancy in inflection in these examples. For the graphical differentiation and marking of a different word type, however, other correspondences are used and the stem spelling is altered. It turns out that the stronger a morphological boundary is, the more likely it is that the stem spelling is preserved.

This applies to compounds and partly to derivational products. The weaker the morphological boundary, the more likely the stem spelling is to be altered, which is often the case in inflection. The domain-dependent analysis of stem alternations with the parameters [± stem preservation in inflection], [± stem preservation in derivation] and [± stem preservation in compounding] thus provide further criteria with the help of which the general strength and extent of stem constancy can be made cross-linguistically comparable.

2.3 Positional localization of deviations

In order to comprehensively record the properties of stem constancy, it is required to determine where deviations in spelling occur within the word stem. Generally speaking, these can occur stem initial, stem internal or stem final, which leads to the possible parameters [± stem initial deviation], [± stem internal deviation] and [± stem final deviation]. A deviation in stem final position, which only concerns one letter, such as the <z>-replacement in Dutch *reizen – (ik) reis* ("travel – (I) travel") or the deletion of the final letter as in English *shake – shaking*, is a comparably small deviation from the stem spelling. With regard to the reading process, a stem internal deviation from stem constancy, which the reader encounters earlier in the reading process than a stem final deviation, is likely to cause greater difficulty for word recognition than stem final changes. Therefore, stem internal deviations can be considered to be more severe deviations from the principle of stem constancy.

As expected, hardly any deviations occur in stem initial position in the corpus data. Of all deviating wordforms in German, only *alt – älter* ("old – older") varies stem initially, in Dutch some double vocal letters in stem initial positions deviate: *oog – ogen* ("eye – eyes"), *uur – uren* ("hour – hours"), *eten – eet* ("eat – (I) eat"), and in French, only the initial accents vary, e.g., *être – était – été* ("to be – was – been"). Analyzing all deviating wordforms in all four languages, initial deviations account for 2%. Stem internal deviations account for 33.5% and consist mostly of vowel changes in strong verbs, e.g., *verlieren – verloren* ("lose – lost"), *give – gave, vind – vond* ("find – found"). Internal plus final deviation in one stem account for 21.8%, e.g., *nehmen – nimmt* ("take – (he) takes"), *feel – felt, blijven – bleef* ("stay – (I) stayed"), *prend – pris* ("take – took"). Deviations in stem final position account for 30.9% and suppletive forms account for 11.8% of all deviating wordforms.[3] Figure 1 gives an overview of the positional distribution of the stem deviations in the four languages:

3. Regarding the corresponding tokens, suppletive wordforms account for most of the deviating tokens (40.8%), then final deviation (24.5%), internal deviation (19%), internal + final deviation (13.4%) and stem initial deviation (2.3%).

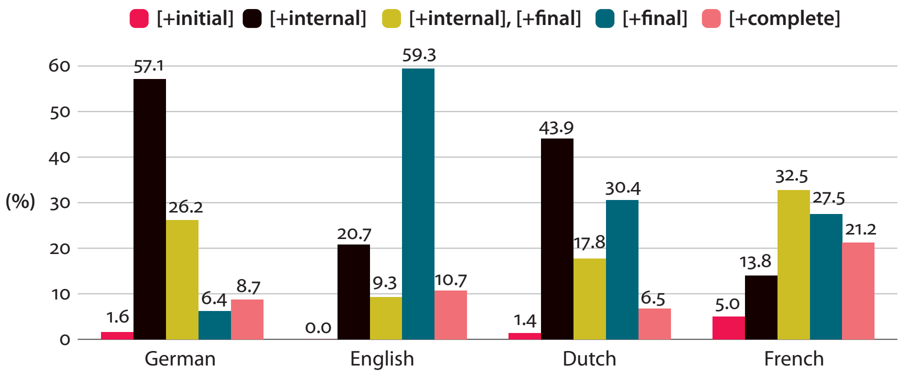


Figure 1. Comparison of the positional distribution of stem deviations in German, English, Dutch and French

It is remarkable that internal deviations are most frequent in German; these are vowel changes in strong verbs. Furthermore, stem final deviations are most frequent in English, including mainly systematic <e>-deletion and <y>-replacement (cf. Berg 2013). The systematicity behind the deviations will be examined in more detail in the following section.

2.4 Systematicity of deviations

Of all deviating types, 80.2% are located among the verbs; of all deviating tokens, 93.3% are located with verbs. The smallest percentage of stem deviation is with nouns (4.9% of all deviating types, 1.7% of all deviating tokens). Now let's have a look at what kind of deviations from stem constancy there are and whether they can be systemized.

German

Counting all types of deviations from stem constancy individually, vowel change is the most frequent type of deviation in German (49.7% of all deviations), e.g., *verlieren* – *verloren* ("lose – lost"), *gilt* – *galt* ("apply to – applied to"), *betreffen* – *betroffen* ("affect – affected"). All stem deviations concerning vowel changes are due to the inflection of strong verbs. The second most common type of deviation is consonant change (18.2%), e.g., *schließen* – *schloss* ("close – closed") and umlauts (17,6%), e.g., *jung* – *jünger* ("young – younger"), *alt* – *älter* ("old – older"). In 7.6% of all deviations double consonant letters are involved, e.g., *treffen* – *traf* ("meet – met"), *kommen* – *kam* ("come – came"). These changes always go along with a vowel change: the pronunciation of the wordform differs completely from the lemma and there is no corresponding two syllable form which could serve as a source for keeping the double consonant spelling from the infinitive form. 7.6% of all deviations are suppletive forms, e.g., *sein – ist – sind* ("be – is – are"). Almost all deviations from stem constancy spelling are due to phonological alternations. Only umlaut changes from <a> to <ä> as in *alt – älter* ("old – older") are morphological deviations from stem constancy, because <e> (*elter) would be a phonographically possible alternative spelling. Thus, German uses [+ graphical resemblance of stems], which is morphologically explainable in 17.6% of all deviating stem spellings. 82.4% of all stem deviations are due to phonological alternations.

English

The three major deviations from stem constancy in English are <y>-replacement (e.g., *city – cities*), <e>-deletion (e.g., *use – using*) and consonant doubling (e.g., *big – bigger*) (cf. Berg 2013; Huddleston & Pullum 2002: 1575ff.). We now take a look at how common the respective deviation types are for the most frequent 1,000 types examined in our corpus study (see Figure 2).

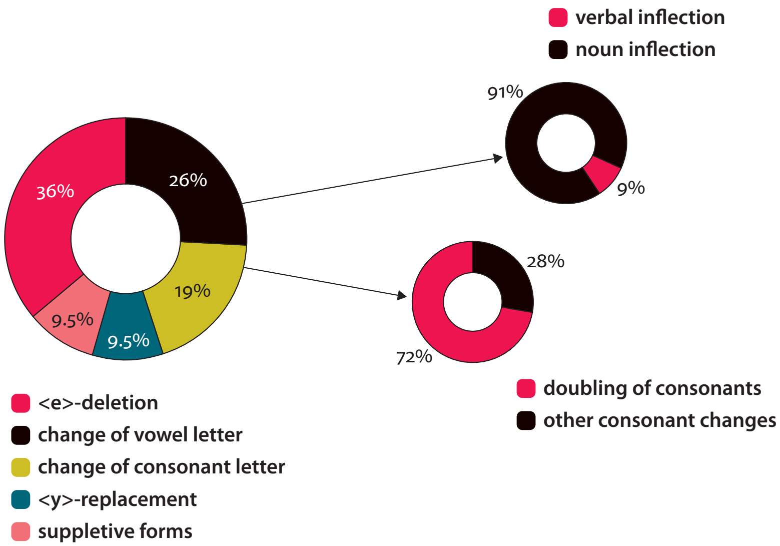


Figure 2. Types of deviations in English stem spelling[4]

The most frequent deviation from stem constancy is final <e>-deletion, as, for example, in *come – coming, use – used* and *survive – survival.* Changes of vowel letters occur mostly in verbal inflection (e.g., *sing – sang – sung*); only a few nouns have vowel changes (e.g., *man – men, foot – feet*). Changes of consonant letters

4. Stems that contain two deviations (e.g., change of consonant letter + <e>-deletion) are counted individually for each type of deviation.

concern on the one hand the doubling of letters (e.g., *big – bigger*), and on the other hand other consonant replacements (e.g., *build – built, make – made, leave – left*). <y>-replacement (*lady – ladies*) and suppletive forms (e.g., *be – am – are – was*) each account for 9.5% of all deviating stem spellings.

Dutch

As mentioned before, in Dutch vowel and consonant letters are systematically doubled or simplified when occurring in different syllable environments (internal position, syllable final position). Of all deviations from stem spelling, the simplification of letters is most frequent (35.3%). The doubling of letters accounts for 21.2% of all deviations from stem constancy. The doubling of letters almost exclusively concerns vowels, and 87% of these occur in verbal inflection, e.g., *nemen – neem, eten – eet*. When the syllable gets closed due to an inflectional suffix, the vowel needs to be doubled to keep the vowel quantity from the infinitive. The simplification of letters concerns 39% vowels and 61% consonants – all of the consonantal simplifications occur in verbal inflection, e.g., *snappen – snap, willen – wil, zeggen – zeg*. To keep the vowel quantity from the infinitive, it is unnecessary to keep the double consonant spelling in many inflectional forms. Furthermore, double consonant spellings in word final position are not allowed in general in Dutch (De Schryver & Neijt 2012; Nunn 1998). Letter changes such as *zien – zag* ("see – saw") *or vind – vond* ("find – found") make up 32.6% of all deviations. These letter changes occur in the inflection of strong verbs. In the cases of <v>/<f>- and <z>/<s>-change, final devoicing is made graphematically visible, e.g., *lezen – lees* ("read – (I) read"), *leven – leef* ("live – (I) live"). This replacement occurs systematically as soon as <v> or <z> would be in syllable final position and accounts for 5.5% of all deviations from stem constancy. Suppletive forms also account for 5.5% of all deviations from stem constancy. Table 5 lists some examples for each type of deviation:

Table 5. Examples for different types of stem deviations in Dutch

Deviation from stem spelling	**Examples**	**Percentage of all deviations**
simplification	rood – rode, week – weken, zeggen – zeg, lukken – luk, bellen – bel	35.3%
doubling	man – mannen, wit – witte, dik – dikke, snel – sneller, plan – plannen, maken – maak, slapen – slaap	21.2%
letter change	hebben – heef, kopen – gekocht, vallen – viel, nemen – nam, vragen – vroeg	32.6%
<v>/<f>- and <z>/<s>-change	blijven – blijf, geven – geef, schrijven – schrijf, lezen – lees, reizen – reis	5.5%
suppletive forms	zijn – ben – is – war, doen – deed	5.5%

In Dutch, the simplification and doubling of letters, as well as the <v>/<f>- and <z>/<s>-change are systematically predictable. These deviations from stem constancy are not arbitrary but can be explained systematically and make up 62% of all deviations. Only letter changes and suppletive forms cannot be derived solely from the spelling of the lemma stem. These spellings are reactions to a different phonetic realization deviating from the lemma stem. They make up 38% of all deviations from stem constancy in Dutch.

French

The comparison of the number of identical stem spellings in Table 1 has shown that 92% of the stems of the wordforms in French are written identically to the lemma stems. Now we take a look at the few deviations from stem constancy and examine if and how they can be systematized.

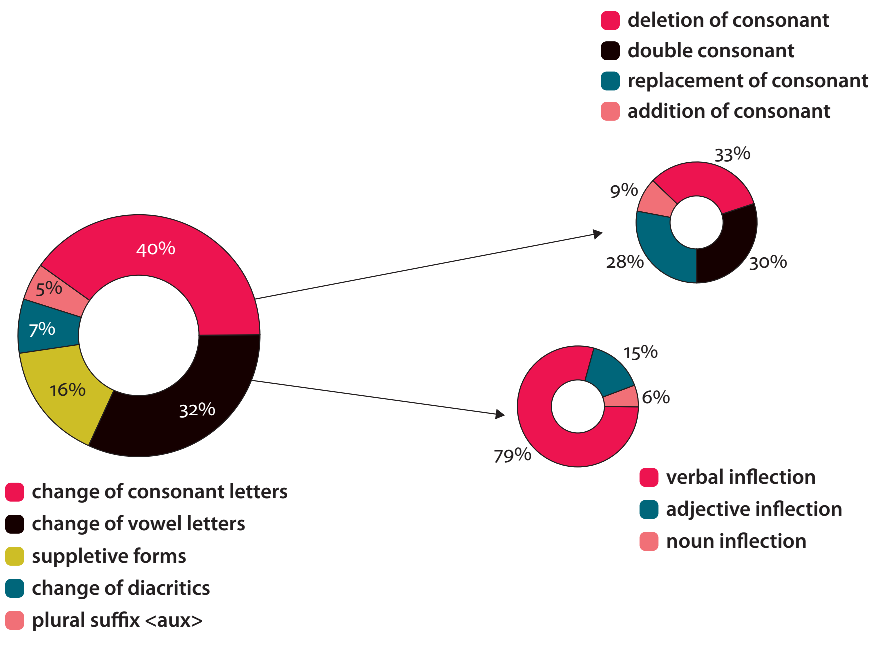


Figure 3. Types of deviations in French stem spelling[5]

Figure 3 shows that there are multiple different forms of deviation from stem constancy. In contrast to the other three languages, French also has deviations due to diacritics alone, e.g., *être* – *étais* – *été* ("to be – was – been"). What also stands out

5. Stems that contain two deviations (e.g., change of consonant letter + change of vowel letter) are counted individually for each type of deviation.

is a special form of plural formation; words that end in <a(i)l> in singular form their plural with <aux>, e.g., *travail – travaux* (“work – works”), *journal – journaux* (“newspaper – newspapers”), *local – locaux* (“place – places”) (cf. Grevisse & Goosse 2008: 677ff.; Catach, 2003: 94). Regarding the consonant changes in detail, it can be seen that the deviations from stem constancy are distributed among the deletion of consonants (e.g., *connaître – connu,* “to know – known”), the replacement of consonants (e.g., *pouvoir – pourrais,* “may – might”), the addition of consonants (e.g., *faire – faisons,* “to make – (we) make”) and double consonants (e.g., *appeler – appelle*, “to call – (I) call”). As with the other languages, the vowel deviations from stem constancy are mainly due to verbal inflection (e.g., *pouvoir – peux,* “may – (I) may”, *prendre – pris*, “to take – taken”). In addition, there are some adjectives (e.g., *nouveau – nouvelle,* “new” masculine – feminine) and a few nouns (e.g., *travail – travaux,* “work” singular – plural) whose stem is altered by a vowel change in different inflectional forms. Although French shows the fewest deviations from stem constancy compared to German, English, and Dutch, the types of deviations are manifold. All deviations in the stems of the wordforms are connected with a pronunciation deviating from the lemma stem. The principle also applies in French that if the phoneme-grapheme correspondences allow it, the stems of a word should always be written identically (Grevisse & Goosse 2008: 86). Homography avoidance strategies, however, usually occur with function words: *la – là* (feminine article – “there”), *ou – où* (“or – where”), *sur – sûr* (“on – sure”) (cf. Catach 2003: 66ff.; Meisenburg 1996: 203f.; Grevisse & Goosse 2008: 108f.).

Types of stem deviations in language comparison

There are deviation types that can be found in all four languages. These include consonant changes and vowel changes. Especially in the case of vowel changes, an average of 92.5% of all deviations can be assigned to verbal inflection. The interesting point about the suppletive forms is that in all four languages they are partly inflectional forms of the same verbs. Forms of *be* alone account for about 94% of all suppletive tokens. Suppletive forms can emerge and persist more easily with frequent wordforms (cf., e.g., Nübling 1999). To take up the double consonant example of the beginning again, it turns out that double consonant spellings are treated differently in the individual languages. In English, the consonant is usually simple when it is in the final position (e.g., *big*); within words it is doubled (e.g., *bigger*). In French double consonant spelling varies and has several functions, e.g., marking of the pronunciation, identification of the feminine form, identification of the diminutive (cf. Catach 2003: 272ff.). Sometimes it shows morphology visibly (e.g., diminutive in *maisonette*) but not always (e.g., *horrible,* “terrible”). In addition, there are language-specific deviations from the stem constancy: the umlaut in German occurs systematically if there is a related form with <a>. The <z>/<s>- and

<v>/<f>-change in Dutch also occurs systematically. <y>-replacement and final <e>-deletion in English are not completely systematic; they are domain-dependent and also suffix-dependent (cf. Berg 2013). In French, diacritics can deviate. Table 6 provides a summarizing overview of the different types of deviation from stem constancy and the respective percentage of each language therein:

Table 6. All deviations sorted according to the deviation type and the respective proportion in the individual languages*

change of vowel, 42%	language-specific deviations, 19%	change of consonant, 17%	double consonants, 14%	suppletive forms, 8%
DUT, 46%	EN, 58%	DUT, 28%	DUT, 64%	FR, 29%
GER, 27%	GER, 21%	FR, 26%	FR, 14%	EN, 28%
EN, 15%	DUT, 11%	GER, 26%	GER, 13%	DUT, 24%
FR, 12%	FR, 10%	EN, 20%	EN, 9%	GER, 19%

* The table reads as follows: vowel changes make up 42% of all deviations from stem constancy in general. Dutch makes up 46% of all vowel changes.

The proportion of Dutch in the vowel changes is composed of 36% vowel doublings, e.g., *maken* – *maak* ("do – (I) do"); 26% vowel simplifications, e.g., *uur* – *uren* ("hour – hours") and 38% vowel changes, e.g., *kiezen* – *kozen* ("choose – chosen").

3. Summary and preview of future research

In this chapter we examined the principle of stem constancy in four languages. We set up possible parameters that can be used to compare the phenomenon of stem constancy across languages. The corpus analysis showed that stem constancy differs in strength in German, English, Dutch and French. There are different types of deviations and these are systematic in different ways.

Stem constancy is highest in French with 92% of stems written identically. However, the types of deviations are manifold. In German, the number of identical stems is 87.4%. Deviations are mostly stem internal and concern vowel changes in the inflection of strong verbs or umlauts (67.3% of all deviations). In English, there are 85% identical stem spellings. Deviations occur most frequently in word final position and often concern <e>-deletion or vowel changes in the inflection of strong verbs. <e>-deletion and <y>-replacement are domain-dependent deviations from stem constancy and are also suffix-dependent (cf. Berg 2013). In Dutch, the amount of [+ identical stem spelling] is lowest with 78.6%. However, the deviations are not domain-dependent, but depend on the syllable structure. Thus, the duplication and simplification of letters and also the changes of <v>/<f> and <z>/<s> can be derived systematically. 62% of all deviations belong to this group.

With the help of the established parameters, we were able to investigate stem constancy more closely and compare the properties and limitations of stem spelling in German, English, Dutch and French. We can use these parameters to evaluate how strong stem constancy is in a language, looking at [± identical stem spelling] and [± optical resemblance of stems]. The next step is to look at the deviations: are deviations domain-dependent ([± stem deviation in inflection/derivation/compounding])? Where do deviations occur ([± stem initial/internal/final deviation])? Are the types of deviation systematic? The formulation and testing of such parameters now allows the investigation of further languages with regards to how stem constancy works in these languages. If stems are written to a high degree identically, these morphological spellings can be reading aids and support the segmentation of complex wordforms (cf. Aronoff et al. 2016). And if spellings indicate more than mere pronunciation – e.g., the constant spelling despite final devoicing (e.g., *Tag* – *Tage*, "day – days") in German or the consistent spelling of the preterite suffix as <ed> in English despite varying pronunciation (e.g., *repeated, begged, looked*) (cf. Carney 1994: 19), morphological information can be made visible and support processing.

Further research could adopt this comparative perspective and utilize the parameters established as the basis for investigating the characteristics and limitations of stem constancy in a broader (typological) scope. In comparative linguistics, two or three languages are usually compared, and similarities and differences are pointed out. Due to this very narrow scope chosen in comparative studies, it is possible to make comprehensive and very detailed language comparisons (cf. König 2012; Eichinger 2012). The aim of language typology, on the other hand, is "the classification of languages or components of languages based on shared formal characteristics" (Whaley 1997: 7). Language typology uses cross-linguistic comparisons with as many languages as possible to classify languages or aspects of languages and to determine formal characteristics of languages (cf. Whaley 1997: 77; Daniel 2011: 44f.). By identifying patterns, regularities and implications, it should be possible to deduce typological universals which could be considered fundamental organizational principles of human languages (cf. Haspelmath et al. 2001: VII–VIII; Velupillai 2012: 16–17). Greenberg formulated some universal principles of human language, as in "if a language has inflection, it has derivation" (1963: 73), to give just one example. However, due to the large number of languages examined, typological comparisons mostly remain on a very general level. Haspelmath (2010) argues that comparative concepts must be used in language typology in order to make cross-linguistic generalizations. These concepts or comparative parameters must be applicable across multiple languages, i.e., they must be based exclusively on simple, universally usable concepts, conceptual-semantic concepts, generally formal concepts and other extralinguistic comparative concepts. The determination of these parameters can only be achieved through small-scale language comparisons. Contrastive studies are therefore a necessary complement of language typology, and this is how we approach stem constancy in this chapter. This language-contrastive approach thus provides a basis for comprehensive typological investigations on stem constancy and typological graphematics in general.

However, some research has already been done on comparative graphematics in recent years. Fuhrhop (2018) demonstrated how a comparative graphematics of German and multiple European languages can be developed. One starting point on the road to developing a comparative typology of graphematics is to examine established writing principles in different languages. This was the approach we have taken in this chapter, investigating stem constancy. Other comparative approaches have already provided fruitful insights into the characteristics of different writing systems and created a path to a typological graphematics. Meisenburg (1996) gives a description of the diachronic writing systems of the Romance languages; Lindqvist (2001) provides a graphematic study of the Scandinavian languages. In addition, there are several language-comparative studies that examine particular graphematic aspects or phenomena concerning characteristics of spelling: Berg & Fuhrhop (2011) cross-linguistically examined the constitution of syllable nuclei and diphthongs. Fuhrhop (2017) investigates the spelling of irregular verbs and their degree of visible morphological information. Further investigating and characterizing phenomena of visual morphology, Fuhrhop (2020) investigates verbal inflection in the Romance languages. Berg (2017) analyzed the morphological coding and decoding of regular verbs, contrasting the unequivocal decoding of English inflectional suffixes with the ambiguity of German inflectional suffixes. Berg (2019) gives a detailed examination of the writing of morphemes in German and English, contrasting, for example, the uniformity and unambiguity of the inflectional and derivational suffixes in German and English. Overall, comparative graphematics is a relatively new research area that has a lot of potential for further investigation. One long-term aim is to establish a typological graphematics through using cross-linguistically applicable parameters which have been set up in comparative studies.

References

Aronoff, Mark, Kristian Berg & Vera Heyer. 2016. Some implications of English spelling for morphological processing. *The Mental Lexicon* 11(2). 164–185. https://doi.org/10.1075/ml.11.2.01aro

Berg, Kristian. 2013. Graphemic alternations in English as a reflex of morphological structure. *Morphology* 23(4). 387–408. https://doi.org/10.1007/s11525-013-9229-1

Berg, Kristian. 2017. Sichtbare Flexionsmorphologie im Englischen und Deutschen: Der Umgang mit Variation in der Schreibung von Stämmen und Affixen. In Nanna Fuhrhop, Renata Szczepaniak & Karsten Schmidt (eds.), *Sichtbare und hörbare Morphologie*, 9–42. Berlin: De Gruyter. https://doi.org/10.1515/9783110528978-002

Berg, Kristian. 2019. *Die Graphematik der Morpheme im Deutschen und Englischen*. Berlin: De Gruyter. https://doi.org/10.1515/9783110604856

Berg, Kristian, Franziska Buchmann, Katharina Dybiec & Nanna Fuhrhop. 2014. Morphological spellings in English. *Written Language and Literacy* 17(2). 282–307. https://doi.org/10.1075/wll.17.2.05ber

Berg, Kristian & Nanna Fuhrhop. 2011. Komplexe Silbenkernschreibungen im Englischen im Vergleich mit dem Deutschen. *Linguistische Berichte* 228. 443–466.

BNC (2007). *The British National Corpus, version 3 (BNC XML Edition)*. BNC Consortium, Bodleian Libraries, University of Oxford. Available online at www.natcorp.ox.ac.uk

Carney, Edward. 1994. *A survey of English spelling*. London: Routledge.

Catach, Nina. 2003. *L'orthographe française*, 3rd edn. Paris: Nathan.

CLARIN.SI (2017). *Common Language Resources and Technology Infrastructure/ Le Monde Corpus (Release 28.10.2017)*. Ljubljana: Jožef Stefan Institute. Available online at https://www.clarin.si/noske/

Cummings, Donald W. 1988. *American English spelling: An informal description*. Baltimore, MD: Johns Hopkins University Press.

Daniel, Michael. 2011. Linguistic typology and the study of language. In Jae J. Song (ed.), *The Oxford handbook of linguistic typology*, 43–68. Oxford: Oxford University Press.

De Schryver, Johan & Anneke Neijt. 2012. *Handboek spelling*. Mechelen: Plantyn.

Dürscheid, Christa. 2016. *Einführung in die Schriftlinguistik*, 5th edn. Göttingen: Vandenhoek & Ruprecht.

Eichinger, Ludwig M. 2012. Deutsch im Sprachvergleich: Grammatische Kontraste und Konvergenzen. In Lutz Gunkel & Gisela Zifonun (eds.), *Deutsch im Sprachvergleich: grammatische Kontraste und Konvergenzen*, vvi–xvi. Berlin: De Gruyter. https://doi.org/10.1515/9783110284768.vii

Eisenberg, Peter. 2016. Der Buchstabe und die Schriftstruktur des Wortes. In Angelika Wöllstein (ed.), *Duden – die Grammatik: Unentbehrlich für richtiges Deutsch*, 9th edn., 61–94. Berlin: Dudenverlag.

Fuhrhop, Nanna. 2020. Visible verbal morphology: Morpheme constancy in Germanic and Romance verbal inflection. *Morphology*. https://doi.org/10.1007/s11525-020-09372-4

Fuhrhop, Nanna. 2017. Sichtbare Morphologie in der Flexion der starken und unregelmäßigen Verben im Deutschen und Englischen. In Nanna Fuhrhop, Renata Szczepaniak & Karsten Schmidt (eds.), *Sichtbare und hörbare Morphologie*, 43–76. Berlin: De Gruyter. https://doi.org/10.1515/9783110528978-003

Fuhrhop, Nanna. 2018. Graphematik des Deutschen im europäischen Vergleich. In Angelika Wöllstein, Mechthild Habermann, Manfred Krifka & Peter Gallmann (eds.), *Grammatiktheorie und Empirie in der germanistischen Linguistik*, 587–616. Berlin: De Gruyter. https://doi.org/10.1515/9783110490992-020

Fuhrhop, Nanna, Franziska Buchmann & Kristian Berg. 2011. The length hierarchy and the graphematic syllable: Evidence from German and English. *Written Language and Literacy* 14(2). 275–292. https://doi.org/10.1075/wll.14.2.05fuh

Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of language*, 58–90. Cambridge, MA: MIT Press.

Grevisse, Maurice & André Goosse. 2008. *Le bon usage: Grammaire française*, 14th edn. Brüssel: De Boeck & Larcier.

Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3). 663–687. https://doi.org/10.1353/lan.2010.0021

Haspelmath, Martin, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.). 2001. *Language typology and language universals: An international handbook*, Volume 1. Berlin: De Gruyter. https://doi.org/10.1515/9783110171549.2

Huddleston, Rodney D. & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316423530

IDS (2014). *Deutsches Referenzkorpus/ Archiv der Korpora geschriebener Gegenwartssprache 2014-I (Release 15.04.2014)*. Mannheim: Institut für deutsche Sprache. Available online at www.ids-mannheim.de/DeReKo

König, Ekkehard. 2012. Zur Standortbestimmung der Kontrastiven Linguistik innerhalb der vergleichenden Sprachwissenschaft. In Lutz Gunkel & Gisela Zifonun (eds.), *Deutsch im Sprachvergleich: grammatische Kontraste und Konvergenzen*, 13–40. Berlin: De Gruyter. https://doi.org/10.1515/9783110284768.13

Lindqvist, Christer. 2001. *Skandinavische Schriftsysteme im Vergleich*. Tübingen: Niemeyer. https://doi.org/10.1515/9783110927085

Meisenburg, Trudel. 1996. *Romanische Schriftsysteme im Vergleich: Eine diachrone Studie*. Tübingen: Narr.

Neef, Martin. 2013. Das Konzept des morphologischen Prinzips und seine Rolle in einer modularen Schriftsystemtheorie. In Martin Neef & Carmen Scherer (eds.), *Die Schnittstelle von Morphologie und geschriebener Sprache*, 9–38. Berlin: De Gruyter. https://doi.org/10.1515/9783110336665.9

Nübling, Damaris. 1999. Zur Funktionalität von Suppletion. In Matthias Butt (ed.), *Variation und Stabilität in der Wortstruktur: Untersuchungen zu Entwicklung, Erwerb und Varietäten des Deutschen und anderer Sprachen*, 77–101. Hildesheim: Olms.

Nunn, Anneke M. 1998. *Dutch orthography*. Den Haag: Holland Academic Graphics.

Oostdijk, Nelleke, Martin Reynaert, Véronique Hoste & Ineke Schuurman. 2013. The construction of a 500-million-word reference corpus of contemporary written Dutch. In Peter Spyns & Jan Odijk (eds.), *Essential speech and language technology for Dutch: Theory and applications of natural language processing*, 219–247. Berlin & Hildesheim: Springer. https://doi.org/10.1007/978-3-642-30910-6_13

Velupillai, Viveka. 2012. *An introduction to linguistic typology*. Amsterdam: John Benjamins. https://doi.org/10.1075/z.176

Whaley, Lindsay J. 1997. *Introduction to typology: The unity and diversity of language*. Thousand Oaks, CA: Sage Publications. https://doi.org/10.4135/9781452233437

CHAPTER 7

Major lexical categories and graphemic weight

Kristian Berg
Universität Bonn

English spelling has a very interesting regularity: there exists a minimum word length for lexical words. Words of this class have to be at least three letters long, even if they consist of only two phonemes and could be spelled with two letters (e.g., *ebb*/*eb, *egg*/*eg). This regularity does not hold for function words (e.g., *a*, *I*, *be*, *he*, *it*, etc.). This means that an important distinction between words (lexical vs. functional) is mirrored in the writing system. In a study of Early Modern English texts, I demonstrate that this regularity evolved gradually over the course of 200 years. This is a case of self-organization in spelling: without explicit guidance or regulation, the pattern emerged in usage. The proposed function of having separate constraints for the length of lexical words and function words is a reading aid.

Keywords: English, lexical words, grammatical words, spelling, self-organization

1. How did we get here?

English spelling is morphographic to some extent. This has been demonstrated for a number of phenomena: for example, the alternation between ⟨-our⟩, ⟨-or⟩, and ⟨-er⟩[1] in British English spelling reflects subtle morphological differences (Aronoff 1978). The form ⟨our⟩ is mostly found with bound bases, and the nouns are mostly inanimate (e.g., *colour*, *favour*); ⟨-or⟩ is a deverbal suffix which denotes animate nouns (e.g., *mediator*, *oppressor*). This distinction is not mirrored in phonology.

Or take the spelling of English suffixes and word endings (Berg & Aronoff 2017, 2018): we have shown that English spellings became both more uniform (variation in the spelling of word forms was reduced) and more unique (suffixes tend to have distinct spellings that set them apart from homophonic word endings – e.g., *nervous* vs. *service*) over time (Berg & Aronoff 2018).

In what follows, I present diachronic evidence for a regularity in the English writing system, the so-called three letter rule. This rule, too, is an instance of

1. I use angled brackets to highlight actual spelling.

https://doi.org/10.1075/cilt.353.07ber

morphographic spelling, albeit a slightly different one. It states that lexical words in English must have a certain minimal length; function words can be shorter. I illustrate the rule and its exceptions (§ 2), speculate about a potential utilization for readers (§ 3), and then trace the spelling of two lexical words and two function words through time (§ 4 and § 5). I show that the three letter rule emerged as a result of another spelling change, namely the transition from complex word final patterns like ⟨dde⟩ (as in ⟨madde⟩) to single letter spellings (as in ⟨mad⟩). I then discuss the findings (§ 6).

2. The three letter rule

The three letter rule concerns the English lexicon, and thus a central part of morphology. It states that English content words must have a minimal length of three letters – even in cases where the regular phoneme-grapheme correspondences would allow for a two letter spelling, as illustrated in (1)

(1) *add, ebb, egg, inn, odd*

This regularity has often been observed in the past. One of the most prominent mentions is in Jespersen's *Modern English Grammar*:

> Another orthographic rule was the tendency to avoid too short words. Words of one or two letters were not allowed, except a few constantly recurring (chiefly grammatical) words: *a*, *I*, *am*, *an*, *on*, *at*, *it*, *us*, *is*, *or*, *up*, *if*, *of*, *be*, *he*, *me*, *we*, *ye*, *do*, *go*, *lo*, *no*, *so*, *to*, (*wo* or *woe*), *by*, *my*. To all other words that would regularly have been written with two letters, a third was added, either a consonant, as in *ebb*, *add*, *egg*, *Ann*, *inn*, *err* – the only instances of final *bb*, *dd*, *gg*, *nn* and *rr* in the language, if we except the echoisms *burr*, *purr*, and *whirr* – or else an *e* […]: *see*, *doe*, *foe*, *roe*, *toe*, *die*, *lie*, *tie*, *vie*, *rye*, (*bye*, *eye*), *cue*, *due*, *rue*, *sue*. (Jespersen 1909: 149)

As Jespersen shows, the three letter rule only holds for lexical words; function words regularly violate it. Lexical words in English have to be of a certain length, and they are "usually bulked up to a minimum of three letters" (Carney 1997: 76). One function may be to avoid homography, as Cook (2014: 70) suggests: "A spoken content word that would logically correspond to two letters is padded out with an extra consonant to avoid confusion with two-letter function words". Note the similar metaphors used in both citations (*bulk up*, *pad out*) – the idea is that words like *ebb* and *odd* are made longer than they have to be.[2] Historically, that is not what happened (see § 5).

2. I would like to thank Rebecca Treiman for this observation.

A more careful inspection reveals that what is at stake here is not length, but syllable weight, as Evertz (2018) shows. The main argument is that even three letters are sometimes not enough, as in the cases in (2):

(2) *tre, *sho, *blu, *bla

Apparently, the rime of the graphemic syllable needs to consist of either two vowel letters (*tree*, *blue*) or of a vowel letter and a consonant letter (*show*, *blah*). Borrowing established terminology from non-linear phonology, we can say that the rime of monosyllabic lexical words must be associated with at least two (graphemic) moras. Each letter in the rime is associated with one mora. This is the first well-formedness constraint for lexical words, which I will call the 'rime constraint'. The second one is that lexical words consist of at least three letters (spellings like ⟨eb⟩ and ⟨eg⟩ comply with the first constraint but are still not heavy enough); I will call this the 'length constraint'.

In today's English, there are only a few exceptions to this regularity (Berg 2019: 162ff.). The CELEX database (which contains 52,447 English lemmas, Baayen, Piepenbrock & Gulikers 1995) lists the following two-letter words:

(3) a. *go, ox, ex*
b. *mu, pi*
c. *do, re, mi, fa, so, la, ti, do, si*
d. *ma, pa, mo, po*

Among these words, only the ones in (3a) are true exceptions to the graphemic weight minimum for lexical words.[3] The words in (3b) and (3c) are of foreign origin (they are Greek letter names and syllable names for the solmization), the ones in (3d) are all shortenings, with *mo* for American English *momentum* or British English *moment*, and *po* for British English *[chamber] pot*, i.e., results of a rather minor word formation type. Not included are *ad* (a shortening), the interjection *hi*, and potentially more counterexamples; however, I assume (a) that the number of such counterexamples is very limited and (b) that they are either the result of minor word formation types or belong to minor lexical classes – i.e., their morphology is special, and so is their spelling.

Thus if we count liberally, there are three exceptions to the minimal weight restriction for lexical words. On the other hand, there are 20 function words among the two-letter words:

3. Note, however, that *go* can be used as a function word to mark future tense, and *ex* started out as a Latinate combining form. Also note that the spelling variant *ax* is not included in CELEX. It is remarkable that *ox*, *ex*, and *ax* all end with ⟨x⟩; it may be desirable to describe ⟨x⟩ as a complex grapheme like ⟨y⟩ (see below).

(4) *an, as, at, id, if, in, it, of, on, or, up, us, be, by, he, me, my, no, to, we*

With *am* and *is*, there are actually more two letter 'word forms' of this type (which do not appear in the results because the CELEX search was for lemmas).

There could be potentially more violations of the weight restriction, however. The list in (5) contains all 57 lemmas in CELEX that consist of two phonemes and that could be spelled with two letters, but which are not:

(5) *add, ass, ebb, eff, egg, ill, inn, odd, off, ore, bee, bow, buy, cue, die, doe, dough, due, dye, fee, foe, ghee, hie, high, hue, key, knee, know, lea, lee, lie, low, lye, mow, nigh, pea, pee, pie, roe, row, rue, rye, sea, see, show, sigh, sow, sue, tea, tee, tie, toe, tow, vie, wee, woe, zee*

For some of these words, the two letter spelling exists, but as a homophonous function word (*in*, *be*, *or*, *by*, *no*). All these cases (except for the only function word in (5), *off*) are evidence for the efficacy of the minimal weight restrictions for lexical words in English.

There are more exceptions for three letter words with a single final vowel in CELEX:

(6) a. *chi, gnu, phi, ski, spa*
b. *cry, dry, fly, fry, ply, pry, shy, sky, sly, spry, spy, sty, try, wry*

The words in (6a) are (relatively) recent borrowings; their graphemic structure is special, and so is their status in the lexicon. The words in (6b), on the other hand, are inconspicuous in this regard; they are (for the most part) frequent English words. What sets them apart from other English words is that they end with single final ⟨y⟩. If we want to maintain our claim about the minimal weight of English lexical words, we have to either allow for the exception of ⟨y⟩, or claim that ⟨y⟩ is heavier than the vowel letters ⟨a, e, i, o, u⟩. Either way, the set of exceptions is limited to two very specific types, recent borrowings and words that end with ⟨y⟩.

English content words must fulfill certain weight constraints, while English function words can be graphemically lighter (by the way, the spelling of German words is parallel in this regard). What is the use of this peculiar regularity? In the following section, I will argue that it can potentially be utilized as a reading aid.

3. Why this regularity?

Minimal weight constraints that only hold for lexical words lead to a set of function words that are 'lighter' than lexical words can be. As such, they can be easily identified by the reader. This potentially helps readers with the parsing of sentences: it tells them which parts are purely grammatical, like *the* and *in*, and where to expect lexical content.

To get a better impression of the number of function words that can be identified at first glance, take two short texts from Alice Munro and John Le Carré, and note the words that violate the length and weight restrictions (highlighted).

> Alfrida. **My** father called her Freddie. **The two of** them were first cousins and lived **on** adjoining farms and then for **a** while **in the** same house. One day they were out **in the** fields **of** stubble playing with **my** father's dog, whose name was Mack. That day **the** sun shone, but did not melt **the** ice **in the** furrows. They stomped **on the** ice and enjoyed its crackle underfoot. How could **she** remember **a** thing like that? **my** father said. **She** made **it up**, **she** said. (Alice Munro, *Family Furnishings*)

> Leamas was not **a** reflective man and not **a** particularly philosophical one. **He** knew **he** was written off – **it** was **a** fact **of** life which **he** would henceforth live with, **as a** man must live with cancer **or** imprisonment. **He** knew there was **no** kind **of** preparation which could have bridged **the** gap between then and now. **He** met failure **as** one day **he** would probably meet death, with cynical resentment and **the** courage **of a** solitary. **He**'d lasted longer than most; now **he** was beaten. **It is** said **a** dog lives **as** long **as** its teeth; metaphorically, Leamas' teeth had been drawn; and **it** was Mundt **who** had drawn them.
> (John Le Carré, *The Spy Who Came in from the Cold*)

A considerable number of running words in texts can be identified as function words on purely graphemic grounds, i.e., without the need for lexical access (29% in *Family Furnishings*, 27% in *The Spy Who Came in from the Cold*). We know that readers often skip function words, but rarely lexical words (Dehaene 2009: 17); their length and graphemic makeup can be used as cues to gauge the position of the next gaze fixation (and accordingly, the length of the saccade).

Of course, not all function words can be formally distinguished from lexical words (e.g., *and*, *from*, *its*, *about*, *you*, *there*). However, of the ten most frequent lemmas in the Corpus of Contemporary American English (CoCA, https://www.english-corpora.org/coca/), eight are lighter than lexical words (*the*, *be*,[4] *of*, *a*, *in*, *to* (infinitive marker), *to* (preposition), *it*).

As with every feature of the writing system that seems to be functional, the question is how it evolved. For the regular correspondences between affix spelling and morphological information, we showed that it evolved gradually in a process of self-organization (Berg & Aronoff 2017). How did the weight minimum for lexical words come about? The remainder of this chapter sets out to answer this question.

4. The forms *be*, *am*, and *is* are lighter than lexical words; the forms *are*, *was*, and *were* are not.

4. Methodology

The data base for this investigation is the large diachronic corpus Early English Books Online (EEBO, https://textcreationpartnership.org/tcp-texts/eebo-tcp-early-english-books-online/). It contains more than 25,000 printed texts between 1500 and 1700, totaling up to 872 million running words. The corpus is neither orthographically normalized nor tagged, and thus it is impossible to search for abstract graphemic patterns (like "consists of a vowel and a double consonant") or parts of speech. As a proxy, I will instead concentrate on two pairs of words that behave differently in today's English, *ebb* vs. *web* and *egg* vs. *leg*. As witnessed by *web* and *leg*, stem final /eC/ can be spelled ⟨eC⟩. However, in biphonemic words like *egg* and *ebb*, this is not the actual spelling. The hypothesis is that the spellings ⟨eg⟩ and ⟨eb⟩ are too light; they violate the length constraint.

To obtain the raw data, I use the Kontext interface (https://kontext.korpus.cz) and perform searches for the following different spellings of *ebb*, *web*, *egg*, and *leg*:

(7) a. *eb, ebb, ebbe*
b. *web, webb, webbe*
c. *eg, egg, egge*
d. *leg, legg, legge*

The results are then annotated for false positives – entries from EEBO with one of the word forms in (7) that are not instances of the lexemes *ebb*, *web*, *egg*, and *leg*, like the following examples:

(8) a. *Wherfore that which is contained vnder the lines AE and* ***EB*** *in commensurable to that which is contained vnder the lines…* (Euclid, 1570)
b. *The time is whilſt Thomas* ***Webbe*** *is Preaching in the publique Congregation…* (Stokes, Edward, Esq., 1653)
c. *The Subſtilar line* ***EG*** *being aw the line XW through the center C…* (Aspley, John, 1662)
d. *Colonell Robert* ***Legge****, Governour of the Town, …* (Vicars, John, 1646)

After filtering these entries, there were 30,454 word forms left, which are analyzed in the next section.

5. Results

The 30,454 forms are distributed as follows among the different lexemes and spelling forms (see Table 1):

Table 1. Absolute (and relative) frequencies of the words *ebb*, *web*, *egg*, and *leg* in three different spellings in the EEBO corpus

	⟨C⟩	⟨CC⟩	⟨CCe⟩	total
ebb	122 (3%)	1558 (41%)	2113 (56%)	3793
web	2519 (72%)	192 (6%)	763 (22%)	3474
egg	249 (2%)	4435 (44%)	5428 (54%)	10112
leg	8106 (62%)	1280 (10%)	3689 (28%)	13075

The total frequencies of *ebb* are in the same order of magnitude as those of *web*, and the same holds for the *egg*/*leg* pair. As for the form variants, the simple coda spelling is only a minor variant for *ebb* and *egg*, while it is the dominant variant for *web* and *leg*.

The more interesting question, of course, concerns the development over time. Figure 1 shows the ratio of today's form among all spellings for the four lexemes (there are no data for *egg* and *web* in 1510, hence the white space).

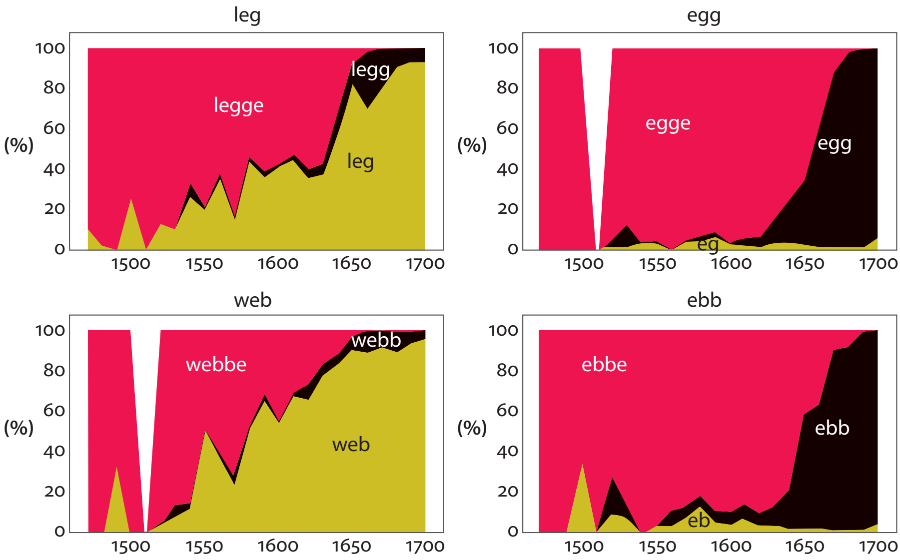


Figure 1. Relative number of word final spelling patterns ⟨CCe⟩, ⟨CC⟩, and ⟨C⟩ among all *leg*/*ebb*/*web*/*egg*-spellings in the EEBO corpus

For *leg* and *web*, we see a gradual transition from ⟨legge⟩/⟨webbe⟩ to ⟨leg⟩/⟨web⟩ between 1500 and 1700. The relatively long duration of this transition makes it improbable that it was governed by spelling authorities; it is more likely an instance of self-organization. The spellings with double final consonants (⟨legg⟩/⟨webb⟩) appear in the 17th century, but they remain marginal for both lexemes and never really catch on. Instead, as Aronoff (p.c.) observed, this pattern is regularly used in the spelling of surnames that would otherwise be homographic with content words: *Webb*, *Kidd*, *Scott*, *Crabb*, *Robb*, *Trapp*.

Interestingly, there is a considerable amount of variation within texts. The variation we see on a global scale (e.g., in Figure 1) comes about (at least in part) because writers use both forms in their texts, as the examples from *A RELATION Of Two several VOYAGES Made into the EAST-INDIES* (by Christoph Frick, written 1700) show:

(9) *[…] for when the others see one of their company thus caught by the* ***Legg****, they won't come near the Coco-Nuts all that day* (p. 45)

(10) *[…] I was immediately shot in the* ***Leg*** *with one Arrow, and with another in my Thigh* (p. 312)

Between 1550 and 1640, in 59% of the texts in which the spelling of *leg* may vary (i.e., texts which contain more than one form of the lexeme), it does actually vary: 59% of these texts are inconsistent with regard to the spelling of *leg*. For *web*, the ratio is 53%. That means that the transition from one form (⟨legge⟩) to another (⟨leg⟩) did not proceed along the lines of texts (with each author consistently using their favorite spelling), but that it involved variation within texts as well. This is a situation we expect to see in a self-organizing system.

For *ebb* and *egg*, the transition is much quicker, and it only starts when *leg*/*web* have mostly switched to their modern forms (around 1640). Here we go from ⟨egge⟩/⟨ebbe⟩ to ⟨egg⟩/⟨ebb⟩. The simple coda spelling ⟨eg⟩/⟨eb⟩ was never a serious contender. There was a short period of time where *leg* and *web* were already mostly spelled ⟨leg⟩, ⟨web⟩, but *ebb* and *egg* were still spelled ⟨egge⟩ and ⟨ebbe⟩.

Why did these spellings have to change at all? After all, ⟨egge⟩ and ⟨ebbe⟩ fulfill both minimal word constraints. My hunch is that double consonant + ⟨e⟩ spellings (as in ⟨legge⟩ and ⟨webbe⟩) were soon regarded as markers for the old way of spelling, the cumbersome, baroque way, which was replaced by the fresh new phonographic simplicity of ⟨leg⟩ and ⟨web⟩. Similarly, the variants ⟨egge⟩/⟨ebbe⟩, I suppose, looked out of fashion and thus had to change. And apparently, ⟨eg⟩/⟨eb⟩ (analogous to ⟨leg⟩/⟨web⟩) was not an option because it left too little graphemic substance.

Both pairs *egg*/*ebb* and *leg*/*web* correlate closely with respect to the ratio of new forms over time (Pearson's r (*ebb*/*egg*): 0.98, $P < 0.001$; Pearson's r (*web*/*leg*): 0.88, $P < 0.001$). That means even though I only investigated two lexemes per category, these two words show a high degree of similarity when it comes to the transition from one form to another, and it is likely that a third lexeme of the same pattern will behave in a similar way.

6. Discussion

In today's English, *leg* and *web* show a pattern that is distinct from phonologically similar words like *egg* and *ebb*: the spellings ⟨leg⟩ and ⟨web⟩ are phonographic spellings, while ⟨egg⟩ and ⟨ebb⟩ are not. The double consonants are usually explicated with reference to a graphemic minimum that English content words must fulfill. As noted in § 2, this is often phrased in terms of 'padding' or 'bulking up'. But historically, we find the opposite development: *egg* and *ebb* were dominantly spelled ⟨egge⟩ and ⟨ebbe⟩ until around 1650. After that, they dropped their final ⟨e⟩. These spellings were thus abridged, not padded – it is only because other words were abridged even more that we see the difference between the groups of lexemes today. We can thus pin down the origin of minimal word constraints in English to the second half of the seventeenth century.

Minimal word constraints are of course not exclusive to written English; there are analogous phonological (prosodic) constraints as well. In spoken English, lexical words are minimally prosodic words, which consist of at least one foot with at least one syllable with at least two moras (see, e.g., McCarthy & Prince 1994, Selkirk 1996). Function words, on the other hand, are often realized as prosodic clitics (Selkirk 1996). But the reflexes of this lexical difference are not isomorphous: in spoken English, function words like *at* regularly appear in reduced/weak forms – but in phrase-final position, they are never reduced (*What did you look át yesterday?*, Selkirk 1996: 200). In written English, the difference is categorical: there is no formal variation. Thus even though both written and spoken language are sensitive to the distinction between function words and lexical words, we cannot derive one from the other because the actual implementation differs.

The same holds for the potential use of the respective minimal word constraints. I suggested above that minimal words in writing could help provide the reader with cues to lexical access, and to potential points of gaze fixation. Selkirk (1996: 187) hypothesizes that the prosodic differences between function words and lexical words could be "exploited by the language learner" whose "first order of business is which words are functional and which are lexical" – a very different function for the same contrast between lexical words and function words we find in speech and writing. This is not surprising: the same grammatical difference has different reflexes in the spoken and written variety, and with them vary the potential uses for the reader/hearer. A simple derivation of written from spoken language cannot do justice to this complex relationship. The English writing system, just as its phonology, is sensitive to this lexical distinction, and spelling thus involves linguistic analysis.

Acknowledgements

I would like to thank the editors of this volume for valuable comments on an earlier version of this chapter, and I would like to thank Mark Aronoff for discussing the three letter regularity regularly with me. On a personal note, Mark is one of the kindest and most generous colleagues I know, fueled by the desire to understand how language works, and never dogmatic. He is truly Jacobsonian in the scope of his interests – nothing linguistic is alien to him. Speaking of Jacobson, I have always wondered whether Mark's interest in writing was (at least in part) in some way attributable to the Prague circle (with his mentor Morris Halle as the proxy). Thank you, Mark!

References

Aronoff, Mark. 1978. An English spelling convention. *Linguistic Inquiry* 9. 299–303.

Baayen, R. Harald, Richard Piepenbrock, & Leon Gulikers. 1995. *The CELEX lexical database (CD-ROM)*. Linguistic Data Consortium. Philadelphia, PA: University of Pennsylvania.

Berg, Kristian. 2019. *Die Graphematik der Morpheme im Deutschen und Englischen*. Berlin: De Gruyter. https://doi.org/10.1515/9783110604856

Berg, Kristian & Mark Aronoff. 2017. Self-organization in the spelling of English suffixes: The emergence of culture out of anarchy. *Language* 93. 37–64. https://doi.org/10.1353/lan.2017.0000

Berg, Kristian & Mark Aronoff. 2018. Further evidence for self-organization in English spelling. *Language* 94(1). e48–e53. https://doi.org/10.1353/lan.2018.0013

Carney, Edward. 1997. *English spelling. Language workbooks*. London: Routledge. https://doi.org/10.4324/9780203318935

Cook, Vivian. 2014. *The English writing system*. London: Taylor & Francis. https://doi.org/10.4324/9780203774236

Dehaene, Stanislas. 2009. *Reading in the brain: The new science of how we read*. New York: Penguin.

Evertz, Martin. 2018. *Visual prosody: The graphematic foot in English and German*. Berlin: De Gruyter. https://doi.org/10.1515/9783110583441

Jespersen, Otto. 1909. *A modern English grammar: On historical principles. Part 1: Sounds and spelling*. Oxon: Routledge.

McCarthy, John & Alan Prince. 1994. Prosodic morphology. In John A. Goldsmith (ed.), *A handbook of phonological theory*, 318–366. Cambridge, MA: Blackwell.

Selkirk, Elisabeth. 1996. The prosodic structure of function words. In James L. Morgan & Katherine Demuth (eds.), *Signal to syntax. Bootstrapping from speech to grammar in early acquisition*, 187–213. New York: Erlbaum.

CHAPTER 8

Word formation in the brain

Data from aphasia and related disorders

Carlo Semenza
University of Padova

Studies in aphasia provide important information about how the brain may represent and process morphologically complex words. The main morphological processes (inflection, derivation and compounding) uncover a fine-grained brain organization. The study of errors, in aphasic syndromes and other disorders like unilateral spatial neglect, clearly determined by failures in morphological processing, has made it possible to clarify several different questions. Evidence was found for stems and affixes to be separately represented. Moreover, evidence for decomposition in processing has also been provided, favoring, however, dual route hypotheses. Information about morphology, it was shown, could persist in absence of the ability to retrieve full phonological forms. Headedness was shown to have psychological reality and neural underpinnings.

Keywords: word formation, aphasia, neglect dyslexia, inflection, derivation, compounding

1. Introduction

The linguistic performance of healthy speakers is opaque to external scrutiny. In contrast, the performance of an aphasic individual consists in most cases of specific patterns of preservation and disruption of linguistic capacities that are believed to reflect the organization of the system. Aphasic conditions thus may offer the researcher an advantageous window from which to observe the working of the language processing system. Indeed, the behavioral consequences of brain damage are independent of the observer's expectations: one may occasionally observe phenomena that are not consistent with available theoretical models. Thus, findings in aphasia, as well as in neuropsychological syndromes in general, whether accidental or resulting from careful experiments, have a heuristic value (Semenza, Bisiacchi & Rosenthal 1988). They might help in adjudicating among different alternatives.

https://doi.org/10.1075/cilt.353.08sem

Aphasiological findings may be compatible with theory A but not with theory B, irrespective of what the observer/experimenter had in mind. Or, even more importantly, aphasiological findings may not be compatible with any of the existing theories, thus forcing the construction of new theories and directing further research in unexpected directions.

Aphasia research must be complemented, in search of converging evidence, with research conducted with other means (e.g., neuroimaging, electroencephalography, eye movement tracking, TMS, etc.). The use of these tools has been growing exponentially in recent years, almost obscuring clinical research. A full description of findings with instrumental techniques is beyond the scope of this chapter. The reader is, however, referred to the extensive review of Leminen et al. (2019), for an update on non-clinical studies. The heuristic merit of observations in aphasia remains invaluable, however.

In the domain of word-formation one can find some of the best examples of the integration of neuropsychological and theoretical (psycho)linguistic models (Allen & Badecker 2001). Indeed research in aphasia provides the best support for the view that knowledge of the component parts of a word, and the rules for combining them, is independent of the knowledge of the word itself (Halle 1973).

2. Neuropsychological syndromes

Neuropsychological syndromes that so far provided information about morphology encompass some aphasic conditions and a peculiar attentional disorder known as 'neglect dyslexia'. In neglect dyslexia, which generally results from a posterior lesion of the right hemisphere, readers ignore the left side of words or sentences. However, it has been repeatedly shown that the syndrome is not merely determined by spatial factors, but is also sensitive to linguistic dimensions. Thus, as reported later in this chapter, neglect of parts of words and sentences would not simply concern what is on the left side of a midline, but tend to respect linguistically significant boundaries.

In aphasia, stems and affixes may be affected independently. Affixation can in fact be selectively impaired in absence of disturbances of the stem (e.g., Goodglass et al. 1972; Miceli et al. 1983). The most important condition where this happens is called 'agrammatism'. Agrammatism is frequently found in non-fluent aphasias due to lesions in the anterior areas of the left hemisphere, typically, but not always, involving Broca's area, traditionally limited to the pars triangularis and the pars opercularis of the lower frontal gyrus. By definition, agrammatism is a disorder that affects the production of free-standing grammatical morphemes (e.g., articles, prepositions, auxiliary verbs, clitics, etc.) and, importantly, affects the production of bound morphemes but not their hosts. Traditionally, agrammatism is contrasted with the so-called ‘paragrammatism’, which is found in the context of fluent aphasias due to lesions to the posterior areas of the left hemisphere, and also affects the use of grammatical morphemes. The distinction between the two classes of phenomena, agrammatism and paragrammatism, is admittedly blurry and somewhat arbitrary. As a general rule, although in agrammatism grammatical morphemes are mostly omitted, in paragrammatism grammatical morphemes are mostly substituted.

One simple explanation for the omission of grammatical morphemes in agrammatism is that patients omit morphemes as a result of a compensatory strategy. In the context of severe articulation problems, people with non-fluent aphasia would consciously or unconsciously resort to ‘telegraphic speech’. Telegraphic speech would consist of an almost exclusive use of content words in their citation form. Importantly, omissions do not, as a rule, result in non-words, as it may happen in omission of the affix. This interpretation cannot explain agrammatism shown in writing or in other tasks that do not require an articulatory effort.

In paragrammatism, substitutions of grammatical morphemes may be the result of problems in word finding. This explanation is in contrast, however, with the fact that decline in word finding does not necessarily correlate with a reduction of paragrammatic errors (Butterworth et al. 1990).

Many authors (e.g., Miceli et al. 1989) have concluded that agrammatism is not a unitary phenomenon. According to this view, agrammatism and paragrammatism do not represent different syndromes but rather designate a wide variety of, more or less, selective deficits affecting free and bound grammatical morphemes with different degrees of severity. Indeed, Miceli et al. (1989) demonstrated how the ability to produce nominal, verbal, and adjectival agreement widely varies across agrammatism cases. Likewise, errors in verbal agreement also vary; although some individuals with agrammatism produce almost exclusively substitutions with the infinitive or the past participle (i.e., the citation forms), in others the proportion of such errors is negligible. The authors concluded that, since one grammatical aspect may be processed independently from another, the study of agrammatism might allow valid inferences on normal grammatical processing only at the single case level. Therefore, there would not be an explanation for agrammatism valid for all cases.

Although the majority of studies have been conducted in English, research on languages with a rich morphological system for instance, Italian, German, or Finnish, where morphology is extremely complex, could provide new insights on the nature of morphological errors. Such studies suggest that what, in English, appear to be omissions of bound morphemes are better interpreted as substitutions. For example, in Hebrew (Grodzinski 1984) or Italian (Miceli et al. 1989), where verbal roots are most often non-words, bound morphemes are never omitted. In light of these observations, apparent omissions observed in English (and, possibly, in other languages without a rich morphology; however, no such data are available in the literature) may indeed be substitutions of a given inflection with a 0- inflection.

Errors affecting bound morphemes are also commonly observed (Patterson 1980) in phonological dyslexia, a condition that is often concomitant with agrammatism. In phonological dyslexia the main problem lies in the inability to apply grapheme-to-phoneme conversion rules. As a result, only words that are already known can be read, via a spared 'lexical route', where full word forms are stored. However, the ability to read depends on the way a word form is stored in the mental lexicon. If, in fact, a given word is stored in a decomposed form (i.e., in the form stem+affix), a person with phonological dyslexia would retrieve the root correctly, but may make morphological errors by missing the affix, substituting it with another affix. In contrast, if the word is stored as a whole (a non-decomposed form), there would be no such errors. Section 4 will review specific studies on this issue.

Cases selectively affecting affixes sharply contrast with cases where the stem is disturbed while the affix is preserved. In some of these cases the inability to produce content words (Caplan, Kellar & Locke 1972; Butterworth 1979; Buckingham 1981; Semenza et al. 1990) is very severely affected. In neologisms, affixation is correct, however, and even perfectly adapted to a syntactically correct sentence frame (e.g., "the stouk galtereed the butrood with a curkl").

The selective disturbance of stems and affixes demonstrated in the above described studies shows that stems and affixes are stored and processed separately in the brain. This finding is a powerful proof in support of linguistic theories like that described in Aronoff (1993).

3. Morphological errors

Morphological errors cannot be easily demonstrated (Badecker & Caramazza 1987). What appear to be morphological errors may be accounted for in various other ways. Semantic problems, phonological problems, and visual similarity may, in fact, cause what may look like authentic morphological mistakes (for a review see Kay 1988).

Miceli & Caramazza (1988) could demonstrate convincing morphological errors in the repetition of adjectives. Their patient, FS, made fewer mistakes when he had to repeat the masculine singular than when repeating the feminine singular and plural forms. Frequency of use and phonological factors were controlled for.

MB, a patient affected by agrammatism and phonological dyslexia, also made convincing morphological errors (Luzzatti, Mondini & Semenza 2001). MB was provided with words altered by an evaluative suffix, for instance, *lettino* "small bed", composed of *letto* "bed" + *-ino*, a diminutive suffix, or *gattone* "big cat", composed of *gatto* "cat" + *-one*, augmentative suffix. MB's performance on these words was compared to performance on pseudo-suffixed words (i.e., simple nouns ending with an orthographic string that is homograph and homophone to an evaluative suffix; e.g., *cervello* "brain", which is not a little *cervo* "deer", and *carbone* "coal", which is not a big *carbo (a non-word)). In the case of suffixed nouns, MB simplified the structure of the target word by stripping the evaluative suffix or paraphrasing it semantically (e.g., *lettino*: "letto, si, … piccolo"! ("bed, yes, … small")). In contrast, MB never made omissions and never paraphrased with pseudo-suffixed words.

Finally, Rastle, Tyler & Marslen-Wilson (2006) had their patient DE read aloud lists of genuinely suffixed words, pseudo-suffixed words, and words with non-morphological embeddings. DE made more stem errors in suffixed words than in other types of words. This effect of morphological status could not be attributed to the relative levels of target imageability and/or frequency.

Another interesting factor influencing morphological errors is the oral or written modality. Rapp, Fisher-Baum & Miozzo (2015) documented a double dissociation in which the production of affixes was disrupted in writing but not speaking or vice versa. The authors concluded that this finding showed that written- and spoken-language systems are independent from a morpho-orthographic operation standpoint.

4. Decomposition and levels in lexical processing

Findings in aphasia allowed for the comparison of two alternative classes of theories about the representation of complex words: the 'full-listing' theories (e.g., Butterworth 1983; Bybee 1985) and the 'decomposition' or 'full parsing' theories (e.g., Taft & Forster 1976; Taft 2004; Libben, Derwing & de Almeida 1999; McKinnon, Allen & Osterhout 2003). Current prevailing opinions are, however, that such theories may not be viewed as contradictory to each other. Several compromise theories have thus been proposed (Caramazza, Laudanna & Romani 1988; Sandra 1990; Zwitserlood 1994; Baayen, Dijkstra & Schreuder 1997; Isel, Gunter & Friederici 2003). Different complex words would be preferentially processed one way or another. For example, very frequently used words would be stored and processed more efficiently in their full form; in contrast, less frequent words would be stored and processed in a decomposed form. Full listing would allow economy in rule application, while decomposition would allow economy of storage.

In neuropsychology the issue of decomposition has been addressed in studies devoted to inflection and derivation (Patterson 1980; Colthearth 1980; Badecker & Caramazza 1987; De Bleser & Bayer 1990; Semenza et al. 1990), and also, extensively, in studies on compounds (Semenza & Mondini 2006; Chiarelli, Menichelli

& Semenza 2007; El Yagoubi et al. 2008; Semenza & Mondini 2010; Semenza et al. 2011b; Semenza & Luzzatti, 2014). The overwhelming evidence provided by these studies is, as reported below, either in favor of full parsing or dual route theories.

Caramazza et al.'s (1985) study on a case of phonological dyslexia showed a marked dissociation in reading different types of non-words. This task is customarily poorly accomplished in such a syndrome, as a consequence of the inability to perform grapheme-to-phoneme conversion. Non-word stimuli could be read significantly better when composed of legal stems and affixes (e.g., *walkest*) than illegal non-words composed of either a real stem and a pseudo-affix (e.g., *walkent*), or of a pseudo-stem and a true affix (e.g., *wolkest*) or other non-words. Since phonological dyslexics read exclusively via the lexical route (alternative to grapheme-to-phoneme conversion and preferred in expert readers for known words), the advantage of legal non-words could only be explained if the representation in the orthographic input lexicon takes the decomposed form.

Words may also be represented in decomposed forms at the level of the orthographic output lexicon (Badecker, Hillis & Caramazza 1989; Mondini et al. 2007). Thus Badecker et al.'s (1989) case DH made morphological errors in writing spontaneously or to dictation. A comparison of errors in writing suffixed words like *ended* versus pseudo-affixed words like *agent* showed fewer errors on the suffixed part of suffixed words. The explanation for such finding is somewhat complicated. DH's impairment was independently demonstrated to be at the level of the 'orthographic output buffer'. The orthographic output buffer is the working memory system that is believed to contain the representation that specifies abstract letter identities and the sequence in which the graphemes in a word appear. Any short-term memory system has a limited capacity and holds information until further production processes come into play. A deficit of such system reduced DH's ability to process a word in function of its length (the longer the word, the heavier the processing burden for the working memory system). The advantage of suffixed over pseudo-affixed words is that the former is fed by the orthographic output lexicon into the orthographic output buffer in decomposed form, consisting of two shorter, more easily processed parts (*end* and *ed*).

Luzzatti et al.'s (2001) agrammatic case MB was impaired in reading regular plural nouns (e.g., *mouths*) with respect to singular nouns (*mouth*), but this difference did not appear in plural dominant nouns (*eyes*). According to Luzzatti et al. (2001), this finding supports the dual-route hypothesis: high-frequency inflected forms like plural dominant nouns would be represented in a non-decomposed format.

Reznick & Friedmann (2015) examined whether and how the morphological structure of written words affects reading in neglect dyslexia. They found that when an affix appeared on the neglected side, it was neglected significantly more often than when the neglected side was part of the root. Root letters on the neglected side were never omitted, whereas affixes were. According to the authors, these findings show an early morphological decomposition of words to their root and affixes at the orthographic-visual analysis stage.

5. Regular and irregular inflection

Processing regular rather than irregular inflection has been proposed to depend, according to a popular set of theories, on the working of different, anatomically distinct, memory systems (Ullman et al. 1997; Ullman 2001; Ullman 2008). The idea was that rule-based language processing, like the production of regular past tense, occurs deep within the left frontal cortex, including Broca's area and the basal ganglia. This area would work on the basis of implicit (procedural) memory systems. The retrieval of irregular verb forms, in contrast, would happen in temporal lobe areas, and be based on an explicit (declarative) memory system. Support for this hypothesis has been provided in a series of studies (e.g., Ullman et al. 1997; Ullman et al. 2005) showing that word-finding difficulties in posterior aphasia, and in people with declarative memory disorders such as Alzheimer's disease, led to more errors with irregular than regular and novel verbs. In contrast, agrammatism in Broca's aphasia, and the impairment of procedures in Parkinson's disease, led to the opposite pattern. Similar patters emerged with receptive tasks (e.g., with studies using lexical decision after priming, Marslen-Wilson & Tyler 1997).

Despite the popularity of the explicit/implicit memory hypothesis, further studies failed to provide empirical support. Thus, cases of agrammatism, all affected by frontal lesions, have been reported showing better performance on regular, as opposed to irregular, past tense (Shapiro & Caramazza 2003; Laiacona & Caramazza 2004; Balaguer et al. 2004; Penke, Janssen & Krause 1999). Furthermore, a meta-analysis by Faroqi-Shah (2007) of studies on the production of morphologically complex regular and irregular verbs in Broca's aphasia failed to show a single consistent dissociation pattern, and over half of the data sets found no difference between regular and irregular verbs.

All of these data indeed seem to indicate a left inferior frontal involvement in processing morphologically complex words, that is not limited to rule-based language transformations. Damage to various processes, including those responsible for the retrieval, interpretation and integration of grammatical features (e.g., tense, person and number agreement of verb forms) would be at the basis of agrammatism (Balaguer et al. 2004). The deficit would affect morphosyntactic mechanisms, independent of the regularity of the verb form.

Other studies addressed irregular morphology from other perspectives. Case MB (Luzzatti et al. 2001), affected by agrammatism and deep dyslexia, showed a specific deficit in reading morphologically complex words. In this case, a series of tasks compared reading of singular nouns, regular plurals (stem + plural suffix), and irregular plurals. Irregularly inflected words were read better than regularly inflected ones, singular nouns were better than plural nouns, and there were fewer errors on irregular plurals than regular plurals. Marked inflected verbs and adjectives were also more affected than the corresponding base forms. These findings demonstrate an impairment of the compositional mechanisms required for processing regularly inflected words. MB's deficit was located in the retrieval of inflectional affixes from the phonological output lexicon.

6. Gender inflection

In languages such as Italian or French, where all nouns have a grammatical gender, gender knowledge must be stored at the lexical level. Luzzatti & De Bleser (1996) studied gender assignment in two Italian speakers, MG and DR, affected by agrammatism. They were required to produce the definite article (*la* for feminine nouns; *il* for masculine nouns) in response to a given noun. In Italian, the general rule is that words with an -*o* ending are masculine, while words an -*a* ending are feminine. Words with the -*e* ending can be either masculine or feminine. In items with a natural gender (e.g., *la madre* [F], "the mother") the right article can be determined on the basis of semantic information. In neutral words with the -*e* ending (e.g., *il sole* [M], "the sun") the gender is entirely lexicalized and cannot be determined on the basis of semantic information or of the phonological/morphological rule. Gender assignment was almost fully preserved in MG, thus showing once more how cases diagnosed with agrammatism may be very different from each other. In contrast, DR systematically attributed the masculine gender to feminine nouns with the opaque ending -*e* (e.g., *il torre instead of *la torre*). DR performed better on derived nouns, in which gender is determined by the derivational suffix.

A deficit with gender was also reported in the aforementioned agrammatic case MB (Mondini, Luzzatti & Semenza 1999). MB made no errors in applying the ending rules, but was impaired in lexically-based gender assignment. The deficit appeared even in those nouns, like *madre,* bearing a natural gender.

These findings seem to suggest that gender assignment mechanisms (rule, semantically, and lexically based) are represented and processed independently from each other.

7. Derivation

Aphasia research on derivation is much more limited. Moreover, some of the claims made on this issue are controversial.

Semenza et al. (1990) studied three cases of fluent aphasia. Their spontaneous speech was characterized by a huge number of neologisms, many of which (like *fratellismo, literally "brotherism") were legal combinations of meaningful parts of real words, including derivational suffixes and prefixes. This finding was interpreted as showing that these individuals, despite their word finding problems, retained control of derivational rules. Productive and less productive derivational affixes were present with a frequency similar to that found in Italian.

Semenza et al. (2002) studied the production of prefixes in two Slovenian speakers with aphasia. The target items were verbs, nouns and adjectives derived by prefixation with one or two prepositions (a very common word-formation rule in Slovenian). In both participants prepositions in prefixes were mostly preserved phonologically. In contrast, many phonological errors affected the remaining portion of the words. The errors respected morphological boundaries. Prefixes, when mistaken, were omitted and substituted with other prefixes. Thus, these errors did not concern single phonemes. Non-prefixed words were not less affected in their initial part than the following portion of the word. These findings show that errors were determined by the morphological structure of the word. Both participants seemed to process morphologically complex words prefixed with a preposition by assembling routines rather than as whole forms.

Derivation, it has been suggested, may be processed with the contribution of the right hemisphere. In some studies, right hemisphere damage was shown to impair derivational tasks (Marangolo et al. 2003; Marangolo & Piras 2008). The evidence is however poor, because the problem could have been determined by generic deficits and contradictory data have been produced. In an fMRI study (Marangolo et al. 2006) on healthy participants, a derivational task, relative to an inflectional task, showed an activation of left fronto-parietal areas. There was no activation of the right hemisphere.

8. The independence of inflection and derivation

An important issue is whether inflection and derivation are autonomous subcomponents of the lexicon and whether they reflect autonomous processes in the brain. The issue is controversial as far as studies in healthy participants are concerned (for an early review see Caplan 1992). Available neuropsychological literature does not provide conclusive results.

Miceli & Caramazza (1988) reported the case of FS, who made morphological errors in spontaneous production and in repetition of single words. In repetition of morphologically complex words, most errors resulted in real words with the same base of the target but with a different affix. The vast majority of errors were substitutions of inflections. In repeating derived words, FS made far fewer derivational errors. Inflection thus seemed much more affected than derivation. This result, however, may have been determined by the discrepancy between inflected words, which are much more frequent, and derived words. Miceli & Caramazza (1988) took into consideration the disproportion between inflectional and derivational items and corrected for the relative frequency. This, notwithstanding the proportion of errors affecting inflections, was still much larger than those affecting derivations, thus apparently supporting the dissociation. However, a primary complication comes from the fact that derived words in Italian are also inflected. A further and more compelling complication is that the probability of making a derivational error is not only a function of the proportion of derived words with respect to inflected words – it is also a function of the proportion of derived words that may appear in a chance response. Therefore, derivational errors may be underestimated not only because of the paucity of derived words in the stimulus set, but also in the set of potential answers. The most convincing finding in favor of a dissociation sparing derivational processes is that FS was almost flawless in repeating the prefix part of prefixed words. Prefixes may be processed differently from suffixes, however.

In a similar investigation, Laine et al. (1995) reported an agrammatic individual who made more morphological errors with inflected than with derived nouns in reading and repetition. Further data on the autonomy of inflection and derivation come from Tyler & Cobb's (1987) study on case DE. DE was able to process derivational suffixes normally, but could not discriminate contextually appropriate and inappropriate inflectional suffixes. Tyler & Marslen-Wilson (1997) further studied DE together with case JG, affected by similar problems, in two priming experiments. Both cases showed difficulties with regularly inflected past tense forms; in contrast, their priming for derivationally related forms was normal.

In summary, evidence for the independence of inflectional and derivational processes in the brain is not conclusive. It rests so far on scant data, limited to possible selective disturbances of inflection. More convincing conclusions could be reached, however, if the opposite pattern of dissociation would also be observed (i.e., a problem with derivation and not with inflection).

9. Morphology without phonology

Some people affected by aphasia may show morphological competence while lacking the ability to retrieve the full phonological form of a word. Typically, non-words may be produced from a neologistic or existent stem and an existing affix that may agree with other parts of speech (e.g., articles) in the sentence context (Caplan et al. 1972; Butterworth 1979; Buckingham 1981; Semenza et al. 1990). These cases suggest that knowledge of word-formation rules is independent from the ability to retrieve the full phonological form.

The best source of evidence showing that lexical retrieval is independent from the access to word-specific aspects of grammar comes from cases of anomia. Badecker, Miozzo & Zanuttini (1995) investigated one such case, Dante, in naming pictures of entities where grammatical gender could not be determined semantically. When Dante was unable to retrieve the phonological form of the noun, he nonetheless always provided the correct article. The authors offered an interpretation of this finding within a dual-stage model of lexical access (e.g., Levelt 1989). Morphological information, represented at the so-called lemma level, would be preserved, while access to the next stage, the lexeme (i.e., the phonological form), was disturbed. An analogous interpretation is given for the so-called 'compound effect' (Semenza & Mondini 2010), a phenomenon that will be described in the next section.

10. Compounding

In picture naming tasks people with aphasia substitute compound words for compound targets (Semenza et al. 1992; Hittmair-Delazer et al. 1994; Semenza, Luzzatti & Carabelli 1997; Blanken 2000; Badecker 2001; Chiarelli et al. 2007), but they seldom produce compound words or compound neologisms instead of simple word targets for items they cannot name. This phenomenon, called 'compound effect', is found for both transparent and opaque compounds. The compound effect has been interpreted (Semenza et al. 2011b) within the framework of the lemma theory (Levelt 1989). Although retrieval of the corresponding phonological form at the lexeme level fails at the lemma-level, morpho-syntactic information (about the compound structure) is preserved.

Likewise, noun-noun compounds are substituted in aphasia by noun-noun compounds, verb-noun compounds by verb-noun compounds, etc. Thus, knowledge about the word structure is preserved in absence of the ability to retrieve the word. Word-formation rules also seem to be preserved in aphasia independently from the retrieval of the phonological form (Hittmair-Delazer et al. 1994; Semenza et al. 1997). For example, in substitution errors of just one of the components, the correctly produced component (either the first or the second) maintains its original position. In the study by Hittmair-Delazer et al. (1994), in verb-noun neologisms, the verb stem correctly appeared in first position, dropping the *-en* ending of the infinitive, thus respecting the rules for German.

The production of compound neologisms replacing existing compound words is a common phenomenon. By definition, the mental lexicon does not contain neologisms. Thus, as argued by Badecker (2001), an intact representation of the morphological structure must initiate the compositional procedure evident in these errors. This is further proof that knowledge of the compound structure is stored in the brain independently of the knowledge of the phonological form.

On the whole, studies in aphasia support full-parsing models (Hittmair-Delazer et al. 1994; Semenza et al. 1997) and dual route models (Mondini et al. 2002), rather than full-listing models. The production of verb-noun compounds (e.g., *schiac-cianoci*, "nut-cracker") in Italian provided one of the most convincing findings in support of decomposition (Mondini et al. 2005; Mondini et al. 2004, Semenza et al. 1997). In many cases of Broca's aphasia, problems are more severe with verbs than with nouns (Semenza 1999). In such cases the verb component in verb-noun compounds is often dropped. This effect is not determined by position, since it does not hold for the first component in noun-noun compounds. Italian verb-noun compounds are nouns. Therefore, if such compounds were decomposed during processing, a greater difficulty for verbs would result in the omission of the verb component.

Decomposition was also demonstrated in the study of prepositional compounds, a common type of compound in Italian. Prepositional compounds are compounds in which the morphological structure is noun-preposition-noun. The linking preposition is often opaque with respect to the meaning of the whole compound (e.g., *film* ***in*** *bianco e nero*, "black and white movie" [*lit.* "film in black and white"], vs. *film* ***a*** *colori*, "color movie" [*lit*. "film on colors"]). Mondini et al. (2005) tested prepositional compounds in people with agrammatism in a series of naming, reading, repetition and completion tasks. Among agrammatic speakers, omission of articles and prepositions is one of the main symptoms. This pattern of impairment clearly emerged when people with agrammatism attempted to retrieve the phonological form of fully lexicalized compounds containing syntactically and semantically opaque prepositional links. Prepositional compounds seem to be (de) composed at some level of processing. The agrammatic deficit can thus damage the retrieval of the linking preposition. However, the opaqueness of the linking preposition must make the whole form representation necessary. Mondini et al. (2005) interpreted their findings as evidence that both the whole word form and the decomposed form must be available before the phonological representation of complex words can be accessed.

Depending on the characteristics of compounds in a given language, the type of aphasia category (e.g., Broca's, Wernicke's, etc.) also plays an important role and interacts with the position of the components in the compound (Semenza & Mondini 2010). Thus, in a study by Semenza et al. (2011b), in Broca's aphasia the first component of a two-component compound was omitted or substituted about three times as often as the second component, while such a difference was not found in the other categories. As aforementioned, the effect was due to the specific difficulty in Broca's aphasia with the first component of verb-noun compounds. Articulation difficulties, typical of Broca's aphasia, were not a sufficient explanation for the specific problem with the first component.

Delazer & Semenza (1998), Badecker (2001) and Semenza et al. (2011b) provided some evidence of the fact that the compound components are retrieved simultaneously. In their studies of anomic aphasia the first and the second component of compounds were affected with the same frequency. No such difference was found in Broca's and Wernicke's aphasia. The authors agree on the idea that at some point in retrieval, one entry activates two separate forms in parallel. In the case of anomic aphasia, the difficulty seems to be in linking separately retrieved component forms to their position in the target frame. This operation must happen at the intermediate stage between semantics and phonology, the so-called lemma level, where the components must be combined and the way they have to be combined is specified. Marelli et al. (2012) further pursued the issue of lemma-level representations, arguing that both decomposed and whole-word representations would exist at this level. In their study, on an individual affected by deep/phonological dyslexia, evidence in favor of decomposed representations came from a replication of the aforementioned effect whereby in verb-noun compounds more errors were made in the verb component than in the noun component. At the same time, there was evidence supporting a whole-word representation theory, because the same individual read verb-noun compounds (e.g., *lavapiatti*, "dishwasher" [*lit.* "wash-dishes"]) better than verb-phrases (e.g., *lui lava piatti*, "he washes dishes") when embedded in sentences.

A study by Mondini et al. (2002) compared the production of compounds to the production of noun phrases. In this study, two individuals with Broca's aphasia were more successful in correctly producing, reading, and repeating inflected adjectives within compounds than in non-lexicalized noun-adjective or adjective-noun phrases. This finding supports the view that in morpho-syntactic deficits different mechanisms are used in producing compounds and noun phrases. The latter require two steps: first, the two lexical items have to be accessed in the mental lexicon and, second, standard morpho-syntactic operations have to be applied in order to achieve grammatical agreement. In contrast, compounds involving adjectives would be accessed as whole words, thus bypassing morphosyntactic operations. Again, all meaningful representations (of the whole word and of its individual components) seem to be activated in retrieval, possibly based upon what Libben (2006) calls maximization of opportunity.

Another important dimension of compounding, headedness, was also addressed. Evidence for the fact that headedness, a dimension conceptualized in theoretical linguistics, has neurological underpinnings was found, albeit with some difficulties.

One problem for aphasiologists has been that it is hard to disentangle the effects of the head from the effect of position. In English or in German, for instance, the head is always in second position. To address the problem, Jarema, Perlak & Semenza (2010) contrasted, across a series of tasks (including naming and reading aloud), the production of French and English compound words differing in the position of the head (left-sided in French, e.g., *certificat médical*, and right-sided in English, e.g., *medical certificate*). Three bilingual individuals with aphasia thus showed preservation of the head with respect to the modifier – although in French left-headed compounds they preserved the left component, in English they preserved the right one.

Marelli et al. (2014) tested naming with both left-headed and right-headed compounds among 45 Italian-speaking individuals with aphasia. The modifier was found to be more difficult to retrieve than the head, but only for head-final compounds.

An alternative opportunity to study the headedness effect in aphasia is offered by neglect dyslexia. Semenza et al. (2011a) found that left-headed compounds were read better, apparently attracting more attention, than right-headed compounds. This result was not attributable to differences in frequency, familiarity, age of acquisition, or imageability, because these factors were controlled for. Semenza et al. (2011a) suggested that attention is captured by the head component after implicit reading of the whole compound. The head would require a relatively lighter processing load than the modifier and would benefit from top-down facilitation. Similar results were obtained by Marelli et al. (2013).

Clinical investigations thus suggest that the head of a compound is processed in the brain differently from the modifier. Converging evidence has been found using instrumental techniques. The electrophysiological correlates of processing compound words were in fact found to vary at the level of the P300 component according to the position of the head (El Yagoubi et al. 2008). Arcara, Semenza & Bambini (2014) used eye movement recording to compare right-headed and left-headed compounds. Right-headed compounds elicited longer total reading times than left-headed compounds. Higher processing costs for right-headed as opposed to left-headed compounds possibly index a reanalysis of the stimulus in order to correctly assign the constituent properties.

11. Conclusion

Research in aphasia in recent decades provides a large body of evidence about word-formation processes and their neurological underpinnings. Errors determined by failure in morphological processing show that stems and affixes are separately represented in the cognitive system and brain. Evidence has emerged of a fine-grained organization, distinguishing the main morphological processes, inflection, derivation, and compounding. Moreover, evidence for decomposition in processing has been found, mostly favoring dual route hypotheses. Importantly, information about morphology has been shown to be independent from the ability to retrieve full phonological forms. Headedness was shown to have psychological reality and neural underpinnings.

Further integration of all of these notions with mainstream linguistics and cross-fertilization among different disciplines is highly desirable. Linguists, on their side, may try to make neuropsychologists aware of issues where data from sources external to linguistics may help them to evaluate different alternatives. Some morphological processes may be conspicuously represented in languages that have not been studied so far. Neuropsychologists should be ready to pick up on such opportunities, selecting those that can be reliably explored with the methods of cognitive neuroscience. Neuropsychologists equipped for such research are relatively few, however, at the present time: neuro-linguistic research is not rewarding in terms of citations nowadays but in exceptional cases. Finally, the mediation of psycholinguistics is also necessary in order to ensure that experiments are properly designed taking into account all possible confounds.

References

Allen, Mark & William Badecker. 2001. Morphology: The internal structure of words. In Brenda Rapp (ed.), *The handbook of cognitive psychology*, 211–232. Philadelphia: Psychology Press.

Arcara, Giorgio, Carlo Semenza & Valentina Bambini. 2014. Word structure and decomposition effects in reading. *Cognitive Neuropsychology* 31(1–2). 184–218. https://doi.org/10.1080/02643294.2014.903915

Aronoff, Mark. 1993. *Morphology by itself*. Stems and inflectional classes. Cambridge, MA: MIT Press.

Baayen, Harald R., Ton Dijkstra & Robert Schreuder. 1997. Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Brain and Language* 37(1). 94–117.
Badecker, William. 2001. Lexical composition and the production of compounds: Evidence from neologistic errors in naming. *Language and Cognitive Processes* 16(4). 337–390. https://doi.org/10.1080/01690960042000120
Badecker, William & Alfonso Caramazza. 1987. The analysis of morphological errors in a case of acquired dyslexia. *Brain and Language* 32(2). 278–305. https://doi.org/10.1016/0093-934X(87)90129-5
Badecker, William, Michele Miozzo & Raffaella Zanuttini. 1995. The two-stage model of lexical retrieval: Evidence from a case of anomia with selective preservation of grammatical gender. *Cognition* 57(2). 193–216. https://doi.org/10.1016/0010-0277(95)00663-J
Badecker, William, Argye E. Hillis & Alfonso Caramazza. 1989. Lexical morphology and its role in the writing process: Evidence from a case of acquired dyslexia. *Report of the Cognitive Neuropsychology Laboratory* 46. The Johns Hopkins University.
Balaguer, Ruth D., Albert Costa, Nuria Sebastian-Galles, Montse Juncadella & Alfonso Caramazza. 2004. Regular and irregular morphology and its relationship with agrammatism: Evidence from two Spanish-Catalan bilinguals. *Brain and Language* 91(2). 212–222. https://doi.org/10.1016/j.bandl.2004.02.007
Blanken, Gerhard. 2000. The production of nominal compounds in aphasia. *Brain and Language* 74(1). 84–102. https://doi.org/10.1006/brln.2000.2338
Buckingham, Hugh W. 1981. Where do neologisms come from? In Jason W. Brown (ed.) *Jargonaphasia*, 39–62. New York: Academic Press. https://doi.org/10.1016/B978-0-12-137580-5.50009-4
Butterworth, Brian. 1979. Hesitation and the production of verbal paraphasias and neologisms in jargon aphasia. *Brain and Language* 8(2). 133–161. https://doi.org/10.1016/0093-934X(79)90046-4
Butterworth, Brian. 1983. *Language Production. Volume 2: Development, Writing and Other Language Processes*. New York: Academic Press.
Butterworth, Brian, Marta Panzeri, Carlo Semenza & Tiziana Ferreri. 1990. Paragrammatism: A longitudinal study of an Italian patient. *Language and Cognitive Processes* 5(2). 115–140. https://doi.org/10.1080/01690969008402101
Bybee, Joan. 1985. *Morphology. A study of the relation between meaning and form*. Amsterdam: Benjamins. https://doi.org/10.1075/tsl.9
Caplan, David. 1992. *Language. Structure processing and disorders*. Cambridge, MA: MIT Press.
Caplan, David, Lucia Kellar & Simeon Locke. 1972. Inflection of neologisms in aphasia. *Brain* 95(1). 169–172. https://doi.org/10.1093/brain/95.1.169
Caramazza, Alfonso, Alessandro Laudanna & Cristina Romani. 1988. Lexical access and inflectional morphology. *Cognition* 28(3). 297–332. https://doi.org/10.1016/0010-0277(88)90017-0
Caramazza, Alfonso, Gabriele Miceli, M. Caterina Silveri & Alessandro Laudanna. 1985. Reading mechanisms and the organization of the lexicon: Evidence from acquired dyslexia. *Cognitive Neuropsychology* 2(1). 81–114. https://doi.org/10.1080/02643298508252862
Chiarelli, Valentina, Alina Menichelli & Carlo Semenza. 2007. Naming compounds in Alzheimer's disease. *The Mental Lexicon* 2(2). 259–269. https://doi.org/10.1075/ml.2.2.07chi
Colthearth, Max. 1980. Reading, phonological recoding and deep dyslexia. In Max Coltheart, Karalyn E. Patterson & John C. Marshall (eds.) *Deep Dyslexia*, 197–226. London: Routledge & Kegan Paul.

De Bleser, Ria & Josef Bayer. 1990. Morphological reading errors in a German case of deep dyslexia. In Jean L. Nespolous & Pierre Villard (eds.), *Morphology, phonology and aphasia*, 32–59. Berlin: Springer. https://doi.org/10.1007/978-1-4613-8969-9_3

Delazer, Margarete & Carlo Semenza. 1998. The processing of compound words: A study in aphasia. *Brain and Language* 61(1). 54–62. https://doi.org/10.1006/brln.1997.1878

El Yagoubi, Radouane Y., Valentina Chiarelli, Sara Mondini, Gelsomina Perrone, Morena Danieli & Carlo Semenza. 2008. Neural correlates of Italian compounds and potential impact of headedness effect: An ERP study. *Cognitive Neuropsychology* 25(4). 559–581. https://doi.org/10.1080/02643290801900941

Faroqi-Shah, Yasmeen. 2007. Are regular and irregular verbs dissociated in non-fluent aphasia? A meta-analysis. *Brain Research Bullettin* 14(74). 1–13.

Goodglass, Harold, Jean B. Gleason, Nancy Ackerman-Bernhalz & Mary R. Hide. 1972. Some linguistic structures in the speech of a Broca's aphasic. *Cortex* 8(2). 191–212. https://doi.org/10.1016/S0010-9452(72)80018-2

Grodzinski, Yosef. 1984. The syntactic characterization of agrammatism. *Cognition* 16(2). 99–120. https://doi.org/10.1016/0010-0277(84)90001-5

Halle, Morris. 1973. Prolegomena to theory of word formation. *Linguistic Inquiry* 4(1). 3–16. https://doi.org/10.1162/ling_a_00313

Hittmair-Delazer, Margarete, Barbara Andree, Carlo Semenza & Ria De Bleser. 1994. Naming by German compounds. *Journal of Neurolinguistics* 8(1). 27–41. https://doi.org/10.1016/0911-6044(94)90004-3

Isel, Frederic, Thomas C. Gunter & Angela D. Friederici. 2003. Prosody-assisted head-driven access to spoken German compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29(2). 277–288.

Jarema, Gonia, Danuta Perlak & Carlo Semenza. 2010. The processing of compounds in bilingual aphasia: a multiple-case study. *Aphasiology* 24. 126–140. https://doi.org/10.1080/02687030902958225

Kay, Janice. 1988. On the origin of morphological errors in two cases of acquired dyslexia. *Linguistics* 26(4). 669–697. https://doi.org/10.1515/ling.1988.26.4.669

Laiacona, Marcella & Alfonso Caramazza. 2004. The noun-verb dissociation in language production: Varieties of causes. *Cognitive Neuropsychology* 21(2–4). 103–123. https://doi.org/10.1080/02643290342000311

Laine, Matti, Jussi Niemi, Päivi Koivuselkä-Sallinen & Jukka Hyönä. 1995. Morphological processing of polymorphemic nouns in a highly inflecting language. *Cognitive Neuropsychology* 12(5). 477–502. https://doi.org/10.1080/02643299508252005

Leminen, Alina, Eva Smolka, Jon A. Duñabeitia & Christos Pliatsikas. 2019. Morphological processing in the brain: The good (inflection), the bad (derivation) and the ugly (compounding). *Cortex* 116(7). 4–44. https://doi.org/10.1016/j.cortex.2018.08.016

Levelt, Willem J. M. 1989. *Speaking. From intention to articulation*. Cambridge, MA: MIT Press.

Libben, Gary. 2006. Why study compound processing? An overview of the issues. In Gary Libben & Gonia Jarema (eds.), *The representation and processing of compound words*, 71–95. Oxford: Oxford University Press.

Libben, Gary, Bruce L. Derwing & Roberto G. de Almeida. 1999. Ambiguous novel compounds and models of morphological parsing. *Brain and Language* 68(1–2). 378–386. https://doi.org/10.1006/brln.1999.2093

Luzzatti, Claudio & Ria De Bleser. 1996. Morphological processing in Italian agrammatic speakers: Eight experiments in lexical morphology. *Brain and Language* 54(1). 26–74. https://doi.org/10.1006/brln.1996.0060

Luzzatti, Claudio, Sara Mondini & Carlo Semenza. 2001. Lexical representation of morphologically complex words: Evidence from an Italian agrammatic patient. *Brain and Language* 79(3). 345–359. https://doi.org/10.1006/brln.2001.2475

Marangolo, Paola & Fabrizio Piras. 2008. Dissociations in processing derivational morphology: The right basal ganglia involvement. *Neuropsychologia* 46(1). 196–205. https://doi.org/10.1016/j.neuropsychologia.2007.07.025

Marangolo, Paola, Chiara Incoccia, Luigi Pizzamiglio, Umberto Sabatini, Alessandro Castriota-Scanderbeg & Cristina Burani. 2003. The right hemisphere involvement in the processing of morphologically derived words. *Journal of Cognitive Neuroscience* 15(3). 364–371. https://doi.org/10.1162/089892903321593090

Marangolo, Paola, Fabrizio Piras, Gaspare Galati & Cristina Burani. 2006. Functional anatomy of derivational morphology. *Cortex* 42(8). 1093–1106. https://doi.org/10.1016/S0010-9452(08)70221-1

Marelli, Marco, Silvia Aggujaro, Franco Molteni & Claudio Luzzatti. 2012. The multiple-lemma representation of Italian compound nouns: A single case study of deep dyslexia. *Neuropsychologia* 50(5). 852–861. https://doi.org/10.1016/j.neuropsychologia.2012.01.021

Marelli, Marco, Silvia Aggujaro, Franco Molteni, & Claudio Luzzatti. 2013. Understanding the mental lexicon through neglect dyslexia: A study on compound nounreading. *Neurocase* 19(2). 128–144. https://doi.org/10.1080/13554794.2011.654222

Marelli, Marco, Giusy Zonca, Antonella Contardi & Claudio Luzzatti. 2014. The representation of compound headedness in the mental lexicon: A picture naming study in aphasia. *Cognitive Neuropsychology* 31(1.2). 26–39. https://doi.org/10.1080/02643294.2013.860024

Marslen-Wilson, William D. & Lorraine Tyler. 1997. Dissociating types of mental computation. *Nature* 387(6633). 592–594. https://doi.org/10.1038/42456

McKinnon, Richard, Mark Allen & Lee Osterhout. 2003. Morphological decomposition involving non-productive morphemes: ERP Evidence. *NeuroReport* 14(6). 883–886. https://doi.org/10.1097/00001756-200305060-00022

Miceli, Gabriele & Alfonso Caramazza. 1988. Dissociation of inflection and derivational morphology. *Brain and Language* 35(1). 24–65. https://doi.org/10.1016/0093-934X(88)90100-9

Miceli, Gabriele, Anna Mazzucchi, Lisa Menn & Harold Goodglass. 1983. Contrasting cases of Italian agrammatic aphasia without comprehension disorder. *Brain and Language* 19(1). 65–97. https://doi.org/10.1016/0093-934X(83)90056-1

Miceli, Gabriele, Caterina Silveri, Cristina Romani & Alfonso Caramazza. 1989. Variation in the pattern of omissions and substitutions of grammatical morphemes in the spontaneous speech of so-called agrammatic patients. *Brain and Language* 36(3). 447–492. https://doi.org/10.1016/0093-934X(89)90079-5

Mondini, Sara, Claudio Luzzatti & Carlo Semenza. 1999. Grammatical gender in an Italian agrammatic patient. *Brain and Cognition* 69(3). 278–281.

Mondini, Sara, Gonia Jarema, Claudio Luzzatti, Cristina Burani & Carlo Semenza. 2002. Why is “red cross” different from “yellow cross”? A neuropsychological study of noun-adjective agreement within Italian compounds. *Brain and Language* 81(1–3). 621–634. https://doi.org/10.1006/brln.2001.2552

Mondini, Sara, Claudio Luzzatti, Giusy Zonca, Caterina Pistarini & Carlo Semenza. 2004. The mental representation of Verb-Noun compounds in Italian: Evidence from a multiple single-case study in aphasia. *Brain and language* 90(1–3). 470–477. https://doi.org/10.1016/S0093-934X(03)00458-9

Mondini, Sara, Claudio Luzzatti, Paola Saletta, Nadia Allamano & Carlo Semenza. 2005. The mental representation of prepositional compounds: Evidence from Italian agrammatic patients. *Brain and Language* 94(2). 178–187. https://doi.org/10.1016/j.bandl.2004.12.005

Mondini, Sara, Roberta Baronio, Valentina Chiarelli, Morena Danieli, Radouane El Yagoubi, Gelsomina Perrone & Carlo Semenza. 2007. Noun-noun compounds in the access to the phonological output buffer. *Brain and Language* 103(4). 57–58. https://doi.org/10.1016/j.bandl.2007.07.043

Patterson, Karalyn E. 1980. Derivational errors. In Max Coltheart, Karalyn E. Patterson & John C. Marshall (eds.), *Deep Dyslexia*, 286–306. London: Routledge.

Penke, Martina, Ulrike Janssen & Marion Krause. 1999. The representation of inflectional morphology: Evidence from Broca's aphasia. *Brain and Language* 68(1–2). 225–232. https://doi.org/10.1006/brln.1999.2103

Rapp, Brenda, Simon Fisher-Baum & Michele Miozzo. 2015. Modality and morphology: What we write may not be what we say. *Psychological Science* 26(6). 892–902. https://doi.org/10.1177/0956797615573520

Rastle, Katleen, Lorraine Tyler & William Marslen-Wilson. 2006. New evidence for morphological errors in deep dyslexia. *Brain and Language* 97(2). 189–199. https://doi.org/10.1016/j.bandl.2005.10.003

Reznick, Julia & Naama Friedmann. 2015. Evidence from neglect dyslexia for morphological decomposition at the early stages of orthographic-visual analysis. *Frontiers Human Neuroscience* 9(497). 1–19.

Sandra, Dominiek. 1990. Processing and representational aspects of compound words in visual word recognition: An experimental approach and a methodological appraisal. *Quarterly Journal of Experimental Psychology* 42°. 529–567. https://doi.org/10.1080/14640749008401236

Semenza, Carlo. 1999. Lexical semantic disorders in aphasia. In Gianfranco Denes & Luigi Pizzamiglio (eds.), *Handbook of Neuropsychology*, 215–244. London: Erlbaum.

Semenza, Carlo, Patrizia Bisiacchi & Victor Rosenthal. 1988. A function for cognitive neuropsychology. In Gianfranco Denes, Carlo Semenza & Patrizia S. Bisiacchi (eds.), *Perspectives on cognitive neuropsychology*, 3–30. London: Lawrence Erlbaum Associates.

Semenza, Carlo, Brian Butterworth, Marta Panzeri & Tiziana Ferreri. 1990. Word formation: New evidence from aphasia. *Neuropsychologia* 28(5). 499–502. https://doi.org/10.1016/0028-3932(90)90075-Y

Semenza, Carlo, Brian Butterworth, Marta Panzeri & Margarete Hittmair Delazer. 1992. Derivational rules in aphasia. *Proceedings of the Berkeley Linguistic Society Annual Meeting*, 435–440. https://doi.org/10.3765/bls.v18i1.1570

Semenza, Carlo, Claudio Luzzatti & Simona Carabelli. 1997. Morphological representation of compound nouns: A study on Italian aphasic patients. *Journal of Neurolinguistics* 10(1). 33–43. https://doi.org/10.1016/S0911-6044(96)00019-X

Semenza, Carlo, Luisa Girelli, Marti, Jan Kobal & Anton Mesec. 2002. Derivation by prefixation in Slovenian: Two aphasia case studies. *Brain and Language* 81(1–3). 242–249. https://doi.org/10.1006/brln.2001.2521

Semenza, Carlo & Sara Mondini. 2006. Neuropsychology of compound words. In Gary Libben & Gonia Jarema (eds.), *The representation and processing of compound words*, 71–95. Oxford: Oxford University Press.

Semenza, Carlo & Sara Mondini. 2010. Compound words in neuropsychology. In Susan Olsen (ed.), *New impulses in word formation*, 331–348. Hamburg: Buske.

Semenza, Carlo, Giorgio Arcara, Silvia Facchini, Francesca Meneghello, Marco Ferraro, Laura Passarini, Cristina Pilosio, Giovana Vigato & Sara Mondini. 2011a. Reading compounds in neglect dyslexia: the headedness effect. *Neuropsychologia* 49(11). 3116–3120. https://doi.org/10.1016/j.neuropsychologia.2011.07.020

Semenza, Carlo, Serena De Pellegrin, Irene Battel, Martina Garzon, Francesca Meneghello & Valentina Chiarelli. 2011b. Compounds in different aphasia categories: A study on picture naming. *Journal of Clinical and Experimental Neuropsychology* 33(10). 1099–1107. https://doi.org/10.1080/13803395.2011.603691

Semenza, Carlo & Claudio Luzzatti. 2014. Combining words in the brain: The processing of compound words. *Cognitive Neuropsychology* 31(1–2). 1–7. https://doi.org/10.1080/02643294.2014.898922

Shapiro, Kevin & Alfonso Caramazza. 2003. The representation of grammatical categories in the brain. *Trends in Cognitive Sciences* 7(5). 201–206. https://doi.org/10.1016/S1364-6613(03)00060-3

Taft, Marcus. 2004. Morphological decomposition and the reverse base frequency effect. *Quarterly Journal of Experimental Psychology* 57A. 745–765. https://doi.org/10.1080/02724980343000477

Taft, Marcus & Kenneth I. Forster. 1976. Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior* 15(6). 607–620. https://doi.org/10.1016/0022-5371(76)90054-2

Tyler, Lorraine K. & Howard Cobb. 1987. Processing bound grammatical morphemes in context: The case of an aphasic patient. *Language and Cognitive Processes* 2(3/4). 245–262. https://doi.org/10.1080/01690968708406934

Tyler, Lorraine K. & William D. Marslen-Wilson. 1997. Disorders of combination: processing complex words. *Brain and Language* 60(4). 48–50.

Ullman, Michael T. 2001. A neurocognitive perspective on language: the declarative/procedural model. *Nature Review Neuroscience* 2(10). 717–26. https://doi.org/10.1038/35094573

Ullman, Michael T. 2008. The role of memory systems in disorders of language. In Brigitte Stemmer & Harry A. Whitaker (eds.), *Handbook of the neuroscience of language*, 189–198. Oxford: Elsevier. https://doi.org/10.1016/B978-0-08-045352-1.00018-5

Ullman, Michael T., Susan Corkin, Maria Coppola, Gregory Hickok, John H. Growdon, Walter J. Koroshetz & Steven Pinker. 1997. A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience* 9(2). 266–276. https://doi.org/10.1162/jocn.1997.9.2.266

Ullman, Michael T., Roumyana Pancheva, Tracy Loved, Yee Eiling, David Swinney & Gregory Hickok. 2005. Neural correlates of lexicon and grammar: Evidence from the production, reading, and judgment of inflection in aphasia. *Brain and Language* 93(2). 185–238. https://doi.org/10.1016/j.bandl.2004.10.001

Zwitserlood, Pienie. 1994. Processing and representation of Dutch compounds: Effects of semantic transparency. *Language and Cognitive Processes* 9(3). 341–36. https://doi.org/10.1080/01690969408402123

CHAPTER 9

The suffixing preference

A preliminary report on processing affixes in Georgian

Alice C. Harris[1] and Arthur G. Samuel[2,3]
[1]University of Massachusetts Amherst / [2]Stony Brook University / [3]Ikerbasque, Basque Center on Cognition, Brain, and Language

The suffixing preference refers to the observation that cross-linguistically suffixes are more abundant than prefixes (Greenberg 1963). Hawkins & Cutler (1988) explain this preference in part by noting that spoken word recognition relies heavily on the beginnings of words, making it advantageous to have no prefix. To test the Hawkins-Cutler hypothesis in Georgian, we carried out lexical decision experiments, a standard kind of experiment in which the participant is presented with real words and nonce words and must identify which is which. In general, responses to Georgian words with prefixes were more accurate and/or faster than to comparable words with suffixes. These results suggest that prefixes may be easier to process than suffixes, contrary to the universalist claim stated above.

Keywords: suffixing preference, word recognition, processing, lexical decision, Georgian

The suffixing preference refers to the observation that there are more languages that use only suffixes than ones that use only prefixes, and that among languages that use both, suffixes are more abundant (Greenberg 1963). Hawkins & Cutler (1988) explain this preference in part by noting that spoken word recognition relies most heavily on the beginnings of words, making it advantageous to have no prefix. There has been a great deal of work since then on differences between prefixes and suffixes, but nearly all of this work has been carried out on a single language – English (with a little work on a handful of European languages that are very similar to English – Dutch, German, Spanish, French, Italian). Exceptions include Finnish (e.g., Bertram, Laine & Karvinen 1999) and Sesotho (Kgolo & Eisenbeiss 2015). Early work in psycholinguistics assumed that processing was universal; processing methods used in English could reasonably be assumed to be used in other languages. However, Cutler (1985, 2009) shows that, in fact, languages can differ

https://doi.org/10.1075/cilt.353.09har

in processing, and subsequent work by others has confirmed this. This fact makes it essential to consider the processing of affixes in languages other than western European ones.

To test the Hawkins-Cutler hypothesis in Georgian, we carried out lexical decision experiments, a standard kind of experiment in which the participant is presented with real words and nonce words and asked to identify which is which. In general, responses to Georgian words with prefixes were more accurate and/or faster than to comparable words with suffixes. These responses suggest that prefixes may be easier to process than suffixes, contrary to the universalist claims of Hawkins & Cutler (1988).

Both of the experiments described here are part of a larger project on affixes and their precursors, clitics. The project includes additional work on affixes of Georgian, work on affixes of English, and work on clitics in Udi and European Portuguese. This paper is our first report of results from this project.

We provide an introduction to the Hawkins-Cutler hypothesis and subsequent work in § 1 and an introduction to relevant aspects of the morphology of Georgian in § 2. The first experiment is discussed in § 3, and the second in § 4. Section 5 provides a concluding discussion.

1. The Hawkins-Cutler hypothesis and subsequent work

In a series of papers, Cutler, Gilligan, and Hawkins clarify word order and morpheme order universals and make a proposal to account for them. Hawkins & Gilligan (1988) show that languages in which the object precedes the verb in basic sentences (henceforth OV order) and ones in which the adposition regularly follows the noun phrase (NP + Po order) have suffixes. On the other hand, languages in which the object follows the verb in basic sentences (VO order) and ones in which the adposition regularly precedes the noun phrase (Pr + NP order) may have prefixes or suffixes. Hawkins & Gilligan (1988) suggest that the occurrence of suffixes with OV and NP + Po order and the occurrence of prefixes with VO and Pr + NP order is due to the Head Ordering Principle, the principle that heads are ordered identically in morphology and syntax. To account for the fact that OV and NP + Po languages are almost universally suffixing, while VO and Pr + NP languages are approximately evenly split between prefixing and suffixing, Cutler, Hawkins & Gilligan (1985) and Hawkins & Cutler (1988) propose that the ordering STEM – AFFIX is preferred because the greater salience of the beginning of the word makes it optimal for stems to occur there. Moreover, this order is said to mirror the order in which the parts of a word are processed.

As evidence that the beginnings of words are the most salient, Cutler et al. (1985) and Hawkins & Cutler (1988) cite previous work that shows that "beginning portions are the most effective cues for successful recall or recognition of a word …; and the effects of distorting the beginning of a word are much more severe than the effects of distorting later portions …" (1985: 131). In addition, Cutler et al. (1985) cite research that supports the concept of the 'uniqueness point', the point at which a word can be identified with certainty, and the importance of other 'left-to-right' phenomena (Marslen-Wilson & Welsh 1978; Marslen-Wilson 1980). They go on to cite evidence that while the beginning is most salient, the end of the word is next in salience, with the middle of the word being least salient. This evidence is persuasive, but none of the experiments cited by these authors in regard to this issue specifically relate to affixes. Most of the experimental items have neither suffixes nor prefixes, so the conclusions cannot relate to those affixes specifically.

None of the experiments specifically cited by Cutler et al. (1985) compare prefixes with suffixes. We know of no research that specifically tests their hypothesis that words with suffixes are preferred in processing over words with prefixes. While Cutler et al. do not report experiments of their own, their hypothesis predicts that suffixed words are easier to understand (faster and more accurate) than prefixed words. If it is to explain the suffixing preference, their hypothesis must predict that these results are true cross-linguistically.

Cutler et al. (1985: 724) mention that the number of languages on which experiments of the types needed have been carried out is extremely small, and they call for more research. Sixteen years later Feldman & Larabee (2001: 689) similarly note the need to study a variety of languages. "Claims about processing asymmetries between prefixed and suffixed forms ultimately need to be systematically evaluated in a variety of languages as well as in a variety of modality configurations and tasks so that language universals as well as the idiosyncrasies of particular languages and of particular tasks can be reconciled". Amenta and Crepaldi, in their 2012 review of written word identification, observe that there is (still) insufficient evidence that directly compares prefixes with suffixes. We would add that there is also insufficient evidence making this direct comparison with respect to spoken language. The small amount of direct evidence is described below.

A number of EEG studies of the effects of syntactic illformedness, mainly in German and Dutch, have reported an early anterior negativity when the word category is revealed by a prefix, with a relatively later negativity when the word category is marked by a suffix (e.g., Friederici, Hahne & Mecklinger 1996, which used both written and auditory stimuli; Friederici & Weissenborn 2007, a survey of studies using both auditory and written stimuli; van den Brink & Hagoort 2004, a study using auditory stimuli).[1] These EEG results have been interpreted as indicating that

1. Anterior negativity is brain activity characterized by negative-going waves generated by brain regions toward the front of the head. It is considered early if it peaks at around 200 milliseconds after a stimulus onset.

a prefixed word is processed more rapidly than a suffixed word (Bridgers & Kucinik 2017), contrary to the Hawkins-Cutler hypothesis. Other EEG studies found that word-category information is processed earlier than lexical-semantic information, even when the former is indicated by the suffix, and the latter by the stem (Friederici et al. 1996). Most of these studies used stimuli where the ill-formedness occurred in the context of a long, complex sentence.

Bridgers & Kucinik (2017) studied visual recognition of Italian derived words with the goal of learning about the interaction of affix position (prefix vs. suffix) and complexity. They argue that words containing category-changing (derivational) affixes are more complex than ones containing derivational affixes that are not category-changing. They find that in Italian, complex words containing non-category-changing affixes show greater differences between prefixes and suffixes than those that are category-changing. They find response times (RT) for prefixes longer (slower) than for suffixes and accuracy greater for suffixes than for prefixes. For nonce words with real Italian affixes, accuracy was higher for suffixed words than for prefixed words. RT was difficult to interpret for nonce words. Thus, their results support the Hawkins-Cutler hypothesis, which predicts that suffixes are easier to process (faster and more accurate) than prefixes.

The studies reviewed above are only somewhat relevant here, since most are studies of written words, and many studied affixed words in the context of complex sentences. Written words are generally presented all at once, while auditory words are necessarily presented sequentially, unfolding in time, as in natural speech. We have conducted our experiments on the effects of prefixes and suffixes in spoken language, since this is what could be relevant to explaining the suffixing preference, and since the Hawkins-Cutler hypothesis relates to the spoken word. The research we report here is on affixed words in simple, single-word statements.

Some other linguists have proposed to account for the suffixing preference in different ways. Givón (1971) proposes a historical explanation, and Hall (1988) proposes a historical interpretation of the Cutler et al. (1985) explanation. Hana & Culicover (2008) propose an acquisition explanation based on complexity, while St. Clair, Monaghan & Ramscar (2009) base their acquisition explanation on the relative efficiency of prefixes and suffixes in cuing of grammatical category. Himmelmann (2014) proposes a prosodic explanation of the same phenomenon. We do not feel that these proposals are necessarily incompatible with an explanation based on processing.

2. Overview of the morphology studied

Georgian has very flexible word order, with OV most basic, and with postpositions (NP + Po). It has very complex verbal morphology involving both prefixes and suffixes; most forms of most lexemes require two or more affixes. In order to construct minimal comparisons, we used in part verbal roots that can occur with a single affix in the present tense – the indicator of subject person and number. For other comparisons we were forced to use forms that contained additional affixes, extraneous from our point of view. We compared an inflectional prefix-suffix pair and a derivational pair, both in the verb.

The prefix *v-* marks a first person subject, while the suffix *-s* marks a third person singular subject. A paradigm for the present indicative for the verb types used here is given in Table 1, where the second person singular has no affix.

Table 1. Paradigm for present indicative for some verb types

	Singular	Plural
1st	*v*-STEM	*v*-STEM-*t*
2nd	STEM	STEM-*t*
3rd	STEM-*s*	STEM-*en*

Georgian also has agreement with direct and indirect objects, but since the third person direct object marking is zero, the forms in Table 1 represent complete sentences, even for verbs that require two arguments.

The prefixes *a-, i-, e-*, and *u-* occur in some forms with a variety of meanings, so that a stem may consist of any one of these prefixes (termed 'version vowels') and a root, e.g., *a-c'er* the stem "write on it", from *c'er* the root "write". A variety of suffixes occur on many verbs in the present indicative and certain other tense-aspect-mood forms (TAMs); these suffixes include *-eb, -ob, -av, -am*, and other less common ones. A present indicative stem may have a form such as *xat-av* "draw". Many verbs combine *a-* with *-eb* or *-ob*, such as *a-cx-ob* "bake". (The stem and root forms cited in this paragraph are also present tense forms with a 2nd person singular subject.)

The prefix *i-* (accompanied in certain TAMs by the suffix *-eb*) forms middles or passives. Grammars and the specialist literature generally treat this together with the suffix *-d* (also accompanied in certain TAMs by *-eb*) and with a third allomorph, zero (e.g., Mač'avariani 1959, Šaniʒe 1973: 280–313). All three are traditionally called formants of passives, though only some examples are passive. Inasmuch as each verb root generally occurs with only one of these three markers, they are considered allomorphs in the traditional literature. The prefix *e-* replaces *i-* in passives having an indirect object, not illustrated here or occurring in our critical stimuli. Table 2 gives examples of all three types.

Table 2. Examples of the three types of 'passives' (Šaniʒe 1973: 288–290)

Root	Present tense of 'passive'	Related active
-c'er- "write"	i-c'er-eb-a "it is written"	c'er-s "he, she, it writes"
-k'et- "make, do"	k'et-d-eb-a "it is made, done"	a-k'et-eb-s "he, she, it makes, does"
-tb- "heat"	tb-eb-a "it is heated"	a-tb-ob-s "he, she, it heats"

The column labeled "Present tense of 'passive'" gives the grammatical form available for the roots listed. The affixes *-eb-a* cannot be considered formants of the passive, since (i) they occur with other verbs that are not passive (see preceding paragraph), and (ii) they do not occur in all forms of the passive; for example, *da-i-c'er-o-s* "may it be written", *ga-v-k'et-d-i* "I was made, done", *ga-tb-nen* "they were heated". (Rather, *-eb* occurs in certain TAM forms, and *-a* is a marker of third person singular subjects with verbs of certain types in certain TAM forms.)

According to the method developed in Dryer (2013), the inflectional suffixing index for Georgian is six, or 75% of the inflectional affixing index, making it a language with a 'moderate preference for suffixing'. The way this is worked out is shown in Appendix G. If this is valid, we might expect a suffixing advantage in experiments with inflectional morphology in Georgian.

This section summarizes only the Georgian morphology that is relevant to the issues raised here.

3. Experiment 1

The first experiment is a simple lexical decision experiment, in which subjects are asked to distinguish between real words and nonce words. Lexical decision is a standard way to measure the speed and accuracy of word recognition. In this case our goal is to discover whether subjects identify words with a prefix more or less quickly and accurately than a word with a suffix.

3.1 Materials

We compiled a set of stimuli that would be as simple as possible, so that, to the extent allowed by the language, our results would not be tainted by the presence of extraneous morphology.

We compared the verbal prefix *v-*, first person subject, with the suffix *-s*, third person singular subject. Some linguists will object that these are not comparable, since the first person is a speech act participant, while the third person is not. They are, however, distinct inflectional morphemes, one a prefix and the other a suffix.

Each is a single consonant, and neither adds a syllable. For example, from the verb *bana* (lemma) "bathe", we used the stimuli *v-ban* "I bathe" and *ban-s* "he, she, or it bathes". We compiled 24 prefixal items of this sort and 24 suffixal; this set of critical items is given in Appendix A. We constructed a set of nonce words of the same size using the real prefix *v-* and the real suffix *-s*. The complete set of nonce words is presented in Appendix B.[2] Together there are 96 items in this set.

These 48 critical words and 48 matched nonce items were randomly mixed in with an additional 96 real words and 96 nonce items, for a total of 144 real words and 144 nonce items.

A number of studies have shown that speakers' estimates of frequency are reasonably accurate and can be substituted for text-based calculations (Feldman, Frost & Pnini 1995; Schreuder & Baayen 1997; Kgolo & Eisenbeiss 2015). We gave a list of the real-word stimuli (among a set of additional words) to five randomly selected subjects and asked them to rate the stimuli for frequency on a scale of 1–5, very infrequent to very frequent. The order of items was randomized for each subject. The ratings of one participant had to be eliminated because it was clear that the subject had reversed the scale. The averages of the remaining four subjects are listed for each item by lemma in Appendix A.

All experimental items were recorded by a native speaker of Georgian using a Tascam DR-40 digital recorder and a Shure SM-10A-CM head-mounted microphone in a sound booth.

3.2 Subjects and methods

The experimental subjects were 51 native speakers of Georgian who live in the capital city, Tbilisi, or its environs. Three of these were excluded from the analysis because their error rates were very much higher than those of all of the other subjects. We did not collect demographic data, but 12 appeared to be male, and all except two appeared to be in their 20's. All subjects had self-declared normal hearing. Participants were paid $5 or the equivalent in Georgian currency. Experiments were conducted in a room of the University of Tbilisi, loaned to us for this purpose by the Georgian Language Institute. Experiment 1 was the first of five experiments in which these subjects participated on the same day.

We used aural stimuli, which subjects heard over earphones. Subjects pressed one labeled computer key for a "yes, this is a real word" answer and a different labeled key for a "no, this is not a real word" response. Items were randomized for

2. Eight nonce words were excluded from the analysis due to experimenter error. They are marked with # in Appendix B.

each subject. Subjects heard only Georgian from the time they entered the room.[3] After they were instructed in how to do the experiment, we ran a short practice session. The practice contained 16 items, 10 of them real words and 6 nonce words. One real word had the prefix *v*-, but the verb stem was more complex than that of the critical items in the experiment. None had the suffix -*s*.

3.3 Results and discussion

The results of this experiment are summarized in Figure 1. The left-hand panel shows that prefixed words had an error rate of 29.2%, while the same lexemes with suffixes had an error rate of 33.5%. The right-hand panel indicates reaction times of 1353 ms for prefixed words and 1392 ms for suffixed words. Thus, prefixed words were identified more accurately and more quickly than suffixed words. The difference in accuracy was statistically significant, $F(1, 47) = 9.93$, $p < .05$, and the response time difference was marginally so, $F(1, 47) = 3.58$, $p < .07$.

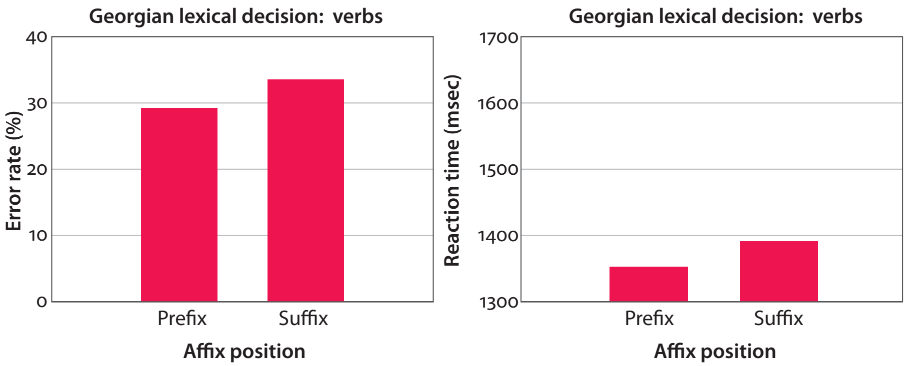


Figure 1. Mean error rates and reaction times for simplex real words with the inflectional prefix *v*- and the inflectional suffix -*s*

Figure 2 shows the corresponding results for the nonce items. Overall, response times were noticeably slower than they were for the real words, $F(1, 47) = 13.41$, $p < .05$. This difference is quite typical of what is found in many lexical decision tasks. The much lower error rate for the nonce items than for real words, $F(1, 47) = 100.62$, $p < .05$), is much less common, and presumably reflects the clearly illegal patterns in the nonce items. It may also reflect an issue with the word stimuli that we will

3. There was one exception who insisted on speaking English at some point during the experiments, and who was then answered in English.

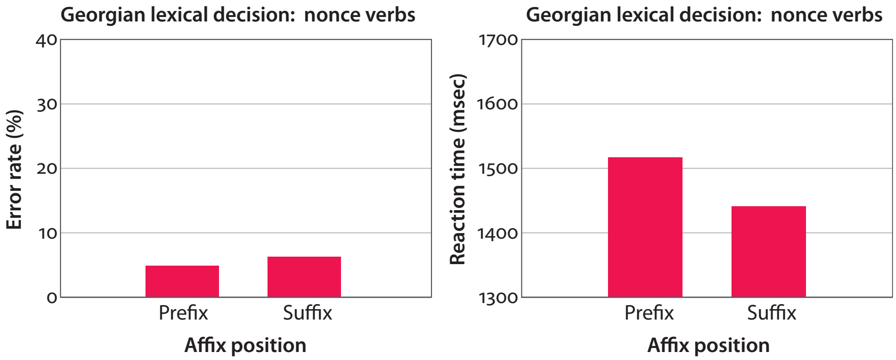


Figure 2. Mean error rates and reaction times for simplex nonce words with the inflectional prefix *v-* and the inflectional suffix *-s*

discuss shortly. The significantly slower responses for nonce items with a prefix than with a suffix, $F(1, 47) = 19.11$, $p < .05$), suggests that when listeners heard a legal onset (prefix), rejecting the item was more difficult than if they had heard the illegal pattern before encountering a legal suffix.

Experiment 1 was designed to assess the Hawkins-Cutler hypothesis using the most direct test available in Georgian: verbs that differed from each other by virtue of having either the prefix *v-* or the suffix *-s*. Their hypothesis would predict that listeners would have more difficulty recognizing words with a prefix, as the prefix potentially disrupts the critical onset portion of the words. Instead, we observed the reverse: the prefixed words were processed more accurately and more quickly than the suffixed words.

Given this outcome, we thought it would be prudent to conduct additional tests of the hypothesis. In order to have the most direct possible test in Experiment 1 (i.e., a single-phoneme affix that either precedes or follows the stem), we had to make some compromises. For example, the lemma frequencies of some of the verbs used in this experiment were quite low (see Appendix A). In addition, some of the simplex verbs in Georgian have begun to be used with a suffix *-av*, and in order to create the most direct test, we preferred to avoid items with extraneous affixes. For example the verb *breg-s* is no longer used; instead speakers use *breg-av-s* "falls with a crash". We had a consultant approve the words in our stimuli, but we learned later that some of our stimuli were more common with the suffix *-av*, possibly leaving the subject to wonder whether the stimulus lacking the suffix was a real word or not. This evolution is a plausible reason for the relatively high error rate for the real words. Therefore, we conducted a second experiment, one that included two additional tests of the Hawkins-Cutler hypothesis. These tests involved more complex stimuli than those used in Experiment 1.

4. Experiment 2

Like Experiment 1, this was a lexical decision experiment, but the stimuli were differently structured. We constructed one set of stimuli that used the same pair of affixes used in Experiment 1 (the prefix *v-* versus the suffix *-s*), but the new stimuli had additional required affixes. We built a second stimulus set based on the prefix *i-* versus the suffix *-d*, with this second set also including required additional affixes. For both sets of stimuli, the core question is as before: Do items with prefixes pose a greater challenge in recognition because their (by hypothesis, critical) onsets are disrupted?

4.1 Materials

Complex stems with v- *or* -s
The stimuli in the first set are similar to those in Experiment 1, except that these have an additional prefix and suffix in all forms – the prefix *a-* or *i-* and one suffix from the set *-eb, -ob, -am, -av, -ev*. For example, among the critical items in this set are *v-a-k'et-eb* "I do, make" and *a-k'et-eb-s* "she, he, it does, makes", each with the prefix *a-* and the suffix *-eb*, in addition to the critical affixes. All forms are transitive active in the present indicative. A group of 24 items in the first person subject form with the prefix *v-* are paired with another 24 items in the third person singular subject form with the suffix *-s*, as in Experiment 1. These are matched with nonce words with the same real affixes. Appendices C and D list these 48 real and 48 nonce stimuli, respectively.

Derivational prefix i- *or suffix* -d
Stimuli in the second set are composed of items formed with the prefix *i-* or the suffix *-d*. Both affixes require the additional suffixes *-eb* and *-a*. Since they are the same in this regard, the prefix *i-* and suffix *-d* are directly comparable. However, the prefix introduces an additional syllable, while the suffix does not. To control for this, we constructed 24 pairs matched for number of segments in the root, and 24 pairs matched for number of syllables in the word. For example, *i-sm-eb-a* "is drinkable" has only two consonants in its root, *-sm-,* and *xv-d-eb-a* "meets someone" has the same number in its root, *xv-*. Through investigator error, two of the pairs do not match as intended. On the other hand, the pair *i-nt-eb-a* "lights" and *mcir-d-eb-a* "is reduced, diminished" each have three syllables. All items in both subsets have four morphemes. All 48 pairs of stimuli in this set are listed in Appendix E. We constructed nonce items that were matched to these words, and all 96 nonce items are listed in Appendix F. Combined, items in this set comprise 192 experimental items. Since there is some discussion in the literature (e.g., Bridgers & Kacinik 2017)

suggesting a difference between affixes that trigger changes in word category and those that do not, it is worth noting that neither *i-* nor *-d* conditions a change in word category.[4]

In sum, there were 48 words and 48 nonce words (with complex stems) based on *v-* and *-s,* and 192 items based on *i-* or *-d.* These 288 critical items were combined with experimental items on another issue, additional nonce words, and real-word fillers for a total of 528 experimental items in this experiment. Half of these were real words, and half were nonce words.

After they had completed the lexical decision task, we asked three subjects to rate each real-word item in the experiment for frequency using the same 1–5 scale used in Experiment 1. Lists were randomized for each subject, and the average of these subjects' ratings is listed for each item in Appendices C and E.

4.2 Subjects and methods

Subjects for this experiment were 41 native speakers of Georgian living in Tbilisi or nearby, none the same as the earlier subjects. Data from four were not included in the analyses due to performance that was at or near chance. Experiments were conducted in the same room as before, but this was the only experiment this time. Subjects were paid 10 GEL each (a little less than $5).

Subjects heard only Georgian from the time they entered the room. Stimuli were again presented aurally over headphones. Subjects were instructed to press one labeled computer key to indicate a real word and another for a nonce word. They were given a practice session, comprising 12 items, half of these nonce words. Two real items and two nonce items were from the set of complex items with *v-* and *-s*, including a prefixed and a suffixed version of the same stem. One item with a derivational prefix *i-* was included, and one with the derivation suffix *-d*. Remaining items matched items that had been included as part of another issue.

4. There may be a difference of opinion on this. Some sources say that a verb such as *civ-d-eb-a* "becomes cold" derives from the associated adjective *civ-i* "cold", while others analyze it as derived from the transitive active *a-civ-eb-s* "makes it cold". We take the position that the passives are derived from the transitive actives because of verbs such as *suk-d-eb-a* "becomes fat" with the related *a-suk-eb-s* "makes s.t. fat". Neither form contains the whole of the associated adjective, *msukan-i* "fat", which contains an additional /m/ word-initially and an additional /an/ stem-finally. Some, such as *xv-d-eb-a* "meets s.o." correspond to neither a transitive active verb nor to an adjective or noun.

4.3 Results and discussion

Complex stems with v- *or* -s

As the left side of Figure 3 shows, accuracy was much higher for these longer words than it had been for the simpler words in Experiment 1. There was a small but significant accuracy advantage for the suffixed items, $F(1, 36) = 10.17$, $p < .05$. Response times did not differ significantly as a function of affix position, $F(1, 36) = 1.73$, n.s.

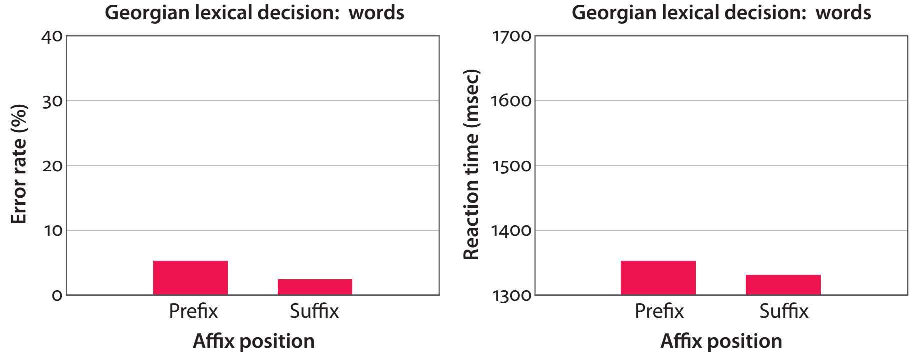


Figure 3. Mean error rates and reaction times for real complex verbs with the inflectional prefix *v*- and the inflectional suffix -*s*

Figure 4 shows the corresponding results for rejecting the nonce items. Again, performance was very accurate, with low error rates for both prefixed and suffixed items. There was no difference in accuracy between the two stimulus types, $F(1, 36) = 2.68$, n.s. In contrast, response times were significantly different, $F(1, 36) = 20.18$, $p < .05$. Note that unlike what we found for the simpler items in Experiment 1, listeners rejected prefixed nonce items more quickly than suffixed ones. It may be that in the context of hearing many complex forms in this experiment (versus simpler forms in Experiment 1), an initial legal prefix was not as strong a cue for lexicality.

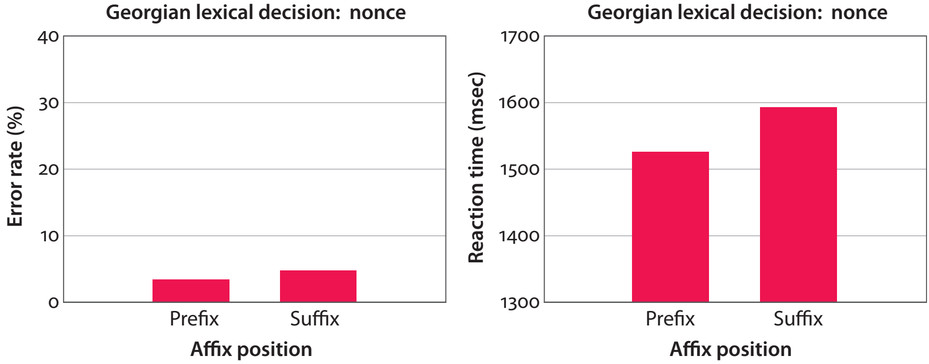


Figure 4. Mean error rates and reaction times for nonce complex verbs with the inflectional prefix *v*- and the inflectional suffix -*s*

Derivational prefix i- *or suffix* -d

Figures 5 and 6 show the accuracy and reaction time results for the complex words that either had the prefix *i*- or the suffix -*d*. Recall that because the prefix introduces an additional syllable, we constructed one set of stimuli for which the words were matched in the number of segments in the root (Figure 5), and one set of stimuli in which the words matched in the number of syllables (Figure 6).

The results for the root-length matched words are very simple to summarize: there was no difference between such words with a prefix and such words with a suffix. This was true both for accuracy, $F(1, 36) = 1.05$, n.s., and for reaction time, $F(1, 36) = 1.63$, n.s.

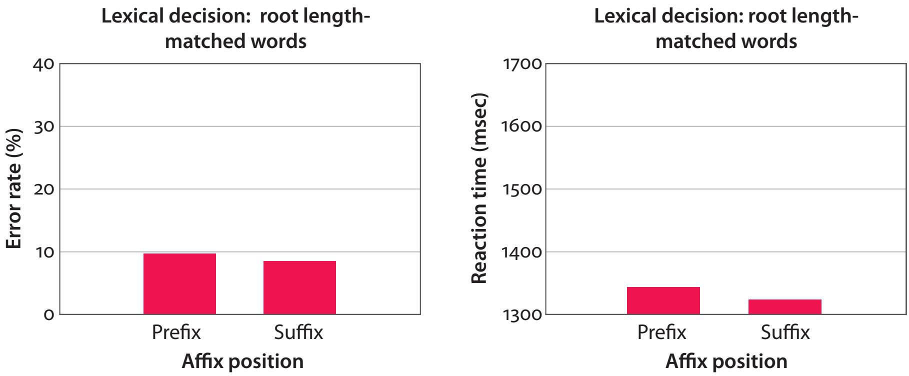


Figure 5. Mean error rates and reaction times for complex real words derived with the prefix *i*-compared with ones derived with the suffix -*d*
(Stimuli were matched on the number of segments in the root.)

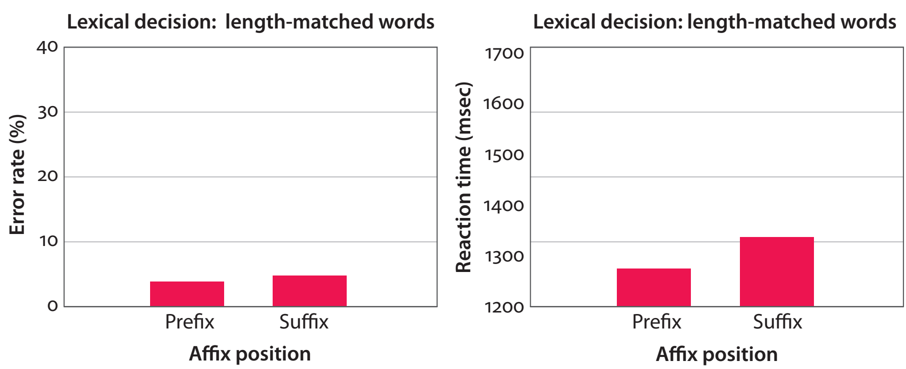


Figure 6. Mean error rates and reaction times for complex real words derived with the prefix *i*-compared with ones derived with the suffix -*d*
(Stimuli were matched on the number of syllables in the word.)

The same lack of a difference was observed in the accuracy results for the words that were matched on the number of syllables, $F(1, 36) = 1.21$, n.s. There was, however, a difference in reaction time for these items, with prefixed items responded to more quickly than suffixed items, $F(1, 36) = 18.54$, $p < .05$. Note that this difference runs counter to the Hawkins-Cutler hypothesis.

Turning to the results for the nonce items (Figures 7 and 8), we find the same pattern of either no difference, or of better performance for prefixed stimuli. For the stimuli matched on the number of segments in the root, accuracy was the same for prefixed and the suffixed items, $F(1, 36) = 0.73$, n.s., while response times were faster for prefixed than for suffixed items, $F(1, 36) = 19.43$, $p < .05$. This advantage

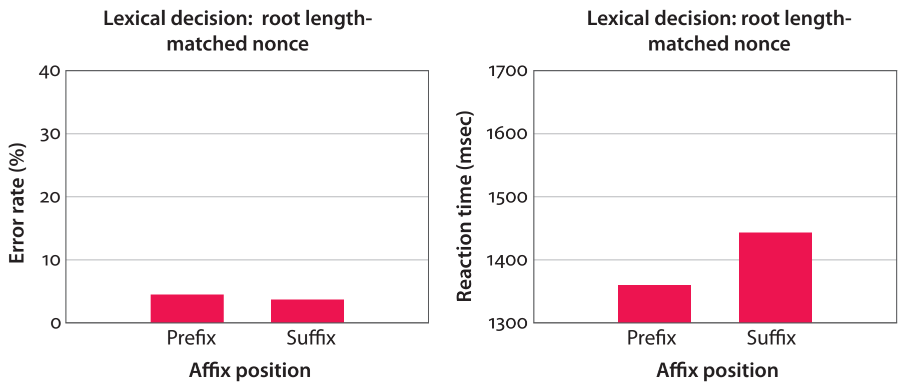


Figure 7. Mean error rates and reaction times for complex nonce words derived with the prefix *i*-compared with ones derived with the suffix *-d*
(Stimuli were matched on the number of segments in the root.)

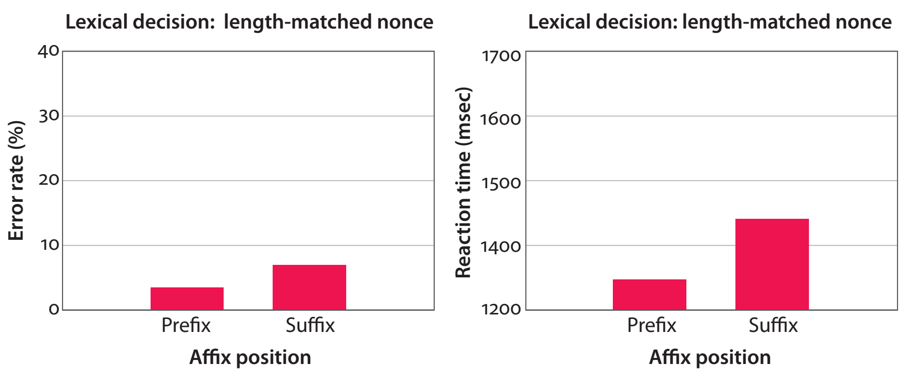


Figure 8. Mean error rates and reaction times for complex nonce words derived with the prefix *i*-compared with ones derived with the suffix *-d*
(Stimuli were matched on the number of syllables in the nonce word.)

for prefixed items was also found in both accuracy, $F(1, 36) = 11.73$, $p < .05$, and reaction time, $F(1, 36) = 22.26$, $p < .05$, for the nonce items that were matched on the number of syllables.

5. Concluding discussion

As described in the Introduction, we are engaged in a project that is motivated by Hawkins & Cutler's (1988) hypothesis. That hypothesis is intended to explain the observation that suffixes are more common than prefixes across languages. The Hawkins-Cutler hypothesis states that processing of suffixed words should be easier than processing of prefixed words because prefixes potentially disrupt processing of a word's most critical location – the onset. There are some prior psycholinguistic studies that are consistent with the hypothesis. However, most of the existing psycholinguistic evidence comes from one language – English – and a substantial portion of this evidence comes from studies of printed rather than spoken words.

Given these limitations, our project includes tests of word recognition in languages other than English – Georgian, European Portuguese, and Udi. The current paper presents our initial findings from the Georgian lexical decision experiments. Our approach was to construct matched items that contrasted in the presence of a prefix (*v*- or *i*-) versus a suffix (-*s* or -*d*). Any given set of stimuli is likely to have some complication, so we developed three different contrasting sets, and in each set we tested both real Georgian words and nonce words that were built with legal Georgian affixes. Our stimuli included both inflectional and derivational morphology.

One set of stimuli was chosen to offer the simplest and most direct test of prefixes versus suffixes – the same items were either prefixed with *v*-, or suffixed with -*s*. A second set preserved these affixes, but also included additional affixes. In the third case, we contrasted the prefix *i*- with the suffix -*d*; this case also included additional (matched) affixes. Despite the substantial variation among the stimulus sets, the results were rather stable, and problematic for the Hawkins-Cutler hypothesis. Considering both stimuli with real-word bases, and those with nonce bases, we have data for prefixes versus suffixes for eight stimulus sets (shown in Figures 1–8). For these eight tests, only two showed a suffix advantage (the nonce stimuli in Experiment 1, using the reaction time measure, and the real word stimuli with complex verbs and *v*- and -*s* in Experiment 2). One test showed no difference (the root-length-matched words for the *i*- versus -*d* stimuli). In all five of the remaining tests, there was actually a significant advantage for the prefixed items over the suffixed items on the accuracy and/or response time measure. That is, our results overall pattern in the opposite direction from what the Hawkins-Cutler hypothesis predicts.

Clearly, we need to consider the results from the broader set of data we have been collecting, with multiple tasks and multiple languages. Just as clearly, these initial results provide no support at all for the universality of the Hawkins-Cutler hypothesis.

Funding

The research reported here was supported in part by grant number BCS 1729256 from the National Science Foundation. Additional support was provided by the Economic and Social Research Council (UK) grant #ES/R006288/1, the Ministerio de Ciencia E Innovación (Spain) grant #PSI2017-82563-P, the Basque Government through the BERC 2018–2021 program and by the Spanish State Research Agency through BCBL Severo Ochoa excellence accreditation SEV-2015-0490, by Ayuda Centro de Excelencia Severo Ochoa (Spain) SEV-2015-0490, and by the University of Massachusetts Amherst.

References

Amenta, Simona & Davide Crepaldi. 2012. Morphological processing as we know it: An analytical review of morphological effects in visual word identification. *Frontiers in Psychology* 3. 232–243. https://doi.org/10.3389/fpsyg.2012.00232

Bertram, Raymond, Matti Laine & Katja Karvinen. 1999. The interplay of word formation type, affixal homonymy, and productivity in lexical processing: Evidence from a morphologically rich language. *Journal of Psycholinguistic Research* 28. 213–226. https://doi.org/10.1023/A:1023200313787

Bridgers, Franca Ferrari & Natalie Kacinik. 2017. How linearity and structural complexity interact and affect the recognition of Italian derived words. *Journal of Psycholinguistic Research* 46. 175–200. https://doi.org/10.1007/s10936-016-9427-1

Cutler, Anne. 1985. Cross-language psycholinguistics. In Ulrich Hans Frauenfelder & Anne Cutler (eds.), special issue of *Linguistics* 23. 659–667. https://doi.org/10.1515/ling.1985.23.5.659

Cutler, Anne. 2009. Psycholinguistics in our time. In Patrick Rabbitt (ed.), *Inside psychology: A science over 50 years*, 91–101. Oxford: Oxford University Press.

Cutler, Anne, John A. Hawkins & Gary Gilligan. 1985. The suffixing preference: A processing explanation. *Linguistics* 23. 723–758. https://doi.org/10.1515/ling.1985.23.5.723

Dryer, Matthew S. 2013. Prefixing vs. suffixing in inflectional morphology. In Matthew S. Dryer & Martin Haspelmath (eds.), *World Atlas of Language Structure*, Ch. 26. https://wals.info/chapter/26. (Accessed 21 July 2019).

Feldman, Laurie Beth, Ram Frost & Tamar Pnini. 1995. Decomposing words into their constituent morphemes: Evidence from English and Hebrew. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21. 947–960.

Feldman, Laurie Beth & Jacqueline Larabee. 2001. Morphological facilitation following prefixed but not suffixed primes: Lexical architecture or modality-specific processes? *Journal of Experimental Psychology: Human Perception and Performance* 27. 680–691.

Friederici, Angela D., Anja Hahne & Axel Mecklinger. 1996. Temporal structure of syntactic parsing: Early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22. 1219–1248.

Friederici, Angela D. & Jürgen Weissenborn. 2007. Mapping sentence form onto meaning: The syntax-semantic interface. *Brain Research* 1146. 50–58. https://doi.org/10.1016/j.brainres.2006.08.038

Givón, Talmy. 1971. Historical syntax and synchronic morphology: An archaeologist's field trip. *Chicago Linguistic Society* 7. 394–415.

Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of language*, 73–113. Cambridge, MA: MIT Press.

Hall, Christopher. 1988. Integrating diachronic and processing principles in explaining the suffixing preference. In John A. Hawkins (ed.), *Explaining language universals*, 321–349. Oxford: Blackwell.

Hana, Jirka & Peter W. Culicover. 2008. Morphological complexity outside of universal grammar. *Ohio State University Working Papers in Linguistics*, https://www.asc.ohio-state.edu/culicover.1/Publications/Hana%20&%20Culicover%20-%20morphological%20complexity.pdf. (Accessed 20 July 2019).

Hawkins, John A. & Anne Cutler. 1988. Psycholinguistic factors in morphological asymmetry. In John A. Hawkins (ed.), *Explaining language universals*, 280–317. Oxford: Blackwell.

Hawkins, John A. & Gary Gilligan. 1988. Prefixing and suffixing universals in relation to basic word order. *Lingua* 74. 219–259. https://doi.org/10.1016/0024-3841(88)90060-5

Himmelmann, Nikolaus P. 2014. Asymmetries in the prosodic phrasing of function words: Another look at the suffixing preference. *Language* 90. 927–960. https://doi.org/10.1353/lan.2014.0105

Kgolo, Naledi & Sonja Eisenbeiss. 2015. The role of morphological structure in the processing of complex forms: Evidence from Setswana deverbative nouns. In Alice C. Harris, T. Florian Jaeger & Elisabeth Norcliffe (eds.), *Laboratory in the field: Advances in cross-linguistic psycholinguistics, special issue of Language, Cognition and Neuroscience* 30. 1116–1133.

Mač'avariani, Givi. 1959. 'Unišno vnebiti' kartvelur enebši. [The 'markerless passive' in the Kartvelian languages]. *Kartvelur enata st'rukt'uris sak'itxebi* 1. 101–129.

Marslen-Wilson, William D. 1980. Speech understanding as a psychological process. In Jean Claude Simon (ed.), *Spoken language generation and understanding*, 39–67. Dordrecht: Reidel. https://doi.org/10.1007/978-94-009-9091-3_2

Marslen-Wilson, William D. & Alan Welsh. 1978. Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology* 10. 29–63. https://doi.org/10.1016/0010-0285(78)90018-X

Šanize, Ak'ak'i. 1973 [1953]. *Kartuli enis gramat'ik'is sapuzvlebi*. [Fundamentals of the grammar of the Georgian language.] Tbilisi: Universit'et'i.

Schreuder, R. & R. Harald Baayen. 1997. How complex simplex words can be. *Journal of Memory and Language* 37. 118–139. https://doi.org/10.1006/jmla.1997.2510

St. Clair, Michelle C., Padraic Monaghan & Michael Ramscar. 2009. Relationships between language structure and language learning: The suffixing preference in grammatical categorization. *Cognitive Science* 33. 1317–1329. https://doi.org/10.1111/j.1551-6709.2009.01065.x

van den Brink, Daniëlle & Peter Hagoort. 2004. The influence of semantic and syntactic context constraints on lexical selection and integration in spoken-word comprehension as revealed by ERPs. *Journal of Cognitive Neuroscience* 16. 1068–1084. https://doi.org/10.1162/0898929041502670

Appendix A. Real verbs with prefix *v-* or suffix *-s*

Pair	Root	Meaning	Prefixed	Suffixed	Lemma frequency
1	ban	bathe	v-ban	ban-s	3.5
2	brʒmed	smelt	v-brʒmed	brʒmed-s	0.75
3	čeč	comb	v-čeč	čeč-s	3
4	čkmetʼ	pinch	v-čkmetʼ	čkmetʼ-s	4.5
5	cvet	wear out	v-cvet	cvet-s	3.75
6	drekʼ	bend	v-drekʼ	drekʼ-s	3
7	pen	spread	v-pen	pen-s	4.5
8	pxan	scratch	v-pxan	pxan-s	3.75
9	gles	smear	v-gles	gles-s	1
10	kʼbeč	bite	v-kʼbeč	kʼbeč-s	4.5
11	kʼrex	graze	v-kʼrex	kʼrex-s	1
12	kleš	wear down	v-kleš	kleš-s	1.5
13	kšen	puff, pant	v-kšen	kšen-s	3.25
14	ɣren	gumble	v-ɣren	ɣren-s	4.75
15	švlepʼ	hit	v-švlepʼ	švlepʼ-s	1
16	tʼex	break	v-tʼex	tʼex-s	4
17	tʼqʼepʼ	slap	v-tʼqʼepʼ	tʼqʼepʼ-s	3
18	cʼmend	clean	v-cʼmend	cʼmend-s	5
19	xvretʼ	perforate	v-xvretʼ	xvretʼ-s	3.75
20	xleč	split	v-xleč	xleč-s	3
21	ɣrec	warp	v-ɣrec	ɣrec-s	2.25
22	pxekʼ	plane	v-pxekʼ	pxekʼ-s	2.75
23	tʼlekʼ	lick	v-tʼlekʼ	tʼlekʼ-s	1.25
24	pʼrex	twist	v-pʼrex	pʼrex-s	2.5

Appendix B. Nonce verbs with prefix *v-* or suffix *-s*

Pair	Prefixed	Suffixed
1	v-bɣen	bɣen-s
2	v-bzečʼ	bzečʼ-s
3	v-pʼcʼkʼekʼ	pʼcʼkʼekʼ-s
4	v-čle čʼ	čle čʼ-s
5	v-xvepʼ	xvepʼ-s
6	v-šxed	šxed-s
7	v-pʼtʼer	pʼtʼer-s
8	v-pler	pler-s
9	v-tʼqʼen	tʼqʼen-s
10	v-šrek	šrek-s
11	v-čkʼvetʼ	čʼkʼvetʼ-s
12	v-kʼret	kʼret-s

Pair	Prefixed	Suffixed
13	v-čkretʼ	čkretʼ-s
14	v-čvlem	čvlem-s
15	v-cvetʼ	cvetʼ-s
16	v-čʼel	čʼel-s
17	v-čʼqʼekʼ	#čʼqʼetʼ-x
18	v-xvletʼ	#xvlepʼ-č
19	v-xvrekʼ	#xvrepʼ-x
20	v-qʼet	#qʼep-k
21	v-qʼvlek	#qʼvlep-c
22	v-znep	#znek-č
23	v-ʒob	#ʒov-x
24	v-kʼber	#kʼber-č

Appendix C. Real complex verbs with prefix *v*- and suffix -*s*

Pair	Meaning	Prefixed forms	Form frequency	Suffixed forms	Form frequency
1	do, make	v-a-k'et-eb	4.7	a-k'et-eb-s	4.7
2	bake	v-a-cx-ob	3.3	a-cx-ob-s	4
3	blame	v-a-bral-eb	3.7	a-bral-eb-s	4
4	hinder	v-a-brk'ol-eb	3.3	a-brk'ol-eb-s	4
5	prepare	v-a-mzad-eb	4.3	a-mzad-eb-s	4.3
6	flatten	v-a-brt'q'el-eb	3	a-brt'q'el-eb-s	2.7
7	tie up	v-a-xv-ev	3	a-xv-ev-s	3
8	turn	v-a-brun-eb	3.3	a-brun-eb-s	3.3
9	spin	v-a-t'rial-eb	3.3	a-t'rial-eb-s	3.7
10	praise	v-a-k-eb	4.7	a-k-eb-s	5
11	lengthen	v-a-grʒel-eb	4.7	a-grʒel-eb-s	4
12	strengthen	v-a-ʒlier-eb	4	a-ʒlier-eb-s	4.3
13	diminish	v-a-k'l-eb	3.3	a-k'l-eb-s	3.7
14	darken	v-a-muk-eb	2.3	a-muk-eb-s	2.3
15	sharpen	v-a-c'vet'-eb	3	a-c'vet'-eb-s	2.7
16	sharpen	v-i-les-av	1.3	i-les-av-s	1.7
17	set up	v-a-c'q'-ob	4.7	a-c'q'-ob-s	4.7
18	invite	v-i-c'v-ev	3.3	i-c'v-ev-s	4.7
19	put in, on	v-a-dg-am	2.3	a-dg-am-s	3
20	raise	v-a-maɣl-eb	1.7	a-maɣl-eb-s	3.3
21	praise	v-a-did-eb	3.3	a-did-eb-s	3
22	blacken	v-a-šav-eb	3.7	a-šav-eb-s	3
23	adorn	v-a-lamaz-eb	3.3	a-lamaz-eb-s	4
24	clean	v-a-suptav-eb	4.3	a-suptav-eb-s	4.3
Frequency averages			3.4		3.6

Appendix D. Complex nonce words with *v*- and -*s*

Pairs	Prefixed forms	Suffixed forms
1	v-a-pšet-eb	a-pšet-eb-s
2	v-a-cr-ob	a-cr-ob-s
3	v-a-bran-eb	a-bran-eb-s
4	v-a-brt'ol-eb	a-brt'ol-eb-s
5	v-a-tx-ev	a-tx-ev-s
6	v-a-prt'el-eb	a-prt'el-eb-s
7	v-a-xn-ev	a-xn-ev-s
8	v-a-brum-eb	a-brum-eb-s
9	v-a-txial-eb	a-txial-eb-s
10	v-a-brʒal-eb	a-brʒal-eb-s
11	v-a-čx-am	a-čx-am-s
12	v-a-ʒlian-eb	a-ʒlian-eb-s

Pairs	Prefixed forms	Suffixed forms
13	v-a-p'l-eb	a-p'l-eb-s
14	v-a-mup-eb	a-mup-eb-s
15	v-a-c'vep'-eb	a-c'vep'-eb-s
16	v-i-ves-av	i-ves-av-s
17	v-a-p'q'-ob	a-p'q'-ob-s
18	v-i-c'k'-ev	i-c'k'-ev-s
19	v-a-bd-am	a-bd-am-s
20	v-a-gaɣl-eb	a-gaɣl-eb-s
21	v-a-rib-eb	a-rib-eb-s
22	v-a-šap-eb	a-šap-eb-s
23	v-a-kamar-eb	a-kamar-eb-s
24	v-a-raptiv-eb	a-raptiv-eb-s

Appendix E. Real words with prefix *i-* or suffix *-d*

Length-matched for root

Pair	Transliteration	Meaning	C's in root	Form frequency	Transliteration	Meaning	C's in root	Form frequency
1	i-sm-eb-a	is drinkable	4	3	xv-d-eb-a	meets s.o.	2	4
2	i-ʒarcv-eb-a	is robbed	2	3.3	ten-d-eb-a	dawn breaks	2	4.3
3	i-šov-eb-a	is gotten, obtained	2	3.3	duɣ-d-eb-a	begins to boil	2	3.7
4	i-ɣeb-eb-a	is painted	2	2.7	muk-d-eb-a	gets dark	2	3
5	i-q'id-eb-a	sells, is for sale	2	4.3	k'et-d-eb-a	gets done	2	5
6	i-k'et'-eb-a	is locked	2	5	civ-d-eb-a	becomes cold	2	3.7
7	i-xur-eb-a	shuts, is covered	2	3.7	ɣam-d-eb-a	night falls	2	4
8	e-jax-eb-a	lashes out at s.o., s.t.	2	4.3	suk-d-eb-a	puts on weight	2	4
9	i-xarš-eb-a	cooks	3	3.7	rbil-d-eb-a	becomes soft	3	2.7
10	i-potl-eb-a	leafs	3	2.3	maɣl-d-eb-a	becomes tall(er), high(er)	3	2.7
11	i-k'umš-eb-a	is compressed, contracted	3	2.3	gril-d-eb-a	becomes cool	3	3
12	i-k'itx-e-ba	is legible	3	3.3	c'itl-d-eb-a	becomes red, s/he blushes	3	2.7
13	i-c'er-eb-a	is written	2	3.3	ber-d-eb-a	grows old	2	4
14	i-mal-eb-a	is hidden	2	3.3	tov-d-eb-a	begins to snow	2	2.3
15	i-p'ov-eb-a	is found, located	2	1.3	c'ux-d-eb-a	becomes sad, sorry	2	4.3
16	i-k'arg-eb-a	gets lost	3	4.3	magr-d-eb-a	gains strength	3	3.3
17	i-xark'-eb-a	is made to pay tribute	3	2	ʒvir-d-eb-a	becomes (more) expensive	3	4.7
18	i-xvet'-eb-a	is cleaned out (of a stable)	3	3.3	rtul-d-eb-a	become complicated	3	3.3
19	i-rkin-eb-a	butts (e.g. calf)	3	1.3	mc'ip-d-eb-a	ripens	3	4
20	i-žlit'-eb-a	is destroyed, massacred	3	1.7	ɣel-d-eb-a	becomes rough (sea), agitated	2	2
21	i-k'rib-eb-a	assembles	3	3.7	mʒim-d-eb-a	becomes heavi(er)	3	4
22	i-c'vel-eb-a	is milked	3	2.7	mc'ar-d-eb-a	becomes (more) bitter	3	3
23	i-smin-eb-a	is heard	3	3	bnel-d-eb-a	gets dark	3	4
24	i-gzavn-eb-a	is sent	4	3.7	kveq'n-d-eb-a	is published, made public	4	3.7
Frequency averages				3.5				3.5

Appendix F. Nonce words with *i-* and *-d*

Length-matched for root

Pair	Prefixed	C's in root	Suffixed	C's in root
1	i-sl-eb-a	2	sl-d-eb-a	2
2	i-nes-eb-a	2	ɣen-d-eb-a	2
3	i-šox-eb-a	2	dus-d-eb-a	2
4	i-reb-eb-a	2	ruk-d-eb-a	2
5	i-t'ib-eb-a	2	t'eg-d-eb-a	2
6	i-t'eg-eb-a	2	cis-d-eb-a	2
7	i-run-eb-a	2	ɣan-d-eb-a	2
8	e-jar-eb-a	2	sup-d-eb-a	2
9	i-xačx-eb-a	3	brug-d-eb-a	3
10	i-totl-eb-a	3	masx-d-eb-a	3
11	i-šk'un-eb-a	3	grik'-d-eb-a	3
12	i-p'itx-e-ba	3	k'itl-d-eb-a	3
13	i-c'ek-eb-a	2	ɣer-d-eb-a	2
14	i-map'-eb-a	2	rov-d-eb-a	2
15	i-xeš-eb-a	2	rux-d-eb-a	2
16	i-k'rat'-eb-a	3	c'agr-d-eb-a	3
17	i-xarp'-eb-a	3	ʒvar-d-eb-a	3
18	i-št'iv-eb-a	3	rtiv-d-eb-a	3
19	i-t'vix-eb-a	3	mtip-d-eb-a	3
20	i-p'lit'-eb-a	3	mq'ar-d-eb-a	3
21	i-k'rit'-eb-a	3	mʒik-d-eb-a	3
22	i-c'vek'-eb-a	3	rc'ar-d-eb-a	3
23	i-krin-eb-a	3	ngrel-d-eb-a	3
24	i-ɣrevn-eb-a	4	pkvin-d-eb-a	4

Length-matched for word

Pair	Prefixed	Syllables	Suffixed	Syllables
1	i-mp-eb-a	3	mcik'-d-eb-a	3
2	i-t'q'r-eb-a	3	moz-d-eb-a	3
3	i-r-eb-a	3	crel-d-eb-a	3
4	i-zm-eb-a	3	clad-d-eb-a	3
5	i-kl-eb-a	3	q'vac-d-eb-a	3
6	i-č'l-eb-a	3	rbad-d-eb-a	3
7	i-c'k'-eb-a	3	brag-d-eb-a	3
8	i-lep-eb-a	3	crid-d-eb-a	3
9	i-mʒl-eb-a	3	sups-d-eb-a	3
10	i-sxr-eb-a	3	t'q'lep'x-d-eb-a	3
11	i-svar-eb-a	4	pxial-d-eb-a	4

Pair	Prefixed	Syllables	Suffixed	Syllables
12	i-xarc-eb-a	4	zriel-d-eb-a	4
13	i-sx-eb-a	3	očr-d-eb-a	3
14	i-ck’-eb-a	3	uxn-d-eb-a	3
15	i-c’q’v-eb-a	3	t’ruv-d-eb-a	3
16	i-p’-ev-a	3	sxmev-d-eb-a	3
17	i-šx-eb-a	3	betr-d-eb-a	3
18	i-mt’r-ev-a	3	napr-d-eb-a	3
19	i-rkm-eb-a	3	mrčan-d-eb-a	3
20	i-rkmen-eb-a	4	tbian-d-eb-a	4
21	i-t’k’and-eb-a	4	paprov-d-eb-a	4
22	i-t’on-eb-a	4	reptav-d-eb-a	4
23	i-t’ečk’-eb-a	4	jian-d-eb-a	4
24	i-kin-eb-a	4	ukad-d-eb-a	4

Appendix G.

The method for evaluating a language’s prefixing vs. suffixing in inflexional morphology is described in Dryer (2013), at https://wals.info/chapter/26. The numbers under the headings “Prefixes” and “Suffixes” are the points assigned on this method, not the number of affixes. Affixes 1–3 are assigned double points, which may all be assigned to one column as in #1 in Georgian (where all case affixes are suffixes), or may be split, as in #2 (where there are both prefixes and suffixes).

	Prefixes	Suffixes
1. case affixes on nouns		2
2. pronominal subject affixes on verbs	1	1
3. tense-aspect affixes on verbs		2
4. plural affixes on nouns		1
5. pronominal possessive affixes on nouns	N/A	
6. definite or indefinite affixes on nouns	N/A	
7. pronominal object affixes on verbs	1	
8. negative affixes on verbs	N/A	
9. interrogative affixes on verbs	N/A	
10. adverbial subordinator affixes on verbs	N/A	

This gives Georgian an inflectional suffixing index of six and an inflectional prefixing index of two. Six is 75% of the total score for affixes.

PART III

Competition, inheritance, and defaults

CHAPTER 10

Feature-based competition

A thousand years of Slavonic possessives

Greville G. Corbett
Surrey Morphology Group, University of Surrey

Competition takes many forms. A newly identified type of competition involves the featural specification of one of the competitors as a key factor. In the particular instance treated here, whether a given item has a competitor depends on its number (and sometimes its person). We focus on the use of the genitive case versus adjective-like forms in possessive expressions (broadly understood). The data come primarily from the Slavonic languages, where a surprising original system of possessive pronouns competing with personal pronouns has played out rather differently through the family. We find a variety of outcomes, from conservative to highly innovative, with some instances of competitors settling into different niches.

Keywords: possession, genitive, possessive pronouns, possessive adjectives, derivation, Slavonic languages, Russian, Old Church Slavonic, Slovenian, Upper and Lower Sorbian

1. Introduction

'Competition' is widely used in linguistics, at different levels of analysis from semantics to phonetics. In some instances, however, it is uncertain whether competition is being used as a technical term or as 'ordinary language', and this brings opportunities for misunderstandings. In ordinary use, competition tends to suggest an ongoing process, observable over time. This is not the case for all its uses in linguistics: for instance, in classical Optimality Theory, different inputs compete, to give a single output, in principle instantaneously. Aronoff's use of competition (2016, 2019b), linking it to ecology, suggests a process, and the examples I shall discuss also involve ongoing competition. Yet there may be possible combinations of factors (making up a niche) which rule out one of the competitors. We need to be clear about which elements the competitors are: lexemes, rules or candidate

https://doi.org/10.1075/cilt.353.10cor

forms (see Williams 2007: 361–364, and Stump 2019: 260–262, among others). And unlike in human competition, there is no mutual awareness (Aronoff 2019b: 42).

Homing in on competition in morphology, there is a fine account of the use of the term and its development in Gardani, Rainer & Luschützky (2019). But there is an area as yet inadequately occupied, namely the competition between lexemes which is specified in part by the featural specification of one of the competitors. In other words, we find lexemes (typically lexemes which are derivationally related), which compete only in certain featurally-defined contexts. Equivalently, the competition can be seen in terms of a partial overlap of paradigms, which is a non-canonical relation between lexemes. This type of competition is our niche.

2. Feature-based competition

There are many instances of lexical competition, that is, competition between 'whole lexemes'. Each competitor comes with its phonology, inflection class, selectional requirements and so on. We will consider rather competition which is regulated by the feature specification (paradigm cells) of one of the competitors. Consider the two boldened lexemes, in these possessive constructions:

(1) Russian

portret	***mam-y***
portrait(M)[SG.NOM]	mother(F)-SG.GEN

"mother's portrait"

(2)

mam-in	*portret*
mother-POSS.ADJ[SG.NOM.M]	portrait(M)[SG.NOM]

"mother's portrait"

In (1) we have a noun in the genitive, while in (2) there is a possessive adjective *mamin* "mother's", derived from the noun *mama* "mother". This possessive adjective agrees with the head noun *portret* "portrait". In this instance, both constructions are possible, and both are attested in the Russian National Corpus.[1] There are various factors involved in the competition between (1) and (2), and some factors are sufficiently strong to rule out one of the competitors (as discussed in § 3). Our interest is in the fact that (2) is available only in competition with the genitive case of animate nouns like *mama* "mother", and not in competition with the other case values of its source noun. This restriction arguably makes sense, in that the competitor competes for a specific role, as will be discussed in (§ 3). We shall then move to a contrasting

1. Freely available at https://ruscorpora.ru/new/

example: the competition has to be specified according to featural specifications which lack any obvious motivation (§ 4).

It is notable that our examples come primarily from possessive constructions (taken broadly), where one of the competitors is the genitive case. This is understandable since it is the "Case whose basic role is to mark nouns or noun phrases which are dependents of another noun" (Matthews 1997: 144). In the languages to be analysed, however, the genitive has other functions (for the grammatical roles of subject and object, and as the object of a preposition), showing that it forms part of the main case system. For recent discussion of the definition of genitive see Lander (2009).[2]

A good deal has been written, particularly in recent years, on possession and genitives; sometimes there is a European slant, as in Koptjevskaja-Tamm (2003), Stolz et al. (2008), Ackermann, Simon & Zimmer (2018) and Johanson, Mazzitelli & Nevskaya (2019), and there are also the following broader collections: Aikhenvald & Dixon (2013), Börjars, Denison & Scott (2013) and Carlier & Verstraete (2013). Most importantly for our purposes, Nikolaeva & Spencer (2013) and Spencer (2013: 348–356) provide insightful analysis of possession and modification. They distinguish canonical possession and canonical modification, for which they provide a set of criteria. Between these canonical extremes, there are two intermediate, non-canonical constructions: 'modification by noun' and 'alienable possession'. The distinctions are carefully drawn, which then allows discussion of the ways in which particular languages may combine the different types. All this provides a firm underpinning to my topic, which tackles something that Nikolaeva and Spencer leave aside, namely the competition between the different possibilities and specifically the featural contribution to that competition.

3. Motivated feature-based competition

We return to the competition in Russian between *mamy*, as in (1), and *mamin,* as in (2), both of which can be translated as "mother's". A natural analysis is that we are not dealing with forms of the same lexeme (as when, say, different case values may be possible in a given instance), rather these are different lexemes;[3] *mamy* is

2. Note too that the genitive is not the only adnominal case; for other adnominal case values, further afield, notably the proprietive, see Dench & Evans (1988) and Ponsonnet (in preparation). Within Slavonic, some languages use the dative as another means for marking possessors.

3. In the clearest cases this is certainly true. In some instances, the regularity of the competition has led linguists to treat the competing items as belonging to the same paradigm (see footnote 13). For related denominal adjectives in Russian see Spencer (2013: 351–354).

the genitive singular of *mama* "mother", while *mamin* is the singular nominative masculine form of the adjective *mamin* "mother's", an adjective which is derived from *mama*. The paradigms of the two lexemes can be thought of as partially overlapping, in that the possessive adjective may compete with the genitive singular of the noun, but not with other cells of its paradigm. This is diagrammed in Figure 1.

		mama "mother, mum"
SINGULAR	NOM	mam-a
	ACC	mam-u
	GEN	mam-y
	DAT	mam-e
	INS	mam-oj
	LOC	mam-e
PLURAL	NOM	mam-y
	ACC	mam
	GEN	mam
	DAT	mam-am
	INS	mam-ami
	LOC	mam-ax

mamin "mother's"		
M	N	F
mamin	mamin-o	mamin-a
=NOM/GEN	mamin-o	mamin-u
mamin-ogo		mamin-oj
mamin-omu		mamin-oj
mamin-ym		mamin-oj
mamin-om		mamin-oj
mamin-y		
=NOM/GEN		
mamin-yx		
mamin-ym		
mamin-ymi		
mamin-yx		

Figure 1. Competition between noun and possessive adjective in Russian

Before we tackle the key type of competition in Figure 1, we note that the material included in it can serve to indicate just how pervasive competition is. There are various types of competition involved, and we should put aside the following types that are *not* our main concern:

1. There is lexical competition between *mama* "mother, mum" and *mat'* "mother", with *mama* being used both as a proper name and as an ordinary common noun.
2. There is derivational competition between the two main formatives for deriving the possessive adjective, namely *-in-* and *-ov-*.[4] In Russian the competition is resolved according to the inflection class of the base; nouns which inflect like *mama* "mother, mum", whether they denote females or males, take *-in-*, while those which inflect like *otec* "father" take *-ov-*.[5] The latter are less readily

4. For discussion of the development of these and other competing possessive suffixes see Bratishenko (2005) and Matasović (2010: 6–9).

5. Certain other Slavonic languages resolve the competition according to the gender of the base (Corbett 2010: 149–151).

derived, and overall there are restrictions on forming the possessive adjective; it had been considered under serious threat in the modern language, but Šmeleva (2008) shows that earlier predictions of its demise had been unduly pessimistic.
3. There is inflectional competition within the possessive adjective *mamin* "mother's, mum's". We find ongoing competition between two sets of inflections: those of the nominative, and the accusative feminine and neuter, are older, shorter, and follow the inflections of nouns; the rest of the paradigm has newer, longer, adjective-like inflections. The progression has been: instrumental ≥ locative ≥ dative, genitive ≥ accusative ≥ nominative, according to Timberlake (2004: 127); Spencer (2002) discusses a slightly older set of forms.
4. There is inflectional competition between nominative and genitive forms for two cells in the paradigm (indicated =NOM/GEN). The accusative is syncretic with the nominative for inanimates, and with the genitive for animates.

The type of competition that *does* concern us is that the possessive adjective competes with the noun only in respect of the genitive case of the latter, and in the singular; a noun in any other case-number combination has no competitor.[6]

Three points deserve clarification here. First, the possessive adjective has a full paradigm, meaning that it can agree fully with its head; however, the competition with the underived noun lexeme is solely in those environments where the noun would stand in the genitive singular. We can thus distinguish a 'primary' lexeme (here the noun), which determines the range of the competition, and a 'secondary' lexeme (the possessive adjective), which can compete in the specified area only. Second, the possessive adjective competes only with adnominal uses of the genitive. And third, there are numerous factors bearing on the choice between the competitors (discussed in Corbett 1987: 324–326, Kopčevskaja-Tamm & Šmelev 1994 and Janda, Nesset & Say 2019: 193–203); for cross-linguistic discussion, including Slavonic data, see O'Connor, Maling & Skarabela (2013).[7] A key factor in the Russian competition is syntax: the possessive adjective is very restricted in the syntactic dependents it can have. Similar restrictions hold in almost all the Slavonic languages.[8] These are the possibilities for Russian:

---

6. There is a considerable literature on the diachronic competition of genitive singular versus possessive adjective; see particularly Richards (1976), Bratishenko (1998) and Eckhoff (2009, 2011), and further references there.

7. The English possessive alternants have attracted an enormous amount of work; for a useful survey see Rosenbach (2014) and for an interesting account of the oblique genitive (as in *a friend of John's*) see Payne & Berlage (2014).

8. The interesting exceptions, particularly in Upper Sorbian, can be found in Corbett (1987); for recent discussion see Nikolaeva & Spencer (2020: 132–139, 331–333).

(3) Russian
portret *tvo-ej* ***mam-y***
portrait(M)[SG.NOM] your-F.SG.GEN mother(F)-SG.GEN
"your mother's portrait"

(4) **tvo-ej* ***mam-in*** *portret*
your-F.SG.GEN mother-POSS.ADJ[SG.NOM.M] portrait(M)[SG.NOM]
"your mother's portrait"

Whatever other conditions are met or not met, the presence of the type of syntactic dependent in (4) makes the possessive adjective impossible, and the genitive (as in (3)) is the only possibility.[9]

Of all the factors, we are concerned with the one involving the inflectional paradigm, that is, the fact that competition is restricted to the genitive singular of the primary lexeme. Now there is a reason why it is the genitive singular which is in competition; in other words, this is motivated competition. First, the genitive is the case most often used for attributive modification; hence it appears in situations where the noun approaches most closely to adjectival function (though recall that the genitive is also a core case.) Second, favouring the singular number is bound up with the fact that possessive adjectives are typically used when their referent is specific, which is most likely with the singular number.

4. Unmotivated feature-based competition

We now turn to a somewhat different type of competition, also feature-based, but where the basis for the featural competition is not evidently motivated. This comes from Old Church Slavonic, the earliest source of evidence for the Slavonic languages. Proto-Slavonic is a reconstruction, while Old Church Slavonic shows traits of one of the branches of Slavonic, namely South Slavonic. Hence it is not the direct ancestor of Russian (which belongs to East Slavonic) but it is the attested variety of archaic Slavonic. The Old Church Slavonic corpus consists of translations of Greek liturgical texts. The translations were made in the ninth century and the earliest extant copies of those translations are reckoned to date from the late tenth century (Huntley 1993: 125). Consider one of the personal pronouns and its corresponding possessive pronoun as laid out in Figure 2.

9. Interestingly, however, comitative phrases, like *Aleša s Mašej* "Alesha with Masha" / "Alesha and Masha", do allow equivalents with a possessive adjective: *Alešina s Mašej stat'ja* literally "Alesha's with Masha article" (Kopčevskaja-Tamm & Šmelev 1994); similar parallels are found with the possessive pronoun.

		1PL "we"
PLURAL	NOM	my
	ACC	nasъ
	GEN	nasъ
	DAT	namъ
	INS	nami
	LOC	nasъ

		1PL POSSESSIVE "our"		
		M	N	F
SINGULAR	NOM	našь	naše	naša
	ACC	=NOM/GEN	naše	našǫ
	GEN	našego		našeję
	DAT	našemu		našeji
	INS	našimь		našejǫ
	LOC	našemь		našeji
DUAL	NOM	naša	naši	
	ACC	naša	naši	
	GEN	našeju		
	DAT	našima		
	INS	našima		
	LOC	našeju		
PLURAL	NOM	naši	naša	našę
	ACC	našę	naša	našę
	GEN	našixъ		
	DAT	našimъ		
	INS	našimi		
	LOC	našixъ		

Figure 2. Competition between personal pronoun (first plural) and possessive pronoun in Old Church Slavonic

Notes:
i. the transcription of Old Church Slavonic follows Huntley (1993)
ii. there are two reduced vowels ь (front) and ъ (back)
iii. the forms for Figures 2 and 3 are from Huntley (1993: 144)

As we saw earlier for the Russian possessive adjective (Figure 1), the competition, as might be expected, is with the genitive.[10] But now compare (Figure 3) the forms for the first person dual ("we two").

		1 DU "we two"
DUAL	NOM	vě
	ACC	na
	GEN	naju
	DAT	nama
	INS	nama
	LOC	naju

Figure 3. The personal pronoun (first dual) and possessive pronoun in Old Church Slavonic

10. Again, there are related issues that we leave to one side, in particular, the existence of clitic pronouns, reflexive pronouns, and the possessive dative.

For this feature combination, there is no distinct possessive pronoun. The genitive of the personal pronoun is used, with no further inflection, for all gender/number/case possibilities. Here is an example:

Old Church Slavonic (Codex Marianus, early eleventh century)[11]

(5)	*syn-ъ*	***naju***
	son-SG.NOM	1DU.GEN
	"our son"	(of the two parents)

Example (5) is from John 9, verse 20. The previous verse has this comparable example (Flier 1974: 72):

(6)	*syn-ъ*	***vaju***
	son-SG.NOM	2DU.GEN
	"your son"	(of the two parents)

There is no distinct possessive pronoun form, but the genitive serves that function. It stands after the noun, like a normal genitive phrase. Usually possessive adjectives (like *isusovъ* "Jesus's") are also postposed.[12] Most significantly, inflecting possessive pronouns (like *našь* "our" in Figure 2) are postposed. Vaillant (1977: 266) found 20 postposed instances to just one preposed; this follows the Greek original in terms of order (the Greek forms are the genitive pronoun). This postposed order is unlike ordinary adjectives, which can be in either position, with preposed being more normal (Vaillant 1977: 266). Much more detail on this, with statistics on the use in different Old Church Slavonic sources and comparison with the Greek original can be found in Večerka (1989: 77–85).

We therefore find a different and particularly interesting type of competition, for the function of possessive pronoun: there may be distinct forms (as in Figure 2), or the genitive of the ordinary personal pronoun may be used (as in Examples (5) and (6)). Given the two possibilities, we should examine the full range. To accommodate the complete picture, Figure 4 has number in the columns and person and case in the rows. Two case values are given for the personal pronoun, the remaining four are indicated "…"; similarly, for the inflecting possessive pronouns, the forms which are comparable to *našь* "our" in Figure 2 are indicated "…".

Figure 4 should be read as follows: two representative forms of the personal pronouns are given on the left. In the possessive pronoun section on the right, the

11. http://titus.uni-frankfurt.de/texte/etcs/slav/aksl/marianus/maria.htm

12. Eckhoff (2018: 40–42) demonstrates how closely the ordering of the possessive adjective (overwhelmingly postnominal) follows the Greek original.

		personal pronoun			possessive pronoun		
		SG	DU	PL	SG	DU	PL
	NOM	azъ	vě	my	ADJECTIVAL		
1	GEN	mene	naju	nasъ	moi …	naju	našь …
	…	…	…	…			
	NOM	ty	va/vy	vy	ADJECTIVAL		ADJECTIVAL
2	GEN	tebe	vaju	vasъ	tvoi …	vaju	vašь …
	…	…	…	…			
	NOM	(tъ)	(ta)	(ti)			
3M	GEN	jego	jeju	jixъ	jego	jeju	jixь
	…	…	…	…			
	NOM	(ta)	(tě)	(ty)			
3F	GEN	jeję	jeju	jixъ	jeję	jeju	jixь
	…	…	…	…			
	NOM	(to)	(tě)	(ta)			
3N	GEN	jego	jeju	jixъ	jego	jeju	jixь
	…	…	…	…			

Figure 4. Personal pronouns and possessive pronouns in Old Church Slavonic

Notes:

i. Full forms are given (there are also some clitic forms).

ii. The basic data can be found in Vaillant (1964: 147–150), Schmalstieg (1983: 63–65) and Huntley (1993: 143–144).*

iii. The expected nominative forms of the third person pronoun are no longer in use in Old Church Slavonic (though preserved in the relative pronoun); their place was taken by a demonstrative, *tъ* "that", as given in parentheses, or *onъ* "that".

* For data on the forms, see Olander (2015: 77–80), and for more examples of the use of dual genitive pronouns see Žolobov (1998).

cells labelled "ADJECTIVAL" indicate a full paradigm, comparable to that given in Figure 2, while the shaded cells indicate forms which do not inflect further: they are identical to the shaded genitive cells on the left.

The remarkable distribution made clear in Figure 4 is that specified by Vaillant (1964: 149); there are distinct, adjectival possessive pronouns complemented by the genitive of the personal pronoun. This competition is based on a surprising feature specification: first and second persons singular and plural (but not dual). This distribution is not motivated in a straightforward way. Markedness is confounded in that we have the most marked number value (dual) combined with the least marked person value (third).

The interest of the Old Church Slavonic situation can be seen if we set up some canonical (idealized) baselines (Round & Corbett 2020). In the canonical world, lexical choice is free; in particular, it does not depend on competing lexical items; so if we take this perspective, the existence of competition is already non-canonical. And while we may be used to finding dedicated possessive pronouns, we also find the canonical situation – that with no competition. Thus, as one instance, the Indo-Aryan language Palula of northern Pakistan has no unique possessive pronoun, rather the possessive function is always expressed by the genitive of the pronoun, personal or demonstrative (Liljegren 2016: 143). Second, if a given language does have possessive pronouns, the canonical situation would then be for it to have a full set of distinct forms (then the competition would be between full lexemes). This is where Old Church Slavonic is so interesting, in having some unique possessive pronouns and some slots occupied by genitives. A third criterion to apply comes from the typological literature: given the factors that are established as part of the more general competition with possessives (as in O'Connor, Maling & Skarabela 2013), and the prevalence of possessors being expressed as pronouns rather than nouns (demonstrated for Czech by Křivan 2019), we might expect that a language would have a complete possessive pronoun set before other possessives would be available. Old Church Slavonic does not meet this expectation. It has possessive adjectives derived from nouns (comparable to those discussed for Russian in § 3); indeed, according to Vaillant (1977: 52), for nouns denoting persons (unless they are modified) it is almost obligatory to use the possessive adjective rather than the genitive.[13] Thus in Old Church Slavonic the possessive adjective, with forms distinct from those of the corresponding noun, is in advance of the possessive pronoun which lacks them for dual and third person.

5. The nature of the competition

We need to be clear about the type of competition in question, to establish that it is indeed a special type. We are analysing competition between lexemes, which is determined by the featural specification of the primary lexeme. Given that it is restricted to a given featural specification (particular cells), we need to establish that we are not dealing with overabundance, where "two (or more) inflectional forms are available to realize the same cell in the inflectional paradigm of a lexeme (i.e., to express the meaning arising from the combination of the lexical meaning of the

13. Given this high productivity, Trubetzkoy (1937: 16) considered that the possessive adjective should be treated as a part of the paradigm of the noun – just as participles are considered to be part of a verb paradigm. Isačenko (1954: 288–289), working on Slovak, followed Trubetzkoy. While productivity is an important criterion, we should also note that the possessive adjective involves a change of word-class membership. Further discussion and sources are given in Corbett (1987: 305–307). A fine account of the factors determining the distribution of possessive adjective versus adnominal genitive in Old Church Slavonic is provided by Huntley (1984).

lexeme and the morphosyntactic and morphosemantic feature values that define the cell)" (Thornton 2019a: 1, see also 2019b). Could we argue that *mama* "mother" in (2) shows overabundance in the genitive singular, having the genitive *mamy* on the one hand, and the whole paradigm of *mamin* on the other, as cell-mates? Not according to Thornton's definition, since there is more to the competition than "the combination of the lexical meaning of the lexeme and the morphosyntactic and morphosemantic feature values that define the cell" (2019a: 1). The key point remains that *mama* "mother" and *mamin* "mother's" (§ 3), and *ja* "I" and *moj* "my" (§ 4), are pairs of different lexemes belonging to different parts of speech.[14]

It may seem evident that we are dealing with two lexemes (but note the discussion in footnote 13). It is also worth saying that there are some phenomena which share aspects of the competition we are discussing but where the balance of evidence is for a single lexeme. First consider suppletion (more directly relevant with the pronouns than with the noun and possessive adjective discussed in § 3, since there the forms are phonologically closer). Now canonical suppletion (as in *go* ~ *went*), where we find maximal regularity in semantics paired with maximal irregularity in form, is a phenomenon involving a single lexeme; it is therefore not directly relevant. But there are more complex (and less canonical) instances of suppletion, which need closer attention. For instance, we might imagine that a lexeme cannot have a cell which itself involves another paradigm, but we find this in some suppletive cells of the Archi personal pronoun (Corbett 2015: 160–161), and arguably in some participial forms in more familiar languages. Thus this apparent peculiarity of our examples can also be found within single lexemes. Even different parts of speech can combine over time to create a single suppletive item, as in the famous case of Italian *uscire* "go out", which arise from a combination of the stems of Old Italian *escire* "go out" and *uscio* "doorway" (Maiden 1995). Since a paradigm-within-a-paradigm can be found in single lexemes, and since there is evidence for items of different parts of speech combining to give a suppletive item, could it be that the two sorts of Slavonic competition will lead, via a stage of overlapping suppletion (Juge 2019: 392–394), to new suppletive items? This is unlikely. As we shall see in § 6, the competing items (personal pronoun and possessive pronoun) are moving apart in various ways, settling into distinct niches and differentiating themselves, rather than coming together.

14. For 'rival forms' taken more generally see Baayen et al. (2013).

6. The competition in diachrony

We know in outline how the competition between the genitive of the personal pronoun and the possessive pronoun plays out over time, but the detail is yet to be understood. In brief, the trend within Slavonic has been towards a possessive pronoun, distinct from the personal pronoun. That is, a possessive pronoun with a full set of distinct forms. The possessive pronoun tends towards drawing the same distinctions of gender, number and case as other adjectives, even if the inflections do not match those of ordinary adjectives.[15] The possessive pronoun typically stands before the noun (like ordinary adjectives).

6.1 A cautionary note

Allowing for the inherent problems of the Old Church Slavonic situation (recall from § 4 that it is not the parent language, Proto-Slavonic, and that the texts are translations), we have a reasonable picture of the starting situation, as summarized in Figure 4. The modern Slavonic languages, though interestingly different as we shall see, might suggest a consistent direction of change, as sketched above. However, as yet we have insufficient data to justify that position; we have to admit the possibility of changes and reversals in the intervening period.

Within a wide-ranging study of the expression of possession in the languages of Europe, Stolz et al. (2008: 371–374) investigate the Adjectival versus Genitive split within pronominal possessives. The criterion is simply agreement or lack of it. There are further differentiating tests that might be applied, but the picture that emerges using their criterion is interesting and relevant. First, there are languages, including Indo-European languages, which have only adjectival forms, and some which only use the genitive of the pronoun. This suggests that changes between those two extremes have gone in two different directions. Now consider languages where the two are in competition (Figure 5).

The picture is very suggestive. Note that all the languages involved are Indo-European languages, sharing their origin. There could be further interest concerning the dual.[16] And as we shall see from the Russian data (discussed in § 6.2.2), there can be evidence for an uninflecting possessive pronoun which is distinct from its source personal pronoun in syntactic behaviour. The conclusion here is that we should beware of assuming a straightforward progression in a single direction.

15. In terms of Stump (2016: 104), the content paradigms match, the form paradigms also match, and differences are restricted to the realized paradigms.

16. For its place in the history of Icelandic see Guðmundsson (1972: 67–68).

	Possessor						
Language	**Reflexive**	**1SG**	**2SG**	**1PL**	**2PL**	**3SG**	**3PL**
Faroese	A	A	A	G	G	G	G
Icelandic	A	A	A	G	G	G	G
Latvian	A	A	A	G	G	G	G
Danish	A	A	A	A	G	G	G
Norwegian	A	A	A	A	G	G	G
Swedish	A	A	A	A	A	G	G
Polish	A	A	A	A	A	G	G
Russian	A	A	A	A	A	G	G
Rumanian	A	A	A	A	A	A/G	A/G
Italian	–	A	A	A	A	A	G

Figure 5. Adjectival vs. Genitive split of possessive paradigms (Stolz et al. 2008: 372)

6.2 The key developments

We know the outline of the development, but there is much to be learned about how the different factors play out. There are four key developments to bear in mind. Each of the four is only partly understood, since the textual record and the research carried out are both patchy. As the different languages have been researched to different degrees, we cannot yet know how the developments interact in each case. I lay out the essentials, since one of the effects of each development is to differentiate the possessive pronoun from the genitive of the personal pronoun, and so to delineate their niches.

6.2.1 *Word order*

Recall that in the earliest texts, possessive pronouns stand after the noun (corresponding in large measure to the Greek original from which Old Church Slavonic texts were translated, Večerka 1989: 77–85); also postposed are genitive personal pronouns, like other genitive dependents, and possessive adjectives. Ordinary adjectives were mainly preposed. Throughout the Slavonic family there has been a trend for the inflecting possessive pronoun, the genitive personal pronoun when functioning as a possessive, and the possessive adjective (but not the ordinary genitive) to move to preposed position, aligning with and typically preceding ordinary adjectives. However, the change was not straightforward and little has been clearly established. For instance, for Russian, Hanne Eckhoff kindly extracted data from the Tromsø Old Russian and OCS Treebank (TOROT) corpus (Eckhoff & Berdičevskis 2015),[17] and the picture shows considerable variation, from the birchbark texts of

17. Personal communications of 24 October 2019.

the 11th to 14th centuries right up to texts of the late 18th century. Overall there was still a majority of examples being postposed, for both the agreeing possessive pronouns and the non-agreeing genitive pronouns; see also Minlos (2012). Nevertheless, in modern Russian, and in other modern Slavonic languages, where the possessive pronoun (agreeing or not) stands before the noun, this is a clear factor differentiating the possessive pronoun from the genitive of the personal pronoun (when not functioning as a possessive), which follows the noun.

6.2.2 *The* N-*factor*

The *N*-factor (taken from the title of Hill 1977) is a useful label for a series of linked phenomena, difficult to analyse but clearly a part of our story. The *N*- is a prothetic *n*- gained by Slavonic personal pronouns, in certain forms under different conditions. The start of this complex process, ongoing in some languages, is before the earliest texts. Originally it involved just three prepositions: *vŭn "in", *sŭn "with", and *kŭn "to(ward)". When these occurred preceding a pronoun with an initial front glide, the following resyllabification occurred (Billings 1996: 62; see also Hill 1977: 305–312 and references there):

(7) *kŭn jemu* → *kŭ. ñe. mu*
to 3SG.DAT
"to(ward) him"

From these modest beginnings, the *N*- has spread dramatically in some languages, as more and more prepositions have triggered it; Hill (1977) documents this in luxuriant detail for Russian. He shows how the environment for *N*- has narrowed: while previously the preposition and pronoun had only to be adjacent for the *N*- to appear, it is now restricted to instances where the pronoun is governed directly by the preposition (Hill 1977: 227–242);[18] this is an important development, as we shall see.[19] Thus in some languages the *N*- depends on the presence of a preposition. As a result, the pronoun has two forms, one for normal use, such as Russian *im* (instrumental) "by him", contrasting with *s nim* "with him", with the *N*-. But there are other languages where the full form of the pronoun has generalized the *N*-. As one instance, in Serbo-Croat the third singular masculine instrumental is *njim*, whether or not it is governed by a preposition. Both types of development are relevant to our story.

---

18. In Russian the environment is expanding in another way in that there are now instances of the *N*- appearing when preposition and pronoun are not adjacent (Daniel 2015, Philippova 2018: 18, 28, 34–35).

19. In some languages this use of the *N*- has been lost (notably in Belarusian); a reversal of the type alluded to in § 6.1.

Consider first the *N*- which appears when the pronoun is directly governed by a preposition. This becomes important when the genitive form of the pronoun, serving as a possessive, stands before the noun. Consider these phrases:

(8) Russian (Hill 1977: 28)

a. *dlja n-ee*
for *N*-3SG.F.GEN
"for her"

b. *dlja ee roditel-ej*
for 3SG.F.GEN parent-PL.GEN
"for her parents"

In (8a), the preposition governs the pronoun directly, and so we find the form with *N*-, while in (8b), the preposition does not directly govern the pronoun. The main situation in which there will be a contrast is precisely that in (8), where there is a contrast between the ordinary pronoun and the genitive in possessive pronoun use. Thus the *N*-factor, in the relevant languages, plays a role in differentiating the uninflecting possessive pronoun from the descendant of its source, the genitive of the personal pronoun. The other development, in which the *N*- becomes a permanent segment of the full pronoun, may in turn provide part of the stem from which an inflecting possessive pronoun is formed (we shall see some examples in § 6.3 and § 6.5).[20]

6.2.3 *Derivational reforming of the possessive pronoun*

In some instances, we find derivational means providing new stems for the possessive pronoun; these may build on the stem with the initial *N*- discussed in § 6.2.2. We find this in the South Slavonic languages, Bulgarian, Macedonian and

20. Ferrell (1958: 97) makes the interesting claim that those languages which have no special *N*- forms for the personal pronoun following prepositions are those which have developed distinct forms for the possessive pronoun. He sees the motivation in the problem of 'the intercalated genitive after a preposition'. In phrases of the type: preposition-pronoun(GENITIVE)-noun, there could be ambiguity as to whether the preposition governed the pronoun directly. There were two solutions: (i) create a distinct possessive pronoun for the missing cells of the paradigm, or (ii) treat the *N*- as marking direct dependence of the pronoun on the preposition (since then the lack of an *N*- in the intercalated structure resolves the ambiguity). There are three main issues with Ferrell's suggestion. First, the number of instances (languages) is small; in particular the first solution is found largely in the South Slavonic branch, and so we need to establish how many independent developments are represented. Second, it needs to be demonstrated that the chronology of the preposing of the genitive pronoun and the spread of *N*- as marking direct dependence of the pronoun is consistent with the causal link Ferrell suggests. And third, the patchy development of agreeing possessive pronouns sketched in §§ 6.3–6.4 does not fit well with his suggestion.

Serbo-Croat (also in Slovenian, for which see § 6.5). Let us take Serbo-Croat[21] first as the most straightforward. Alongside the expected *moj* "my", *tvoj* "your (singular)", *naš* "our" and *vaš* "your (plural)" we find *njegov* "his, its", *njen* / *njezin* "her" and *njihov* "their" (Browne 1993: 323–325). These third person forms all have a pronominal form including the initial *N*-, and a derivational affix; they inflect for gender, number and case. This is a clear indicator that the possessive pronoun has fully occupied the earlier niche, since it has achieved a full set of forms, for each person and number, distinct from the personal pronoun.

6.2.4 *Loss of the dual*

The dual has been lost in most of the Slavonic languages, a fascinating story in itself, and one which has left dramatic complexities in the system of numeral phrases (see, among many others, Belić 1932; Suprun 1969; Mel'čuk 1985; Žolobov & Krys'ko 2001; Corbett 2012: 209–210, and references there). For the issue of the possessive pronoun, however, the effect of the loss of the dual is to remove three problematic cells, where the genitive personal pronoun served as the possessive pronoun; by removing these cells, the loss of the dual has a substantial effect in differentiating the possessive pronoun. We return to the interesting situation of languages which preserve the dual, Upper and Lower Sorbian (in § 6.4) and Slovenian (in § 6.5).

6.3 Outcomes in languages where the dual has been lost

Given the starting situation of Common Slavonic, and the different developments sketched in § 6.2, there are various possible outcomes for the relation between personal pronoun and possessive pronoun. The first is that the possessive pronoun is distinguished functionally from the genitive of the personal pronouns, in terms of word order for instance, but that nothing significant happens in terms of the pattern of forms. This is what we find in the West Slavonic language Slovak (Short 1993b: 546–547). That is, it has possessive pronouns for the first and second persons, singular and plural, which agree, and uninflecting forms for the third person (all genders, singular and plural).[22]

Now consider Czech, a West Slavonic language, most closely related to Slovak. The forms are as in Figure 6.

21. Following the 2017 'Deklaracija o zajedničkom jeziku' (http://jezicinacionalizmi.com/deklaracija/), I treat Serbo-Croat as one polycentric language, with four standards, Bosnian, Croatian, Montenegrin, and Serbian. See Corbett & Browne (2018) for a linguistic outline and Bugarski (2012, 2019 and references there) for the difficult complexities of the sociolinguistic background and language status.

22. I give the situation in the standard languages; as Motoki Nomachi points out (personal communication, 1 November 2019) there are inflected forms in dialects of Slovak and Polish.

		personal pronoun		possessive pronoun	
		SG	PL	SG	PL
1	NOM	já	my	ADJECTIVAL	ADJECTIVAL
	GEN	mne	nás	můj …	náš…
	…	…	…		
2	NOM	ty	vy	ADJECTIVAL	ADJECTIVAL
	GEN	tebe	vás	tvůj …	váš…
	…	…	…		
3M	NOM	on	oni / ony		
	GEN	jeho	jich	jeho	jejich
	…	…	…		
3F	NOM	ona	ony	ADJECTIVAL	
	GEN	jí	jich	její …	jejich
	…	…	…		
3N	NOM	ono	ona		
	GEN	jeho	jich	jeho	jejich
	…	…	…		

Figure 6. Personal pronouns and possessive pronouns in Czech (Short 1993a: 470–473)
Note: in the third person masculine plural, *oni* is for animates, and *ony* for inanimates

Czech shows interesting developments. The third person possessive forms have come together in having an initial *je-*. For the third plural, this gives a form which is distinct from the genitive of the personal pronoun, while still not inflecting. The new third singular feminine *její* had a form with the appearance of an adjective, and it has indeed become an adjectival, agreeing form. This leaves a surprising distribution, in which all the forms of the possessive adjective, apart from *jeho*, are distinct, whether inflecting or not, from the genitive form of the personal pronoun.

The changes in Polish, also West Slavonic, are particularly noteworthy. While the basic position is comparable to that in Slovak, Polish has innovated polite pronouns: *pan* (to a man), *pani* (to a woman), *panowie* (men), *panie* (women) and *państwo* (mixed group). Of these, just *pan* has an inflecting possessive *pański* (Bartnicka et al. 2004: 278–280, 287). Three of these polite pronouns use their genitive as a possessive pronoun: *pana* (an alternative to *pański*), *pani* (to a woman) and *państwa* (mixed group). Most interestingly, these are normally preposed, as in the following example:

(9) Polish (Bartnicka et al. 2004: 287)
Czy to jest ***pana*** *samochód*?
Q DEM be.3SG Mr.POSS car[SG.NOM]
"Is that your car?"

This shows that the possessive pronouns have a prenominal slot, and these new forms (equivalent to the genitive, and not inflecting further) are occupying it; this distinguishes them from the ordinary nouns which are their source. For the other two possibilities, *panie* (women) and *panowie* (men), the genitive plural forms *pań* (women) and *panów* (men) are rare (in this polite possessive use); when they do occur they are normally postposed, as the normal noun would be. Other West Slavonic languages will be considered in § 6.4.

If we turn to the East Slavonic languages, the picture is at first similar to Slovak, in that there is the same matching of forms. We start from Russian (Figure 7):

		personal pronoun		possessive pronoun	
		SG	PL	SG	PL
1	NOM	ja	my	ADJECTIVAL moj …	ADJECTIVAL naš…
	GEN	menja	nas		
	…	…	…		
2	NOM	ty	vy	ADJECTIVAL tvoj …	ADJECTIVAL vaš…
	GEN	tebja	vas		
	…	…	…		
3 M	NOM	on	oni	ego	ix
	GEN	ego	ix		
	…	…	…		
3 F	NOM	ona	oni	ee	ix
	GEN	ee	ix		
	…	…	…		
3 N	NOM	ono	oni	ego	ix
	GEN	ego	ix		
	…	…	…		

Figure 7. Personal pronouns and possessive pronouns in modern standard Russian

As discussed in §§ 6.2.1–6.2.2, there are ways in which the uninflecting forms of the possessive pronoun have differentiated themselves from the genitive of the personal pronoun, but in terms of inflecting/non-inflecting, the picture in Figure 7 is as in Old Church Slavonic (minus the dual). And yet, there have been signs of inflecting forms appearing. There is a third plural inflecting form *ixnij* "their", which has been current in the spoken language at least since the 19th century. In their survey of newspaper texts, Graudina, Ickovič & Katlinskaja (1976: 250–251) found it used around 1% of the time in their sample (4/393). It is more common in spoken language; the singular corresponding forms *evonnyj* "his, its" and *ejnyj* "hers" are roundly condemned by Graudina, Ickovič & Katlinskaja (1976: 251), and they do not offer any statistics. Note, then, that the inflecting forms exist, and that there is a significant difference between singular and plural. There is helpful discussion in

Volk (2014), on the grammatical status of the uninflecting forms and the progress of the inflecting forms (slowed by the conservatism of the standard language).

In Ukrainian, we find the basic pattern similar to Russian. But in Ukrainian in the third person plural there is both uninflecting *jix* "their" and inflecting *jixnij*, and the inflecting variant is the more common (Pugh & Press 1999: 176–177). And in Belarusian the third person possessive pronoun has both uninflecting and inflecting forms, with again the inflecting plural forms being ahead of the singular forms in general acceptability (Mayo 1993: 905–907).

In the South Slavonic part of the family we find a contrasting picture. As we saw in § 6.2.3, Serbo-Croat has a full set of possessive pronouns, with the third person pronouns built on stems with an initial *n*-, and this is largely what we find for the whole branch. We shall look at Serbo-Croat's closest relative, Slovenian, in § 6.5. The remaining languages, Bulgarian and Macedonian have a full set of inflecting possessive pronouns. They have both reduced their case system substantially, following the Balkan trend. This has included losing the genitive, which has the effect of making the distinction between possessive pronoun and personal pronoun even clearer. On the other hand, dative clitic pronouns can be used to indicate possession, more in Bulgarian than in Macedonian; for Bulgarian see Scatton (1993: 204–205, 237) and for Macedonian see Friedman (1993: 264–266, 293–294). Bulgarian shows an interesting twist: the third person nominative forms of the pronoun are built on the stem with initial *t*- rather than initial *n*- (singular *toj* (M), *tja* (F), *to* (N), plural *te*); however, the possessive, as in the sister languages has the *n*- in the singular (*negov* "his, its" and *nejn* "her"); on the other hand, in the plural, the possessive is *texen* "their" (dialectally, *nixen* "their" is also found, Feuillet 1999: 155).

None of the languages analysed in this section has retained the dual, and this automatically reduces the potential contrast between personal pronoun and possessive pronoun. The more interesting situation is that where the dual is preserved, since then there is potentially the greatest split between the agreeing and non-agreeing possessives. We turn to these languages next.

6.4 Dual preserved: First outcome: Upper and Lower Sorbian

The languages which have preserved the dual come from different branches of the family: Upper and Lower Sorbian from the West Slavonic branch, and Slovenian from the South Slavonic branch. We start with Lower Sorbian, whose forms are given in Figure 8.

This pattern has lasted a thousand years: it is exactly as in Old Church Slavonic (Figure 4). Even the actual forms are uncannily similar, once regular sound change and current orthography are allowed for.

		personal pronoun			possessive pronoun		
		SG	DU	PL	SG	DU	PL
1	NOM	ja	mej	my	ADJECTIVAL		ADJECTIVAL
	GEN	mje	naju	nas	mój …	naju	naš…
	…	…	…	…			
2	NOM	ty	wej	wy	ADJECTIVAL		ADJECTIVAL
	GEN	tebje	waju	was	twój …	waju	waš…
	…	…	…	…			
3M	NOM	wón	wónej	wóni			
	GEN	jogo	jeju	jich	jogo	jeju	jich
	…	…	…	…			
3F	NOM	wóna	wónej	wóni			
	GEN	jeje	jeju	jich	jeje	jeju	jich
	…	…	…	…			
3N	NOM	wóno	wónej	wóni			
	GEN	jogo	jeju	jich	jogo	jeju	jich
	…	…	…	…			

Figure 8. Personal pronouns and possessive pronouns in Lower Sorbian

Notes:

i. Full forms are given (there are also some clitic forms).

ii. The data are from Janaš (1984: 172–176) and Stone (1993: 622–623), updated according to the 2006 spelling reform.

When we turn to Upper Sorbian, the picture appears to be the same; there are some differences in the forms, which need not detain us (Faßke 1981: 544–545, 554; Stone 1993: 621–625).[23] But there is a surprise; the third singular feminine has an optional inflecting form: *jejny, jejna, jejne* "her", which Faßke calls colloquial. This is the same development as in Czech, except that in Upper Sorbian the adjectival possessive pronoun is in competition with the uninflecting form. Dialect data are also of great interest. The southern dialect of Rodewitz has only the inflecting form of the third singular feminine (it is not optional as in the standard language). And the remaining third singular pronoun (masculine and neuter) has inflecting forms which are used occasionally. Thus, we see the inflecting forms gradually spreading. However, we cannot see the effect on the dual, since this dialect has lost it (Jentsch 1980: 95, 134, 159).[24]

The basic picture is clear: both Upper and Lower Sorbian essentially preserve the system from the earliest Slavonic texts. There is some limited encroachment of inflecting forms, initially in the feminine singular (as we saw also in Czech, §6.3).

23. The uninflecting forms used as possessive pronouns are always prenominal (Faßke 1981: 554).

24. For the effects of language contact (with German) on the system, see Scholze (2008: 67–71).

6.5 Dual preserved: Second outcome: Slovenian

The dual is also preserved in Slovenian, and here the outcome is quite different, as Figure 9 indicates.

		personal pronoun			possessive pronoun		
		SG	DU	PL	SG	DU	PL
1	NOM	jàz	mídva (F médve)	mí (F mé)	ADJECTIVAL	ADJECTIVAL	ADJECTIVAL
	GEN	mêne	náju	nàs	mój …	nájin …	nàš …
	…	…	…	…			
2	NOM	tí	vídva (F védve)	ví (F vé)	ADJECTIVAL	ADJECTIVAL	ADJECTIVAL
	GEN	têbe	váju	vàs	tvój …	vájin …	vàš …
	…	…	…	…			
3M	NOM	òn	ônadva	ôni	ADJECTIVAL	ADJECTIVAL	ADJECTIVAL
	GEN	njêga	njíju	njìh	njegóv …	njún …	njíhov …
	…	…	…	…			
3F	NOM	ôna	ônidve / onédve	ône	ADJECTIVAL	ADJECTIVAL	ADJECTIVAL
	GEN	njé	njíju	njìh	njén …	njún …	njíhov …
	…	…	…	…			
3N	NOM	ôno	ônidve	ôna	ADJECTIVAL	ADJECTIVAL	ADJECTIVAL
	GEN	njêga	njíju	njìh	njegóv …	njún …	njíhov …
	…	…	…	…			

Figure 9. Personal pronouns and possessive pronouns in Slovenian

Standard Slovenian has varieties with tone and without; for the forms according to the tonemic system see Priestly (1993: 407–409) and for the non-tonemic system, as here, see Herrity (2000: 89–90, 94–95). There are additional gender forms in the first and second persons, which I have simply added in to the relevant cell, since they do not affect the relations with the possessive pronoun. For the renewal of the dual within the personal pronoun see Jakop (2008: 56–70).

The key point is that Slovenian has completed the set of inflecting possessive pronoun forms, corresponding to a personal pronoun which preserves the dual.[25] The rise of the dual possessive pronoun forms is detailed in Tesnière's monumental study (1925); they make their appearance relatively late: the earliest example of a dual possessive pronoun dates from 1677 (Tesnière 1925: 274–275). Tesnière also gives detail on the various forms in different Slovenian dialects. Thus Slovenian illustrates well the second possible outcome: each form of the possessive pronoun is distinct from the personal pronoun, and each form inflects for the set of gender, number and case feature values required of an adjective in Slovenian.

25. Tom Priestly (personal communication 18 November, 2019) confirms that this is true too for the dialect of Sele. More generally, completing the paradigms of inflecting possessive pronouns involves the various suffixes used for deriving possessive and relational adjectives from nouns.

7. Conclusion

The notion of competition has provided a novel insight into Slavonic possessives, and equally the Slavonic possessives show the need to recognize a particular type of competition. That new type is competition based on the featural specification of the primary lexeme. The featural specification which is needed to define the competition of the possessive adjective with the noun is motivated, but the featural specification needed for the rather different competition of personal pronoun vs possessive pronoun is rather surprising in Old Church Slavonic. Here first and second person singular and plural have a dedicated possessive pronoun, while pronouns of the third person or the dual number do not. The competition has been playing out for 1000 years. In broad brush terms, both the possessive adjective (§ 2) and the possessive pronoun (§§ 3–6) have made progress in expanding their syntactic niches as opposed to their competitors (nouns in the genitive and pronouns in the genitive, respectively). Overall, the possessive pronoun is the more successful of the two. Step by step, they are developing their morphology to match their syntactic niche. But when we look closely, we see that progress has not been smooth. And the outcomes have been rather different, particularly for the pronouns, as is clear when we compare the languages with the richest systems of personal pronouns: Upper and Lower Sorbian on the one hand and Slovenian on the other. The competition has been complex, and it continues.

Acknowledgements

It is a pleasure to dedicate this paper to Mark Aronoff, who has long been a leader in linguistics, regularly pointing to new areas to investigate (such as competition), often helping us to understand the intellectual path that has brought us to where we are. He demonstrates how to do all this in a way which combines insight with good-nature and wit, as recently in his account of affix hopping and analytical methods (Aronoff 2019a). I thank these colleagues for their help in tackling competition: Matthew Baerman, Wayles Browne, Marina Chumakina, Michael Daniel, Hanne Eckhoff, Tim Feist, Victor Friedman, Marcia L. Haag, Iván Igartua, Steven Kaye, Ranko Matasović, Thomas Menzel, Sedigheh Moradi, Tore Nesset, Motoki Nomachi, John Payne, Tom Priestly, Janie Rees-Miller, Erich Round, Helen Sims-Williams, Andrew Spencer, Gerald Stone, Thomas Stolz, Anna Thornton, as well as Nick Evans whose question about Old Church Slavonic sparked off this research, and Penny Everson and Lisa Mack for help in preparing the manuscript.

Funding

The support of the AHRC (grant: AH/N006887/1: Lexical splits: a novel perspective on the structure of words) is gratefully acknowledged.

References

Ackermann, Tanja, Horst J. Simon & Christian Zimmer (eds.). 2018. *Germanic genitives* (Studies in Language Companion Series 193). Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.193

Aikhenvald, Alexandra Y. & R. M. W. Dixon (eds.). 2013. *Possession and ownership: A cross-linguistic typology*. Oxford: Oxford University Press.

Aronoff, Mark. 2016. Competition and the lexicon. In Annibale Elia, Claudio Iacobini & Miriam Voghera (eds), *Livelli di Analisi e fenomeni di interfaccia. Atti del XLVII congresso internazionale della società di linguistica italiana (SLI): Fisciano, Salerno 26–28 settembre 2013*, 39–51. Roma: Bulzoni Editore.

Aronoff, Mark. 2019a. Canonical syncretism and Chomsky's S. In Matthew Baerman, Oliver Bond & Andrew Hippisley (eds.), *Morphological perspectives*, 138–147. Edinburgh: Edinburgh University Press.

Aronoff, Mark. 2019b. Competitors and alternants in linguistic morphology. In Franz Rainer, Francesco Gardani, Wolfgang U. Dressler & Hans Christian Luschützky (eds), *Competition in inflection and word-formation* (Studies in morphology 5), 39–66. Cham: Springer. https://doi.org/10.1007/978-3-030-02550-2_2

Baayen, R. Harald, Anna Endresen, Laura A. Janda, Anastasia Makarova & Tore Nesset. 2013. Making choices in Russian: Pros and cons of statistical methods for rival forms. *Russian Linguistics* 37. 253–291. https://doi.org/10.1007/s11185-013-9118-6

Bartnicka, Barbara, Björn Hansen, Wojtek Klemm, Volkmar Lehmann & Halina Satkiewicz. 2004. *Grammatik des Polnischen* (Slavolinguistica 5). Munich: Otto Sagner.

Belić, Aleksandar. 1932. *O dvojini u slovenskim jezicima*. Belgrade: Srpska kraljevska Akademija.

Billings, Loren A. 1996. Sandhi phenomena and language change. In Ursula Kleinhenz (ed.) *Interfaces in phonology* (Studia Grammatica 41), 60–82. Berlin: Akademie Verlag.

Börjars, Kersti, David Denison & Alan K. Scott (eds). 2013. *Morphosyntactic categories and the expression of possession* (Linguistik Aktuell/Linguistics Today 199). Amsterdam: John Benjamins. https://doi.org/10.1075/la.199

Bratishenko, Elena. 1998. Morphosyntactic variation in possessive constructions and the accusative in Old East Slavic texts. Toronto, Canada: University of Toronto dissertation.

Bratishenko, Elena. 2005. Denominal adjective formation and suffix hierarchy in Old East Slavic. *Slavic and East European Journal* 49. 361–377. https://doi.org/10.2307/20058299

Browne, Wayles. 1993. Serbo-Croat. In Bernard Comrie & Greville G. Corbett (eds.), *The Slavonic Languages*, 306–387. London: Routledge.

Bugarski, Ranko. 2012. Language, identity and borders in the former Serbo-Croatian area. *Journal of Multilingual and Multicultural Development* 33. 219–235. https://doi.org/10.1080/01434632.2012.663376

Bugarski, Ranko. 2019. Past and current developments involving pluricentric Serbo-Croatian and its official heirs. In Tanneke Schoonheim & Johan Van Hoorde (eds.), *Language variation: A factor of increasing complexity and a challenge for language policy within Europe: Contributions to the EFNIL Conference 2018 in Amsterdam*, 105–114. Budapest: Research Institute for Linguistics, Hungarian Academy of Sciences. Available at: http://www.efnil.org/documents/conference-publications/amsterdam-2018

Carlier, Anne & Jean-Christophe Verstraete (eds.). 2013. *The genitive* (Case and Grammatical Relations across Languages 5). Amsterdam: John Benjamins. https://doi.org/10.1075/cagral5

Corbett, Greville G. 1987. The morphology/syntax interface: Evidence from possessive adjectives in Slavonic. *Language* 63. 299–345. https://doi.org/10.2307/415658
Corbett, Greville G. 2010. Canonical derivational morphology. *Word Structure* 3. 141–155. https://doi.org/10.3366/word.2010.0002
Corbett, Greville G. 2012. *Features*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139206983
Corbett, Greville G. 2015. Morphosyntactic complexity: A typology of lexical splits. *Language* 91. 145–193. https://doi.org/10.1353/lan.2015.0003
Corbett, Greville G. & Wayles Browne. 2018. Serbo-Croat: Bosnian, Croatian, Montenegrin, Serbian. In Bernard Comrie (ed.), *The world's major languages*, 3rd edn., 339–356. London: Routledge. https://doi.org/10.4324/9781315644936-19
Daniel, Michael. 2015. Stem initial alternation in Russian third person pronouns: Variation in grammar. In Vladimir P. Selegej (main editor), *Komp'juternaja lingvistika i intellektual'nye texnologii: po materialam ežegodnoj Meždunarodnoj konferencii "Dialog" (2015)*, I, 95–103. Moscow: RGGU.
Dench, Alan & Nicholas Evans. 1988. Multiple case-marking in Australian languages. *Australian Journal of Linguistics* 8. 1–47. https://doi.org/10.1080/07268608808599390
Eckhoff, Hanne M. 2009. A usage-based approach to change: Old Russian possessive constructions. In Johanna Barðdal & Shobhana L. Chelliah (eds.), *The role of semantic, pragmatic, and discourse factors in the development of case*, 161–180. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.108.10eck
Eckhoff, Hanne M. 2011. *Old Russian possessive constructions: A construction grammar approach*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110255041
Eckhoff, Hanne M. 2018. Quantifying syntactic influence: Word order, possession and definiteness in Old Church Slavonic and Greek. In Jasmina Grkovic-Major, Björn Hansen & Barbara Sonnenhauser (eds.), *Diachronic Slavonic syntax: The interplay between internal development, language contact and metalinguistic factors*, 29–61. Berlin: De Gruyter.
Eckhoff, Hanne M. & Aleksandrs Berdičevskis. 2015. Linguistics vs. digital editions: The Tromsø Old Russian and OCS Treebank. *Scripta & e-Scripta* 14–15. 9–25.
Faßke, Helmut. 1981. *Grammatik der obersorbischen Schriftsprache der Gegenwart: Morphologie*. Bautzen: Domowina.
Ferrell, James. 1958. The systems of the third person pronoun with particular reference to the category of the preposition in Slavic languages. In *American Contributions to the Fourth International Congress of Slavists, Moscow, September 1958*, 77–102. The Hague: Mouton.
Feuillet, Jack. 1999. *Grammaire historique du Bulgare*. Paris: Institut d' Études slaves.
Flier, Michael S. 1974. *Aspects of nominal determination in Old Church Slavic*. The Hague: Mouton. https://doi.org/10.1515/9783110888379
Friedman, Victor A. 1993. Macedonian. In Bernard Comrie & Greville G. Corbett (eds.), *The Slavonic languages*, 249–305. London: Routledge.
Gardani, Francesco, Franz Rainer & Hans Christian Luschützky. 2019. Competition in morphology: A historical outline. In Franz Rainer, Francesco Gardani, Wolfgang U. Dressler & Hans Christian Luschützky (eds), *Competition in inflection and word-formation* (Studies in morphology 5), 3–36. Cham: Springer. https://doi.org/10.1007/978-3-030-02550-2_1
Graudina, Ljudmila K., Viktor A. Ickovič & Lija P. Katlinskaja. 1976. *Grammatičeskaja pravil'nost' russkoj reči: opyt častotno-stilističeskogo slovarja variantov*. Moscow: Nauka.

Guðmundsson, Helgi. 1972. *The pronominal dual in Icelandic* (University of Iceland Publications in Linguistics 2). Reykjavík: Institute of Nordic Linguistics.

Herrity, Peter. 2000. *Slovene: A comprehensive grammar*. London: Routledge.

Hill, Steven P. 1977. *The N-factor and Russian prepositions: Their development in 11th-20th century texts*. The Hague: Mouton. https://doi.org/10.1515/9783110801910

Huntley, David. 1984. The distribution of the denominative adjective and the adnominal genitive in Old Church Slavonic. In Jacek Fisiak (ed.), *Historical syntax*, 217–236. Berlin: Mouton. https://doi.org/10.1515/9783110824032.217

Huntley, David. 1993. Old Church Slavonic. In Bernard Comrie & Greville G. Corbett (eds) *The Slavonic languages*, 125–187. London: Routledge.

Isačenko, Aleksandr V. 1954. *Grammatičeskij stroj russkogo jazyka v sopostavlenii s slovackim: Morfologija: I*. Bratislava: Izdatel'stvo Slovackoj Akademii Nauk.

Jakop, Tjaša. 2008. *The dual in Slovene dialects* (Diversitas Linguarum 18). Bochum: Universitätsverlag Dr. N. Brockmeyer.

Janaš, Pětr. 1984. *Niedersorbische Grammatik* (2nd edn.). Bautzen: Domowina.

Janda, Laura A., Tore Nesset & Sergey Say. 2019. *Prezidentskie vybory* vs. *vybory prezidenta*: How to choose? *Russian Linguistics* 43. 181–204. https://doi.org/10.1007/s11185-019-09214-w

Jentsch, Helmut. 1980. *Die sorbische Mundart von Rodewitz/Spree* (Schriftenreihe des Instituts für sorbische Volksforschung in Bautzen 47). Bautzen: Domowina.

Johanson, Lars, Lidia Federica Mazzitelli & Irina Nevskaya (eds). 2019. *Possession in languages of Europe and North-Central Asia* (SLCS 206). Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.206

Juge, Matthew L. 2019. The sense that suppletion makes: Towards a semantic typology on diachronic principles. *Transactions of the Philological Society* 117. 390–414. https://doi.org/10.1111/1467-968X.12175

Koptjevskaja-Tamm, Maria. 2003. Possessive noun phrases in the languages of Europe. In Frans Plank (ed.), *Noun phrase structure in the languages of Europe*, 621–722. Berlin: Mouton de Gruyter.

Kopčevskaja-Tamm, Marija [=Koptjevskaja-Tamm, Maria] & Aleksej Šmelev. 1994. Alešina s Mašej stat'ja (o nekotoryx svojstvax russkix pritjažatel'nyx prilagatel'nyx). *Scando-Slavica* 40. 209–228. https://doi.org/10.1080/00806769408601049

Křivan, Jan. 2019. Pronominal and adjectival attributive possession in spoken Czech: A usage-based perspective. In Lars Johanson, Lidia Federica Mazzitelli & Irina Nevskaya (eds.), *Possession in languages of Europe and North-Central Asia*, 149–168. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.206.08kri

Lander, Yury A. 2009. Varieties of genitive. In Andrej Malchukov & Andrew Spencer (eds.), *The Oxford handbook of case*, 581–592. Oxford: Oxford University Press.

Liljegren, Henrik. 2016. *A grammar of Palula* (Studies in Diversity Linguistics 8). Berlin: Language Science Press. https://doi.org/10.26530/OAPEN_611690

Maiden, Martin. 1995. A proposito dell'alternanza *esce, usciva*, in italiano. *Lingua Nostra* 56. 37–41.

Matasović, Ranko. 2010. Possessive genitives and adjectives from the historical point of view. In Motoki Nomachi (ed.), *The Grammar of possessivity in South Slavic languages: Synchronic and diachronic perspectives* (Slavic Eurasian Studies 24), 1–12. Hokkaido: Slavic-Eurasian Research Center.

Matthews, Peter H. 1997. *The concise Oxford dictionary of linguistics*. Oxford: Oxford University Press.

Mayo, Peter. 1993. Belorussian. In Bernard Comrie & Greville G. Corbett (eds.), *The Slavonic languages*, 887–946. London: Routledge.

Melʹčuk, Igor' A. 1985. *Poverxnostnyj sintaksis russkix čislovyx vyraženij*. Vienna: Institut für Slawistik der Universität Wien. https://doi.org/10.3726/b12909

Minlos, Philip. 2012. Prenominal and postnominal position of adjective attributes in Old Russian. *Russian Linguistics* 36. 21–40. https://doi.org/10.1007/s11185-011-9087-6

Nikolaeva, Irina & Andrew Spencer. 2013. Possession and modification – a perspective from Canonical Typology. In Dunstan Brown, Marina Chumakina & Greville G. Corbett (eds.), *Canonical morphology and syntax*, 207–238. Oxford: Oxford University Press.

Nikolaeva, Irina & Andrew Spencer. 2020. *Mixed categories: The morphosyntax of noun modification*. Cambridge: Cambridge University Press.

O'Connor, Catherine, Joan Maling & Barbora Skarabela. 2013. Nominal categories and the expression of possession: A cross-linguistic study of probabilistic tendencies and categorical constraints. In Kersti Börjars, David Denison & Alan Scott (eds.), *Morphosyntactic categories and the expression of possession*, 89–121. Amsterdam: John Benjamins. https://doi.org/10.1075/la.199.04oco

Olander, Thomas. 2015. *Proto-Slavic inflectional morphology: A comparative handbook*. Leiden: Brill. https://doi.org/10.1163/9789004270503

Payne, John & Eva Berlage. 2014. Genitive variation: The niche role of the oblique genitive. *English Language and Linguistics* 18(2). 331–360. https://doi.org/10.1017/S1360674314000057

Philippova, Tatiana. 2018. Prepositional repercussions in Russian: Pronouns, comparatives and ellipsis. Beersheba, Israel: Ben-Gurion University of the Negev thesis.

Ponsonnet, Maïa. In preparation. Possession. In Claire Bowern (ed.), *Oxford handbook of Australian languages*. Oxford: Oxford University Press.

Priestly, Tom M. S. 1993. Slovene. In Bernard Comrie & Greville G. Corbett (eds.), *The Slavonic languages*, 388–451. London: Routledge.

Pugh, Stefan M. & Ian Press. 1999. *Ukrainian: A comprehensive grammar*. London: Routledge.

Richards, Karen Rondestvedt. 1976. Toward a history of the possessive in literary Russian: The demise of the possessive adjective. In Sanford B. Steever, Carol A. Walker & Salikoko S. Mufwene (eds.), *Papers from the parasession on diachronic syntax*, 260–273. Chicago: Chicago Linguistic Society.

Rosenbach, Anette. 2014. English genitive variation – the state of the art. *English Language and Linguistics* 18(2). 215–262. https://doi.org/10.1017/S1360674314000021

Round, Erich & Greville G. Corbett. 2020. Comparability and measurement in typological science: The bright future for linguistics. *Linguistic Typology* 24(3). 489–525. https://doi.org/10.1515/lingty-2020-2060

Scatton, Ernest A. 1993. Bulgarian. In Bernard Comrie & Greville G. Corbett (eds.), *The Slavonic languages*, 188–248. London: Routledge.

Schmalstieg, William R. 1983. *An introduction to Old Church Slavic* (2nd edn., revised and expanded). Columbus, OH: Slavica.

Scholze, Lenka. 2008. *Das grammatische System der obersorbischen Umgangssprache im Sprachkontakt: mit Grammatiktafeln im Anhang* (Schriften des Sorbischen Instituts 45). Bautzen: Domowina.

Short, David. 1993a. Czech. In Bernard Comrie & Greville G. Corbett (eds.), *The Slavonic languages*, 455–532. London: Routledge.
Short, David. 1993b. Slovak. In Bernard Comrie & Greville G. Corbett (eds.), *The Slavonic languages*, 533–592. London: Routledge.
Šmeleva, T. V. 2008. Pritjažatel'nye prilagatel'nye: Počemu ne sbyvaetsja vinogradovskij prognoz? In A. Mustajoki, M. V. Kopotev, L. A. Birjulin & E. Ju. Protasova (eds.), *Instrumentarij rusistiki: korpusnye podxody* (Slavica Helsingiensia 34), 358–371. Helsinki: University of Helsinki.
Spencer, Andrew. 2002. Gender as an inflectional category. *Journal of Linguistics* 38. 279–312. https://doi.org/10.1017/S0022226702001421
Spencer, Andrew. 2013. *Lexical relatedness: A paradigm-based model*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199679928.001.0001
Stolz, Thomas, Sonja Kettler, Cornelia Stroh & Aina Urdze. 2008. *Split possession: An areal-linguistic study of the alienability correlation and related phenomena in the languages of Europe* (Studies in Language Companion Series 101). Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.101
Stone, Gerald. 1993. Sorbian (Upper and Lower). In Bernard Comrie & Greville G. Corbett (eds.), *The Slavonic languages*, 593–685. London: Routledge.
Stump, Gregory. 2016. *Inflectional paradigms: Content and form at the syntax-morphology interface*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781316105290
Stump, Gregory. 2019. An apparently noncanonical pattern of morphotactic competition. In Franz Rainer, Francesco Gardani, Wolfgang U. Dressler & Hans Christian Luschützky (eds) *Competition in inflection and word-formation* (Studies in morphology 5), 259–278. Cham: Springer. https://doi.org/10.1007/978-3-030-02550-2_10
Suprun, Adam E. 1969. *Slavjanskie čislitel'nye (stanovlenie čislitel'nyx kak osoboj časti reči)*. Minsk: Belorussian State University.
Tesnière, Lucien. 1925. *Les formes du duel en slovène*. Paris: Champion.
Thornton, Anna. 2019a. Overabundance in morphology. In *Oxford research encyclopedia of linguistics*. Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.554
Thornton, Anna. 2019b. Overabundance: A canonical typology. In Franz Rainer, Francesco Gardani, Wolfgang U. Dressler & Hans Christian Luschützky (eds.) *Competition in inflection and word-formation* (Studies in morphology 5), 223–258. Cham: Springer. https://doi.org/10.1007/978-3-030-02550-2_9
Timberlake, Alan. 2004. *A reference grammar of Russian*. Cambridge: Cambridge University Press.
Trubetzkoy, Nikolai S. 1937. O pritjažatel'nyx prilagatel'nyx (possessiva) starocerkovno-slavjanskogo jazyka. *Zbornik lingvističkih i filoloskih rasprava A. Beliću o četrdesetogodišnjici njegova naučnog rada posvećuju njegovi prijatelji i učenici*, 15–20. Belgrade: Mlada Srbija.
Vaillant, André. 1964. *Manuel du vieux slave: I: Grammaire* (2nd edn.). Paris: Institut d' Études slaves.
Vaillant, André. 1977. *Grammaire comparée des langues slaves: V: La syntaxe*. Paris: Klincksieck.
Večerka, Radoslav. 1989. *Altkirchenslavische (Altbulgarische) Syntax: I: Die lineare Satzorganisation*. Freiburg im Breisgau: U.W. Weiher.
Volk, Vitalij S. 2014. Sintaksis pritjažatel'nyx mestoimenij i ad''ektivnja derivacija. *Acta Linguistica Petropolitana: Trudy instituta lingvističeskix issledovanij* 2. 510–533: https://cyberleninka.ru/article/n/sintaksis-prityazhatelnyh-mestoimeniy-i-adektivnaya-derivatsiya

Williams, Edwin. 2007. Dumping lexicalism. In Gillian Ramchand & Charles Reiss (eds.), *The Oxford handbook of linguistic interfaces*, 353–381. Oxford: Oxford University Press.
Žolobov, Oleg F. 1998. *Symbolik und historische Dynamik des slavischen Duals / Simvolika i istoričeskaja dinamika slavjanskogo dvojstvennogo čisla* (Beiträge zur Slavistik 35). Frankfurt am Main: Peter Lang.
Žolobov, Oleg F. & Vadim B. Krys'ko. 2001. *Dvojstvennoe čislo*. Moscow: Azbukovnik.

CHAPTER 11

Competition in comparatives

A look at Romance scenarios

Anna M. Thornton
Università dell'Aquila

This chapter provides an initial exploration of the phenomenon of competition in comparatives in Romance languages. Latin had both synthetic and periphrastic comparatives, whose distribution was phonologically conditioned. Romance languages have generalized the periphrastic construction to all adjectives, but have also often inherited continuants of all or some of the four synthetic suppletive comparatives *melior* "better", *peior* "worse", *māior* "bigger", *minor* "smaller". Therefore, in the evolution from Latin to Romance languages, the stage is set for competition between synthetic and analytic constructions, such as Italian *migliore* vs. *più buono* "better". This chapter discusses the solutions found in different Romance languages, as explained in descriptive grammars and other works. These solutions span from extinction of the synthetic suppletive forms altogether, to blocking of the analytic construction when the synthetic one is available, to overabundance between synthetic and analytic forms, to the creation of niches in which the two constructions distribute differentially. A corpus-based pilot study of Italian suggests a fruitful way to uncover such niches.

Keywords: comparative, suppletive comparative, synthetic comparative, analytic comparative, competition, niche, Romance languages, *migliore*, *più buono*

1. Comparatives from Latin to Romance

Classical Latin displayed a well-behaved grammar for the formation of comparatives based on adjectives. Comparatives were formed synthetically, by means of the suffix *-ior* (M and F) / *-ius* (N), except in the case of adjectives whose stem ended in a vowel, which formed their comparatives periphrastically, by means of the adverb *magis* positioned before the positive degree form:[1]

1. Cuzzolin (2011: 542) observes that some exceptions to this rule are attested, such as synthetic forms like *strenuior* "more strenuous", *idoneior* "more apt"; and usage of *magis* with forms that normally have a synthetic comparative is attested, like in the following passage by Ovid, cited by

https://doi.org/10.1075/cilt.353.11tho

(1) a. *alt-us*$_{M}$, *alt-a*$_{F}$, *alt-um*$_{N}$ *alt-ior*$_{M/F}$, *alt-ius*$_{N}$
"high, tall" "higher, taller"
b. *idōneus* *magis idōneus*
"apt" "more apt"

Some adjectives had synthetic comparatives formed on the basis of a suppletive stem:

(2) *bonus* "good" *melior* "better"
malus "bad" *peior* "worse"
magnus "big" *māior* "bigger"
parvus "small" *minor* "smaller"

The relation between the positive degree adjectives and the comparatives listed in (2) is traditionally recognized (e.g., by Priscian, who calls such cases "inaequalia" (Hertz 1855: 88)).[2]

Already in Late Latin, there was a tendency to favor analytic comparatives over synthetic ones, and also a tendency to use *magis* or *plus* together with a synthetic comparative, as in the examples in (3) (from Herman 2000: 64):

(3) a. *magis deterius habebat*
"she rather grew worse" [*lit*. "more worse"] (Mark 5.26)
b. *oculi Domini multo plus lucidiores sunt super solem*
"the eyes of the Lord are much brighter [*lit*. "much more brighter"] than the sun" (Sirach 23.28)

The received view about why this happened is expressed by Herman (2000: 64) in the following way:

> there was a tendency in speech to use the explicit comparative adverbs rather than rely on the suffixed forms alone; … the older forms were tending to become semantically devalued, in that their explicitly comparative meaning may no longer have been generally clear.

Salvi (2011: 338): *quid magis est saxo durum, quid mollius unda?* "what is harder (*magis durum* vs. *durius*) than a rock, what is softer (*mollius*) than a wave?". Maltby (2016), based on a thorough investigation of Latin texts, observes that, although Latin grammarians only mentioned adjectives whose stem ends in a vowel as forming the comparative analytically, in fact other kinds of adjectives commonly formed their comparative by preposing *magis*: adjectives ending in *-rus*, *-er* and *-ris*, "long" adjectives (four syllables or more), deverbal adjectives. But the amount of competition between synthetic and analytic comparatives from the same base is negligible in Classical Latin.

2. The notion of suppletion is not found in the Latin grammarians and appears to have been introduced by Hermann Osthoff (1900).

Romance languages eventually abandoned the formation of synthetic comparatives, and generalized the analytic construction, with Ibero-Romance and Romanian continuing the construction MAGIS + positive degree adjectives (4a), and the other languages replacing MAGIS with (the descendants of) PLUS (4b):

(4) "higher / taller" in Romance languages
  a. Portuguese *mais alto*
     Spanish *más alto*
     Catalan *més alt*
     Romanian *mai înalt*
     Occitan *mai naut / p(l)us naut*
  b. French *plus haut*
     Romansh[3] *pli aut*
     Italian *più alto*
     Sardinian *prus artu*

However, descendants of all or at least some of the suppletive synthetic comparatives are also used in most Romance languages; a representative list is given in Table 1:[4]

Table 1. Suppletive synthetic comparatives in Latin and some Romance languages

Latin	**Spanish**	**Catalan**	**French**	**Italian**
MELIOR(EM)	*mejor*	*millor*	*meilleur*	*migliore*
PEIOR(EM)	*peor*	*pitjor*	*pire*	*peggiore*
MĀIOR(EM)	*mayor*	*major*	*(majeur)**	*maggiore*
MINOR(EM)	*menor*	*menor*	*moindre*	*minore*

* *Majeur* "is of learnèd origin" (Maiden 2011: 223).

Therefore, in most individual Romance languages the scenario for a potential competition between the synthetic suppletive comparatives and the corresponding analytic forms is established.[5] This chapter is a first exploratory overview of how the

3. These data are from Rumantsch Grischun.

4. Some languages seem to have synthetic forms only for "better" and "worse": this is apparently the case in Romansh (Anderson 2016: 176), dialects of Central Italy (Loporcaro & Paciaroni 2016: 235) and Sardinian (Mensching & Remberger 2016: 278).

5. While in Latin the so-called 'relative superlative' degree was syncretic with the so-called 'absolute superlative' (*altissimus* "the tallest / very tall"), in the Romance languages investigated here the 'relative superlative' is expressed by preposing the definite article to the 'comparative': e.g., Italian *più buono / migliore* "better", *il più buono / il migliore* "the best". In what follows, I will consider the distribution of synthetic and analytic forms in Romance languages without regard to the distinction between comparative and relative superlative; some of the examples in § 3 contain relative superlatives.

different Romance languages reacted to this case of competition. In analogy to what is done in the canonical approach to the study of linguistic phenomena, I first review the theoretical possibilities concerning the relation between two competing comparative forms in Romance languages; then I cursorily show how the extant situation is described by grammars and other descriptions of various languages; finally, I sketch how a corpus-based study could help to individuate the factors governing the selection of one or the other of two competing forms, and I present a few preliminary results about Italian.

This chapter is obviously not a report from a large-scale investigation of competition between synthetic and analytic forms in Romance languages, which has not been conducted yet; it should be taken rather as a preliminary feasibility study for such an investigation, which would ideally require the cooperative effort of several specialists in the different languages, and produce results whose presentation would largely exceed the space available here. It is my hope that the following observations will inspire other scholars to conduct at least parts of such an investigation.

2. Synthetic and analytic expression of 'better', 'worse', 'bigger' and 'smaller' in Romance: Theoretical possibilities and descriptions in grammars

Any Romance language might theoretically have in its grammar two ways of expressing the concepts of 'better', 'worse', 'bigger' and 'smaller': a regular periphrastic formation with a descendant of MAGIS or PLUS followed by a positive degree adjective, and a synthetic form inherited (or acquired as a learnèd borrowing) from the Latin suppletive synthetic comparative.

Such a situation can in turn give rise to at least four possible different scenarios:

1. the analytic construction blocks the synthetic construction
2. the synthetic construction blocks the analytic construction
3. both constructions exist, and are used interchangeably
4. both constructions exist, and are used in different contexts

Aronoff (2016, 2019) has proposed to analyze cases of potential competition between linguistic forms and constructions in the light of Gause's principle of competitive exclusion, which states that no two species with identical ecological niches can coexist in a stable equilibrium. The result of the competition, however, need not be extinction of one of the competitors: both can thrive if they manage to find themselves a niche. The concept of niche in ecology is defined in the OED Online as "The actual or potential position of an organism within a particular ecosystem, as determined by its biological role together with the set of environmental conditions under which it lives" (OED, 2020); Elton (1927), cited by the OED, defines *niche* as a term used "to describe the status of an animal in its community, to indicate what it is doing and not merely what it looks like". If we replace "animal" with "linguistic form or construction" and "community" with "environment", we have a clear parallel between ecological niches for living species and niches for linguistic forms. Aronoff (2019: 44) exploits this metaphor, defining a niche as a "distribution pattern of resource use"; he observes that "[t]here are many ways in which the differential distribution patterns of resource use can occur in nature", among which "*resource partitioning* …, in which two or more competing species divide up the resource (usually what they consume) along some lines, and its subtype, *spatial partitioning*, in which the resource is a distinct area or habitat that each species occupies". Aronoff concedes that it is not easy to identify the linguistic analogues of resource partitioning: "There is no way to know in advance what will constitute a [linguistic, AMT] niche, but the standard linguistic variables (morphology, phonology, and others) usually provide opportunities" (Aronoff 2016: 42).

If we look at scenarios 1–4 listed at the beginning of this section in this framework, considering the analytic and the synthetic comparative constructions as the analogue of species, we could say that in scenarios 1 and 2 one species has completely ousted the competing one, which has become extinct; one species / construction has expanded its niche to the point that it has occupied the entire environment (habitat), eating up all the resources and causing the other one to die out. In scenario 3, there appear to be sufficient resources for both species / constructions to thrive, at least at the synchronic stage we are observing: both find enough resources to keep alive; in linguistics, this situation has been called 'overabundance' (Thornton 2019). Scenario 4 is the closest analogue of ecological niche differentiation in our field of investigation: the synthetic and the analytic constructions distribute in different environments, where 'environment' can be understood as the set of collocations in which each construction occurs; the 'resources' here would be the comparees in a comparative construction.

All four scenarios are claimed to exist in at least some Romance language, according to grammatical descriptions. Unfortunately, grammars and other descriptive works do not always discuss the matter: for example, only 6 of the 17 "structural overviews" of the individual Romance languages in *The Oxford Guide to the Romance Languages* (Ledgeway & Maiden 2016) mention the question, and in two cases – Romanian and Dalmatian – only to inform us that synthetic forms do not exist (Maiden 2016a, 2016b). Therefore, corpus-based studies are called for. However, it is also instructive to look at what existing descriptions claim when they do discuss the question: these data might orient future corpus searches that could confirm or disconfirm the observations found in grammars, which are often based only on educated guesses or speakers' intuitions. Therefore, following the four scenarios listed at the beginning of this section, I will start with offering a quick review of statements found in some descriptive grammars (and, occasionally, other reference works on the individual languages).

2.1 Scenario 1: The analytic construction blocks the synthetic construction

This situation is represented by Romanian, where descendants of the Latin synthetic comparatives are not in use (Brăescu 2013: 416). Apparently, the same situation held in Dalmatian (Maiden 2016b: 132). Besides, according to some descriptions, it is also represented in other languages, for at least one of the four adjectives. For example, for French Grevisse & Goosse (1980: 157, § 206) list only *bon* "good", *petit* "little" and *mauvais* "bad" has having "des formes spéciales" [special forms] of comparative, thus implying that in French *plus grand* "bigger" [*lit*. "more big"] blocks a synthetic descendant of MAIOR "greater, bigger".

2.2 Scenario 2: The synthetic construction blocks the analytic construction

Scenario 2 is claimed to hold in at least one case in French, while descriptions of Spanish disagree on whether this is the scenario represented by the language.

For French, Grevisse (1986: 897, § 552) writes that the synthetic form *meilleur* replaces the analytic form °*plus bon*, which is not allowed in correct French. The observation that *plus bon* is inadmissible "dans le français correct" leaves us wondering what happens in less than correct French; however, it appears that in at least some varieties of French there is a case in which a synthetic form blocks its analytic alternative.

For Spanish, Alcina, Franch & Blecua (1975: 581) claim that the synthetic forms *mejor, peor, mayor* and *menor* "pueden considerarse por supletivismo comparativos de *bueno, malo, grande* y *pequeño*" [can be considered suppletive comparatives of *bueno* "good", *malo* "bad", *grande* "big" and *pequeño* "small"]; they do not mention the possibility of analytic constructions, thus implying that these are blocked. This extreme characterization of the situation is not shared by other authors: Penny (2002: 130) observes that *mayor* and *menor* "compete with analytic *más grande* and *más pequeño*", and implies that only *mejor* and *peor* block *más bueno* and *más malo*; however, yet others claim that both synthetic and analytic forms are used but in different contexts, as will be discussed in § 2.4.

None of the grammars of Italian I consulted (Migliorini 1952; Serianni 1988; Maiden & Robustelli 2000; Salvi & Vanelli 2004) go as far as claiming a blocking relation of the analytic forms by the synthetic ones, and their views will be discussed under scenarios 3 and 4.

2.3 Scenario 3: Both constructions exist, and are used interchangeably

Both the synthetic and analytic versions of all or some of the four comparatives are claimed to be used more or less interchangeably in Catalan, Spanish, and Italian, but not in French.

For Catalan, Wheeler, Yates & Dols (1999: 99) write: "*Major* "bigger", *menor* "smaller", *millor* "better" and *pitjor* "worse" are irregular synthetic comparative forms for *més gran, més petit, més bo* and *més dolent / més mal*. However, the analytic forms are not only correct but even preferred".

For Spanish, Penny (2002: 130) states that *mayor* and *menor* are in competition with *más grande* and *más pequeño*, as reported in § 2.2.

For Italian, Migliorini's (1952: 66) formulation seems to imply free variation between synthetic and analytic constructions (and even some priority of the analytic ones): "Alcuni aggettivi molto usati hanno, oltre alla forma normale del comparativo e del superlativo, altre forme che si dicono organiche" [Some very frequent adjectives have, besides the normal form of comparative and superlative, other forms which are called organic]; he lists 'organic' (i.e., synthetic) comparatives and superlatives for all four of the adjectives meaning "good", "bad", "big" and "small". Serianni (1988: §V.81) first introduces the so-called 'organic' comparatives for the same adjectives, but then observes that the four adjectives also appear with *più* "more" and suggests that these analytic forms are semantically equivalent with the synthetic forms. However, he then proceeds to comment on the cases in which this semantic equivalence does not hold, which are instances of scenario 4.

2.4 Scenario 4: Both constructions exist, and are used in different contexts

Whenever the existence of two synonymous competing constructions in a given language is observed, authors usually take great pains in finding semantic or socio-stylistic conditions that govern the distribution of the two competing forms (see Thornton 2019: 250 for a list of quotations to this effect). Authors of grammars of Romance languages are no exception.

The kinds of factors which are mentioned as regulating the choice between a synthetic and an analytic comparative are at least the following:

1. Concrete vs. abstract meaning
2. Elevated vs. informal register
3. Specific senses of the base adjective
4. Set phrases

For each of these four factors, I offer below a brief representative illustration of what grammars claim for various languages about at least one of the pairs of competing comparatives.

2.4.1 *Concrete vs. abstract meaning*

Claims have been made for both French and Italian that the choice of competing comparatives has to do with concrete vs. abstract meaning; however, in some cases a different explanation may be more accurate.

For French, Judge & Healey (1995: 303–304) maintain that the pairs *plus mauvais / pire* and *plus petit / moindre* are distinguished by the fact that the synthetic forms are used for more abstract meanings, and the analytic forms for more concrete ones. For example, in the sentence, *la nourriture était encore plus mauvaise que celle de l'internat* "the food was even worse [*lit*. more bad] than the one of the boarding school", they claim that the analytic form *plus mauvaise* refers specifically to the taste of the food. Replace the analytic form *plus mauvaise* with the synthetic form *pire*, however, and the meaning is more abstract, referring not only to the taste of the food but to its quality in general. Similarly, the analytic form *plus petit* "smaller" is used for its concrete meaning in a sentence such as *sa voiture est plus petite que la nôtre* "his car is smaller [*lit*. more small] than ours", but when the entity described is changed to a more abstract concept, then the synthetic form *moindre* is used instead, as in *ses chagrins sont moindres que les nôtres* "his sorrows are smaller than ours".

Serianni (1988, §V.81) also maintains that in Italian synthetic forms one can observe "un prevalere dei significati astratti" [a prevalence of abstract meanings], but his example is a set phrase, *il minore dei mali* "the lesser evil" and therefore does not unequivocally support the hypothesis that *minore* is chosen here because we are dealing with an abstract meaning.

2.4.2 *Elevated vs. informal register*

The factor of register is invoked for French and Italian, but not mentioned for Spanish. For French, Grevisse (1986) claims that a difference in register is at play for both *moindre*, which is said to belong to "langue soignée" [elevated language] (§ 553), and *pire*, which has "une teinte plutôt littéraire" [a rather literary hue] (§ 554; see also Grevisse & Goosse 1980: 157). Maiden & Robustelli (2000: 342) also invoke register as a factor for Italian: "The rather learned and elevated connotations of *maggiore* and *minore* also tend to mean that these forms are avoided in informal registers".

2.4.3 *Specific senses of the base adjective*

This factor is very often invoked, particularly in grammars that have a contrastive orientation and are written for second language learners. The most common senses discussed are the following: 'younger' and 'older' as different from 'smaller' and 'bigger'; 'smaller in size' and 'bigger in size' as different from 'less / more important' or the like; 'good' in the sense 'good-natured, kind'. As will become clear in the following representative selection of observations, conclusions based on the intuitions of one native speaker may be open to dispute without a collection of hard empirical data.

2.4.3.1 *'Younger' and 'older'*

Grammars of Spanish are not in agreement as to which competing form is more acceptable for comparative age. Butt & Benjamin (1988: 79) maintain that Spanish allows both *mi hermano menor* and *mi hermano pequeño* for "my younger brother" with *más pequeño* used to the exclusion of *menor* in predicative position. In other words, one should say *Mi hermano es más joven / pequeño* (not: *menor*) *que yo* "my brother is younger than me"; however, a few lines later they quote this literary example: *Virginia era unos meses menor que yo* (A. Mastretta). In contrast, Kattán-Ibarra & Pountain (1997: 31) state that *menor* is "always used if the sense is 'younger', 'youngest'" and *mayor* is "always used if the sense is 'older, 'oldest'". Clearly, these two grammars of Spanish have opposite views on what construction is or should be used for the senses 'younger', 'youngest' and 'older, 'oldest', and only a corpus-based investigation could settle the question.

2.4.3.2 *'Smaller in size' and 'bigger in size'*

The analytic form is said to be more commonly used for size than the synthetic form in Spanish and French. Butt & Benjamin (1988: 79) maintain that in Spanish "*[m]enor* differs from *mayor* in that it cannot refer to dimensions: *Esta habitación es más pequeña/*menor que ésa* [this room is smaller than that one]". Similarly, for French, Grevisse (1986, § 553) maintains that "Lorsqu'il s'agit de choses concrètes mesurables *plus petit* est préféré" [when talking about concrete measurable things *plus petit* [*lit*. more little] is preferred].

2.4.3.3 *'Good-natured, kind'*

The analytic form is claimed to be more often used for describing character than the synthetic form in Spanish and Italian. Kattán-Ibarra & Pountain (1997: 31) state that in Spanish "*más bueno* is sometimes used in the sense of 'better in character or behaviour'", but not to the complete exclusion of *mejor* (and they add a parallel observation concerning *más malo* and *peor*). Butt & Benjamin (1988: 74) concur, observing that "*Más bueno, más malo* are used of moral qualities though *mejor / peor* are more usual".

Italian also uses the analytic form for character. Maiden & Robustelli (2000: 342) maintain that "*Più cattivo* and *più buono* tend to be used – rather than *peggiore* and *migliore* – when the sense is 'more ill-natured', 'more unpleasant' and 'more good-natured', 'nicer', 'more agreeable'".

Apart from these common phenomena that have parallels in different Romance languages, more idiosyncratic facts are also reported: for example, Grevisse (1986, § 552) observes that in French "[l]orsque *bon* signifie 'naïf, crédule', le comparative *meilleur* paraît difficilement acceptable" [when *bon* means "naïve, credulous" the comparative *meilleur* appears hardly acceptable].

2.4.4 *Set phrases*

As mentioned in § 2.4.1 Serianni's (1988) example in Italian, *il minore dei mali* "the lesser evil", might be considered a set phrase. For Occitan, Patrick Sauzet claims explicitly that certain synthetic comparatives are "restricted to specific set phrases":

> There exist residual synthetic comparatives such as *melhor* / *-a* "better.M/F", and others restricted to specific set phrases, e.g., *la màger part* "the greater part", *pas la mendre causa* "not the slightest thing", but *es mai grand que tu* (***màger que tu*) "he is taller than you", *Vilafranca es mai pichòta* (***mendre*) *que Tolosa.*
>
> (Oliviéri & Sauzet 2016: 330)

(Parenthetically, it should be noted here that these examples of *màger* and *mendre* could be just further examples of the fact that synthetic forms tend not to be used for senses like 'bigger / smaller in size', rather than specific set phrases.).

2.5 Interim conclusions

This brief review of grammatical descriptions of some Romance languages has shown that all the theoretically possible outcomes of a competition between synthetic and analytic comparatives appear to be attested in at least one language, for at least one pair of competing constructions. However, authors do not always agree on the details of the usage of each specific form. For example, as discussed in § 2.4.3.1, two grammars of Spanish (Butt & Benjamin 1988; Kattán-Ibarra & Pountain 1997) have opposite views on the usage of *menor* vs. *más pequeño* when the sense is 'younger'.

Clearly, our understanding of the contexts of usage of each of the competing constructions in each Romance language would benefit greatly from a corpus-based investigation. Such an enterprise exceeds by far the goals of the present chapter and the space available, but in the next section I will report on some preliminary results concerning Italian *più buono* and *migliore*. These data illustrate the kind of information one could obtain from corpus-based studies on our topic and the fact that even a very preliminary investigation uncovers data that have not been noticed by descriptions based only on native competence.

3. *Più buono* vs. *migliore* in Italian: Niches

The frequency of occurrence of *più buono* and *migliore* (in all their inflected forms) in Italian is quite unbalanced, as illustrated in Table 2, which shows data from the LIP corpus of spoken Italian and from *la Repubblica* corpus of written Italian:[6]

Table 2. Frequency of *più buono lit*. "more good" and *migliore* "better" in corpora of spoken and written Italian

Corpus	***Più buono***	***Migliore***	**Ratio**
LIP (spoken Italian, 500K tokens)	4	55	1:14
*la Repubblica** (written Italian, 320M tokens)	554	~72000	1:130

* Data from Santilli (2014: 46).

In both corpora *migliore* is much more frequent than *più buono*, but the magnitude of the disproportion is different, in the range of tens in the spoken corpus and in the range of hundreds in the written corpus.

The small number of tokens in the LIP corpus allows us to conduct an exhaustive scrutiny of the contexts in which the two constructions occur. The 4 tokens of *più buono* divide themselves neatly in two fields: 2 tokens apply to human beings, in a homily by a Catholic priest who talks about becoming "più buoni" during Lent (5a), and 2 tokens apply to wine (specifically, and somewhat surprisingly in my view, to Lancers) (5b):

(5) a. *c' e' quaresima van fatti sacrifici cosi' il signore chiede ti fanno diventare piu' buono e viene $ per diventare piu' buono e allora faro' questo sacrificio*
"it's Lent, we must make sacrifices, that's what the Lord asks, they make you become better [*lit*. more good] and he comes $ to become better [*lit*. more good] and then I will make this sacrifice." (F D 14 1 A)

b. A: *il Lancers e' piu' buono*
"Lancers is better [*lit*. more good]"
B: ** piu' buono **
"better [*lit*. more good]" (R A 1 188 A – 189 B)

6. The LIP corpus of spoken Italian (De Mauro et al. 1993) contains about 500,000 word tokens collected between 1990 and 1992 in four Italian cities (Milan, Florence, Rome, Naples) and representing five different diaphasic situations (face-to-face conversations; telephone conversations; interviews, debates, and classroom interactions; lectures, sermons, and speeches; radio and television programs); the total number of words per city and per diaphasic situation is balanced; the corpus was searched by means of the search tool available at http://badip.uni-graz.at/. The *la Repubblica* corpus is a corpus of journalistic prose, containing about 320,000,000 word tokens, representing 16 years of issues (1985–2000) of the daily newspaper *la Repubblica*; the corpus was searched by means of the search tool available at https://corpora.dipintra.it/public/.

Migliore has a much wider distribution, but interestingly in the LIP corpus it is never used to qualify food or beverages, or human kindness. When applied to humans, *migliore* refers to the fact that these people are good at their job, or are good students, not to their kindness, as shown in the examples in (6):

(6) a. *siete stati i migliori sia come livello del del materiale dal punto di vista tecnico sia come prezzi*
"you were the best both in terms of the quality of the of the material from the technical point of view and in terms of price" (N B 5 40 F)
b. *ho vinto il premio come migliore attore del da da parte della critica*
"I won the prize as best actor of the by by critics" (F E 8 4 B)
c. *devo dire Raffaella e' una delle migliori quindi sono contenta*
"I must say Raffaella is one of the best, therefore I am happy" (R A 9 319 B [teacher speaking to Raffaella's parent])
d. *i migliori impiegati e fanno male*
"the best clerks and they do it wrong" (N A 9 14 B)

Besides, *migliore* refers to humans in 4 tokens of the set phrase *migliore amico* "best friend". Another set phrase which appears in the LIP corpus is *migliori auguri* "best wishes" (2 tokens). Most of the nouns modified by *migliore* are very general (*cosa migliore* "best thing" has 4 tokens), and often refer to ways of doing things or of obtaining results, or to the results themselves, as shown in (7):

(7) *condizioni* "conditions", *frutti* "outcome, results", *linea di condotta* "way of behaving", *maniera* "manner", *modo* "way", *risultati* "results" (4 tokens), *sistema* "system", *successo* "success", *vie* "ways"

Other nouns modified by *migliore* refer to establishments in which something is sold, or to offices: *librerie* "bookstores", *cinema* "movie theatres", *sezione* "office of the city council in Naples"; to speech acts or containers of speech acts: *conclusion* "conclusion", *contributo* "contribution", *giornale* "newspaper", *risposta* "reply"; to general circumstances: *avvenire* "future", *casi* "cases", *situazione* "situation", *stato* "state"; and to some other miscellaneous categories. It is noteworthy that only in a couple of cases does *migliore* refer to concrete entities (other than places): once to paper as the best surface for a certain operation, and once to a drug that was claimed to have slimming effects. No tokens of *migliore* modifying nouns referring to food or to humans with respect to their disposition are found in the LIP corpus of spoken Italian.

The fact that *più buono* is mostly reserved for kindness and food taste was first observed by Santilli (2014), who analyzed all the tokens of *più buono* in the written Italian corpus of *la Repubblica 1985–2000*.

Further support for this conclusion comes from examining the collocates of *più buono* and *migliore* in the same corpus.[7] Among the first 10 collocates of *più buono* in *la Repubblica 1985–2000*, shown in Table 3, six refer to foods and beverages or their treatment (*riscaldare* "warm up"), three to human moral qualities (*generoso* "generous" and *gentile* "kind" indicate qualities normally possessed by someone who is *buono*, and *cattivo* "bad, evil" is *buono*'s antonym in this sense), and one, *natale* "Christmas" occurs with *buono* in the set phrase *buon Natale* "merry Christmas". These findings in Italian are reminiscent of the use of *plus mauvais* to refer to the taste of food in French (§ 2.4.1) and *más bueno* in Spanish referring to character or behavior (§ 2.4.3.3).

Table 3. Top ten collocates of *più buono lit*. "more good" in *la Repubblica 1985–2000* corpus

		Frequency	T-score	MI	logDice
1.	*minestra* "soup"	3	1.731	11.233	6.098
2.	*riscaldare* "warm up"	3	1.731	10.206	5.363
3.	*Coca* "coke"	3	1.731	10.182	5.345
4.	*generoso* "generous"	7	2.641	9.264	4.687
5.	*gentile* "kind"	6	2.445	9.258	4.668
6.	*frutta* "fruit"	3	1.729	9.181	4.507
7.	*natale* "Christmas"	14	3.734	8.987	4.461
8.	*dolce* "sweet, dessert"	5	2.230	8.505	3.942
9.	*vino* "wine"	5	2.230	8.426	3.868
10.	*cattivo* "bad, evil"	14	3.729	8.228	3.716

The first ten collocates of *migliore*, shown in Table 4, present a very different picture, similar to the one observed in the LIP corpus.

The data about collocates show that *più buono* has found a niche, and is used mostly with reference to "better taste" of food or "better disposition" of people; the connection between *più buono* and the semantic field of taste had been discovered already by Santilli (2014) in a corpus-based study, but is not mentioned by the Italian grammars I consulted. *Migliore* has a much wider niche; it is true, as some grammars observe, that it is used for abstract concepts, but it is not limited to these, since it can be used when the comparees are concrete nouns, e.g., denoting places; besides, *migliore* appears in some set phrases, such as *migliore amico* "best friend".

7. The collocates and their accompanying statistics were generated through the search tool available at https://corpora.dipintra.it/public/. The data are presented in terms of decreasing logDice (a statistic measure that expresses the typicality of a collocation), and were generated by searching data from a window of 5 tokens on each side of the lemma, setting the following parameters: minimum frequency in corpus = 5, minimum frequency in given range = 3. For information on the statistical measures displayed in Tables 3 and 4 see https://www.sketchengine.eu/.

Table 4. Top ten collocates of *migliore* “better” in *la Repubblica 1985–2000* corpus

		Frequency	T-score	MI	logDice
1.	*modo* “way”	2348	47.878	6.389	8.340
2.	*condizione* “condition”	1411	37.208	6.723	8.320
3.	*risultato* “result”	1361	36.402	6.237	7.998
4.	*soluzione* “solution”	856	28.946	6.552	7.894
5.	*attrice* “actress”	548	23.304	7.794	7.888
6.	*mondo* “world”	1886	42.671	5.842	7.848
7.	*tradizione* “tradition”	604	24.394	7.073	7.804
8.	*ipotesi* “hypothesis”	836	28.546	6.296	7.747
9.	*uno* “one”	1914	42.873	5.642	7.695
10.	*qualità* “quality”	604	24.308	6.517	7.601

4. Concluding remarks

The brief glimpse at corpus data illustrated in § 3 shows that corpus-based studies could be very productive in discovering whether two competing forms have found niches that allow both to be used. The amount of descriptive work to be done in this area of Romance linguistics is very large; existing grammatical descriptions occasionally mention factors that could constitute a niche for one of the competing comparative constructions, but most of the time the statements we find in grammars are based on the authors’ intuition, not on evidence from corpora; it is likely that more factors could be discovered by systematic corpus-based searches; besides, as is well-known (cf. Biber 2012), analyses of corpora might show that some of the co-occurrence restrictions claimed by the grammars are illusory, and lead to more accurate descriptions.

Acknowledgements

Thanks to Chiara Cappellaro, Greville G. Corbett, Pierluigi Cuzzolin, Enzo Santilli and this volume’s editors, who have read a first draft of this chapter and have made many helpful suggestions for improvement. Particular thanks to Janie Rees-Miller, who helped more than should be asked of any editor. It is a pleasure to dedicate this chapter to Mark Aronoff, whose scholarship and friendship have enriched my life.

References

Alcina Franch, Juan & José Manuel Blecua. 1975. *Gramática española*. Barcelona: Ariel.

Anderson, Stephen R. 2016. Romansh (Rumantsch). In Adam Ledgeway & Martin Maiden (eds.), *The Oxford guide to the Romance languages*, 169–184. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199677108.003.0012

Aronoff, Mark. 2016. Competition and the lexicon. In Annibale Elia, Claudio Iacobini & Miriam Voghera (eds.), *Livelli di analisi e fenomeni di interfaccia: Atti del XLVII Congresso internazionale di studi della Società di Linguistica Italiana (SLI)*, 39–51. Rome: Bulzoni.

Aronoff, Mark. 2019. Competitors and alternants in linguistic morphology. In Franz Rainer, Francesco Gardani, Wolfgang U. Dressler & Hans Christian Luschützky (eds.), *Competition in inflection and word-formation*, 39–66. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-030-02550-2_2

Biber, Douglas. 2012. Corpus-based and corpus-driven analyses of language variation and use. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of linguistic analysis* (1st edn.). Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199544004.013.0008

Brăescu, Raluca. 2013. Adjectives and adjectival phrases. In Gabriela Pană Dindelegan (ed.), *The grammar of Romanian*, 410–431. Oxford: Oxford University Press.

Butt, John & Carmen Benjamin. 1988. *A new reference grammar of modern Spanish*. London: Arnold.

Cuzzolin, Pierluigi. 2011. Comparative and superlative. In Philip Baldi & Pierluigi Cuzzolin (eds.), *New perspectives in historical Latin syntax, Volume 4: Complex sentences, grammaticalization, typology*, 495–605. Berlin: de Gruyter. https://doi.org/10.1515/9783110253412.549

De Mauro, Tullio, Federico Mancini, Massimo Vedovelli & Miriam Voghera. 1993. *Lessico di frequenza dell'italiano parlato*. Milan: Etaslibri.

Elton, Charles S. 1927. *Animal ecology*. London: Sidgwick & Jackson.

Grevisse, Maurice. 1986. *Le bon usage. Grammaire française*. Douzième édition refondue par André Goosse. Paris: Duculot.

Grevisse, Maurice & André Goosse. 1980. *Nouvelle grammaire française*. Paris & Gembloux: Duculot.

Herman, Jozsef. 2000. *Vulgar Latin*. Translated by Roger Wright. University Park, PA: Pennsylvania State University Press.

Hertz, Martin (ed.). 1855. *Prisciani institutionum grammaticarum libri I-XII*. Leipzig: Teubner.

Judge, Anne & F. G. Healey. 1995. *A reference grammar of modern French*. Lincolnwood, IL: NTC.

Kattán-Ibarra, Juan & Christopher J. Pountain. 1997. *Modern Spanish grammar. A practical guide*. London & New York: Routledge.

Ledgeway, Adam & Martin Maiden (eds.). 2016. *The Oxford guide to the Romance languages*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199677108.001.0001

Loporcaro, Michele & Tania Paciaroni. 2016. The dialects of central Italy. In Adam Ledgeway & Martin Maiden (eds.), *The Oxford guide to the Romance languages*, 228–245. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199677108.003.0015

Maiden, Martin. 2011. Morphophonological innovation. In Martin Maiden, John Charles Smith & Adam Ledgeway (eds.), *The Cambridge history of the Romance languages: Structures*, vol. I, 216–267. Cambridge: Cambridge University Press.

Maiden, Martin. 2016a. Romanian, Istro-Romanian, Megleno-Romanian and Aromanian. In Adam Ledgeway & Martin Maiden (eds.), *The Oxford guide to the Romance languages*, 91–125. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199677108.003.0008

Maiden, Martin. 2016b. Dalmatian. In Adam Ledgeway & Martin Maiden (eds.), *The Oxford guide to the Romance languages*, 126–138. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199677108.003.0009

Maiden, Martin & Cecilia Robustelli. 2000. *A reference grammar of modern Italian*. London: Arnold.

Maltby, Robert. 2016. Analytic and synthetic forms of the comparative and superlative from early to late Latin. In James Noel Adams & Nigel Vincent (eds.), *Early and late Latin. Continuity or change?*, 340–366. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781316450826.016

Mensching, Guido & Eva-Maria Remberger. 2016. Sardinian. In Adam Ledgeway & Martin Maiden (eds.), *The Oxford guide to the Romance languages*, 270–291. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199677108.003.0017

Migliorini, Bruno. 1952. *Grammatica italiana*. Florence: le Monnier.

OED. December 2020. Niche, n. and adj. *OED Online*. Oxford University Press, https://www-oed-com.univaq.clas.cineca.it/view/Entry/126748?rskey=PlLDCQ&result=1&isAdvanced=false. (29 December, 2020).

Oliviéri, Michèle & Patrick Sauzet. 2016. Southern-Gallo-Romance (Occitan). In Adam Ledgeway & Martin Maiden (eds.), *The Oxford guide to the Romance languages*, 319–349. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199677108.003.0019

Osthoff, Hermann. 1900. *Vom Suppletivwesen der indogermanischen Sprachen: Erweiterte akademische Rede; akademische Rede zur Feier des Geburtsfestes des höchstseligen Grossherzogs Karl Friedrich am 22. November 1899*. Heidelberg: Wolff.

Penny, Ralph. 2002. *History of the Spanish language* (2nd edn.). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511992827

Salvi, Giampaolo. 2011. Morphosyntactic persistence. In Martin Maiden, John Charles Smith & Adam Ledgeway (eds.), *The Cambridge history of the Romance languages: Structures*, Volume 1, 318–381. Cambridge: Cambridge University Press.

Salvi, Giampaolo & Laura Vanelli. 2004. *Nuova grammatica italiana*. Bologna: il Mulino.

Santilli, Enzo. 2014. Italian comparatives: A case of overabundance? L'Aquila, Italy: University of L'Aquila bachelor's thesis. Available at https://www.academia.edu/9307153/Italian_Comparatives_a_Case_of_Overabundance (21 October 2019).

Serianni, Luca. (with Alberto Castelvecchi). 1988. *Grammatica italiana. Italiano comune e lingua letteraria. Suoni forme costrutti*. Turin: Utet.

Thornton, Anna M. 2019. Overabundance: A canonical typology. In Franz Rainer, Francesco Gardani, Wolfgang U. Dressler & Hans Christian Luschützky (eds.), *Competition in inflection and word-formation*, 223–258. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-030-02550-2_9

Wheeler, Max W., Alan Yates & Nicolau Dols. 1999. *Catalan: A comprehensive grammar*. London: Routledge. https://doi.org/10.4324/9780203300275

CHAPTER 12

Multi-layered default in Ripano

Michele Loporcaro and Tania Paciaroni
University of Zurich

This chapter deals with morphological and syntactic defaults in Ripano (Italo-Romance), using the toolkit of Network Morphology and Canonical Typology. Analyzing noun and adjective inflection, we propose a unitary inheritance hierarchy for nominals which features a general default plus overrides specifying class-specific rules of exponence. The hierarchy accounts for gender assignment along with inflectional classes, as Ripano has overwhelmingly overt gender. This also offers the link to the syntax, because in Ripano overt gender is not context-free but depends on syntactic context, a property never described for any other language of the world. We show that one can distinguish a normal vs an exceptional-case default and conclude on how these relate to the morphological default established in the inheritance hierarchy.

Keywords: default, inheritance, overt/covert gender, normal-/exceptional-case default, Italo-Romance, Network Morphology, Canonical Typology

1. Introduction

A prominent strand of research in theoretical morphology has focused on the notion of default as an effective tool in linguistic analysis, and has pointed to the occurrence of different types of defaults within one and the same system (cf., e.g., Corbett & Fraser 2000; Brown & Hippisley 2012; Aronoff 2013; Brown 2016; Gisborne & Hippisley 2017a, 2017b). Following up along this research line, the present chapter addresses the different defaults in the nominal morphology and morphosyntax of the Italo-Romance variety of Ripatransone (central Italy) whose quite extraordinary agreement system (by Indo-European standards) has been subjected to intensive study in both traditional dialectology and theoretical morphology/syntax (cf., e.g.,

https://doi.org/10.1075/cilt.353.12lop

Parrino 1967; Lüdtke 1976; Harder 1988; Ledgeway 2012: 299–310; D'Alessandro 2017, 2020; Paciaroni & Loporcaro 2018).[1]

In § 2, we propose an inheritance hierarchy for Ripano noun inflection which is fascinatingly more complex than that proposed for Russian in the seminal study by Corbett & Fraser (1993: 126). This hierarchy, as usual in this kind of formalization of inflectional systems (cf., e.g., Brown & Hippisley 2012), provides for a morphological default. We then show (§ 3) that Ripano has two distinct syntactic defaults: a normal-case default and an exceptional-case default (Corbett & Fraser 2000, Evans, Brown & Corbett 2002: 119–123, Brown & Hippisley 2012: 86–106). The former must be understood in the light of the dependency of overt gender marking from syntactic context, a typological singularity of Ripano described in detail in Paciaroni & Loporcaro (2018), while the latter specifies the form selected in case of non-agreement (in the absence of an agreement controller) or for agreement with non-canonical controllers. In terms of exponents, the exceptional-case default is expressed with the same inflection as the morphological default, while the normal-case default has a dedicated morphological expression. Section 4 offers some conclusions.

2. Inheritance and default in nominal morphology

We start by setting up a hierarchy accounting for nominal inflection along the lines of Corbett & Fraser (1993: 126; 2000: 60) and Brown & Hippisley (2012). For the inflection of nouns and adjectives the relevant morphosyntactic features are number and gender. Number has two values, singular and plural, while gender has four values: masculine, feminine, (mass) neuter and non-autonomous neuter, as schematized in (1) (where gender/number agreement with the head noun is manifested on the definite article):

1. Our data sources are three-fold: (a) descriptive work by Parrino (1967), Lüdtke (1976), Harder (1988), Mancini (1993); (b) texts by native speakers: Rossi (1999, 2007, 2008), Lambertelli (2003), Cardarelli (2010); (c) our own fieldnotes collected during several trips to Ripatransone, from 2012 on. Unreferenced examples come from fieldwork (partly available online in DAI) and have been checked against the other sources. The data are given in a broad IPA transcription.

(1) The gender system of Ripano

	SINGULAR		PLURAL	
N	l-ə	ˈpa		Ø
	DEF-N.SG	bread(N)		
M	l-u	ˈka	l-i	ˈka
	DEF-M.SG	dog(M)	DEF-M.PL	dog(M)
NAM	l-u	ˈvratːʃa	l-ə	ˈvratːʃa
	DEF-M.SG	arm(NAN)	DEF-F.PL	arm(NAN)
F	l-e	ˈma	l-e	ˈma
	DEF-M.SG	hand(F)	DEF-F.PL	hand(F)
	"the bread/dog/arm/hand"		"the dogs/arms/hands"	

There are three target genders contrasting only in the singular, since the (mass) neuter is number-defective, as usual in central-southern Italy (cf. Loporcaro 2018: 116–154). Besides, there is a controller gender value, the non-autonomous neuter, as highlighted in the boxes:[2] the forms of the definite article selected by the noun "arm" are the same occurring in the singular with masculine nouns and in the plural with feminine. As argued in Loporcaro & Paciaroni (2011), this class of nominal lexemes constitutes a fourth non-autonomous gender value in a sizeable group of dialects spanning central-southern Italy.

2.1 Noun inflectional classes

Moving on to morphological features, we will assume the analysis of Ripano noun inflectional classes (ICs) – defined as "set[s] of lexemes whose members each select the same set of inflectional realizations" (Aronoff 1994: 182) – presented in Paciaroni & Loporcaro (2018: 167–169), which we briefly recapitulate in what follows (with minor modifications). In the schema in (2) nouns are given in their form in isolation (termed 'strong' paradigm;[3] capital letters stand for root forms, A≠B indicating non-phonologically conditioned allomorphy, A=A lack thereof):

2. We follow Corbett (1991: 151) in assuming the target vs controller gender distinction: "controller genders, the genders into which nouns are divided" vs "target genders, the genders which are marked on adjectives, verbs and so on."

3. Note that strong/weak are often used to qualify different inflectional classes hosting complementary sets of lexemes (as is the case, e.g., with strong/weak Germanic verbs), whereas here, just as in the description of German adjectives, the terms are used to qualify two alternative inflections that one and the same lexical item takes in different syntactic contexts (the same terminology is used, e.g., in Corbett 2006: 95f.).

(2) Noun ICs of urban Ripano. Strong paradigm (full inflection)[4]

Noun ICs of urban Ripano.				Strong paradigm (fullinflection)			
IC	ISC	SG	PL	example		gloss	gender
I		A-*e*	A-a[4]	ˈkaːse	ˈkaːsa	"house/-s"	F
II		A-*u*	A-*i*	ˈfijːu	ˈfijːi	"son/-s"	M
III	a.	A-*a*	A-*i*	ˈpaːʈʂa	ˈpaːʈʂi	"father/-s"	M
	b.	A-*a*	B-*i*	ˈmeːsa	ˈmiːʃi	"month/-s"	M
	c.	A-*a*		ˈmɛːla	----	"honey"	N
IV	a.	A-*u*	A-*a*	ˈvratːʃu	ˈvratːʃa	"arm/-s"	NAN
	b.	A-*u*	B-*a*	ˈwoːvu	ˈɔːva	"egg/-s"	NAN
V		A	B	vaʃˈto	vaʃˈtu	"stick/-s"	M
VI		uninflected		ˈka	ˈka	"dog/-s"	M, F, N

There are six ICs which can be arranged into an inheritance hierarchy after that proposed for Russian by Corbett & Fraser (1993: 126) within Network Morphology. Such a hierarchy provides "for a *default* interpretation of lines of inheritance, such that within a family of nodes inheriting from the same source node, a member may be able to *override* the inheritance of a fact." (Brown & Hippisley 2012: 33; emphasis in the original). The inheritance hierarchy for Ripano nominals (i.e., nouns and adjectives, the latter addressed in § 2.3) is displayed in Figure 1 (where Roman numbers in the node labels correspond to ICs in (2)).[5]

The first branch ending with a terminal node (N_VI) accounts for indeclinables: the main structural criterion for inclusion in this class is that the stem ends in a stressed vowel which does not show any singular/plural alternation (more examples

4. All -*a*'s in the inflections considered in (2) and the following paradigms (as well as the hierarchies in Figures 1–2) are phonemic /ə/'s, which may be realized phonetically as either [ə] or [a], depending on prosodic context. Elderly speakers of urban Ripano tend to realize [ə] utterance-internally vs -[a] prepausally, and hence in citation forms: thus, local lexicographers and the descriptive literature on Ripano use this notation. By contrast, younger urban and rural speakers tend to generalize -[ə]. This -/ə/ contrasts today with final unstressed -/a/ (re)introduced recently, probably on the model of Standard Italian (cf. Paciaroni & Loporcaro 2018: 150), which is never realized as -[ə] and occurs in the feminine singular of determiners, quantifiers and place names, but generally not in nouns and adjectives. Disregarding this detail, we use -*a* for simplicity throughout.

5. The top node is labelled MOR_NOMINAL rather than MOR_NOUN, since no separate node for adjectives, and hence, no branching between MOR_NOUN and MOR_ADJ needs to be part of the lexemic hierarchy for Ripano, for reasons explained in § 2.3.

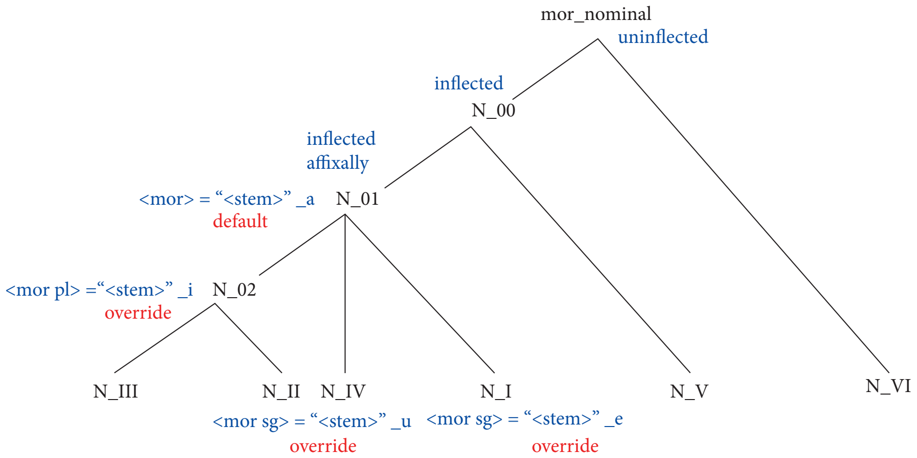


Figure 1. Inheritance hierarchy for Ripano nominals (strong paradigm)

in (3)).[6] All remaining ICs inflect for number (apart from number-defective nouns, obviously, such as the mass neuters in ISC IIIc). The left branch first leads to a node that is labelled N_00 (inflected nouns). Here, the right branch leads to a terminal node N_V, an IC characterized by lacking affixal inflection and by the phonological property of final stress (cf. Rossi 2008: 18f.) – a property shared with class VI – but nevertheless displaying two contrasting forms for the singular vs plural cells, due to the occurrence of a stressed vowel alternation originally caused by metaphony and then rendered opaque as final vowels following /n/ or /r/ were deleted with the entire final syllable: e.g., *lu li'mo/li li'mu* "lemon(M),-s", *lu məra'to/li məra'tu* "bricklayer(M),-s" (compare Italian *limone*, *muratore*). The left branch N_01, on the other hand, dominates ICs displaying affixal inflection, and contains information on the singular and plural default exponent, which is *-a* for both cells. The node N_01 dominates several other nodes: both N_I and N_IV inherit the default exponent *-a* in the plural, but the singular default form is overridden, the overrides being *-e* in N_I and *-u* in N_IV.[7] The node N_02, then, is defined by the override of the plural default value *-a*. This superclass subdivides into N_II and N_III, which inherit the plural exponent *-i* but remain distinct for the exponent of the singular, which is *-u* in N_II (override) and *-a* by default in N_III.

6. Class VI also includes a few nouns (mainly loans) which end in an unstressed vowel (e.g., *'bːɔjːa* "executioner/-s", *bːi'kiːni*/-s).

7. Overrides "captur[e] inflectional classes, syncretism and deponency" (Brown & Hippisley 2012: 36).

The same hierarchy effectively models gender assignment, as Ripano has (almost) generalized overt gender, the indeclinable class VI being the only one which hosts nouns of all three target genders:

(3) a. M *lu/li fəˈrːa* "blacksmith", *kulaˈtu* "colander, strainer", *kuntəˈði* "farmer";
b. F *le/la bːəˈʃe* "lie", *ˈma* "hand", *pənˈtsjo* "rent";
c. N *lə ˈfje* "hay", *ˈgra* "wheat", *saˈpo* "soap", *ˈvi* "wine".

All remaining ICs (see the right-hand column in (2)), host nouns of just one gender, with the sole exception of IC III, which hosts both masculine (IIIa-b) and (mass) neuter nouns (IIIc). Since however mass neuters lack a plural form, this does not cause any deflection from overt gender: a noun whose paradigm reduces to an *a*-ending singular is neuter, while nouns with *a*-singular and *i*-plural are masculine.

The pervasiveness of overt gender makes the Ripano system more canonical than that of Italian – often quoted in the literature in Canonical Typology as a good approximation to a fully canonical system, especially as far as gender agreement in the noun phrase is concerned (cf., e.g., Corbett 2006: 9; Fedden & Corbett 2017: 2; *contra* Thornton 2019) – since Ripano comes nearer "an ideal overt system" (Corbett 1991: 62), which "would have a marker for gender on every noun, with only one marker per gender" thus complying with *Criterion* 2 for canonical agreement:

(4) Canonical agreement typology *Criterion* 2 (Corbett 2006: 11):

controller has overt expression of agreement features > controller has covert expression of agreement features

On the other hand, canonical inflection (Corbett 2009, 2012: 197–199) would require for each lexeme to have a unique stem form across cells, and for each paradigm cell to host just one exponent. Thus, canonical overt gender, with a cumulative gender/number exponent as found throughout Romance (in itself a deviation from canonicity), in a three target-gender system with two number values such as Ripano would appear as follows:

(5) Canonical overt gender (hypothetical) with cumulative gender/number exponence

	SG	PL
N	A-*u*	
M	A-*o*	A-*i*
F	A-a	A-*e*

Ripano, however, departs from canonicity since the morphology still makes more distinctions than required by the syntax, contrasting inflectional classes, while it fails "to make … morphosyntactically relevant distinction[s] … under particular (morphological) conditions" (Baerman, Brown & Corbett 2005: 2), i.e., shows syncretisms (see (2) and the schema in (6)):

(6) Strong paradigm: syncretism pattern A

	I	IV	II	III
SG	*-e*	*-u*	*-u*	*-a*
PL	-a	-a	*-i*	*-i*

The inflections seen in (2) – and modelled in Figure 1 – are not constant across syntactic contexts, because M nouns (ICs II–III), as well as NAN nouns, as far as their singular forms are concerned (IC IV), display inflections different from those in (2) in a complementary subset of syntactic contexts (§ 3). This results in the weak paradigm in (7):

(7) Noun ICs of urban Ripano. Weak paradigm (reduced inflection)

Noun ICs of urban Ripano.				Weak paradigm (reduced inflection)			
IC	ISC	SG	PL	example		gloss	gender
I		A-*e*	A-a	ˈkaːse	ˈkaːsa	"house/-s"	F
II-IV	a.	A-a	A-a	ˈfijːa	ˈfijːa	"son/-s"	M
		A-a	A-a	ˈpaːʈʂa	ˈpaːʈʂa	"father/-s"	M
		A-a	B-a	ˈvratːʃa	ˈvratːʃa	"arm/-s"	NAN
	b.	A-a	A-a	ˈmeːsa	ˈmiːʃa	"month/-s"	M
		A-a	B-a	ˈwoːva	ˈɔːva	"egg/-s"	NAN
	c.	A-a		ˈmɛːla	----	"honey"	N
V		A	B	vaʃˈto	vaʃˈtu	"stick/-s"	M
VI		uninflected		ˈka	ˈka	"dog/-s"	M, F, N (= [ˈpa] "bread")

Syncretism becomes more pervasive in (7), as schematized in (8) (to be contrasted with (6)):

(8) Weak paradigm: syncretism pattern B

	I	III	II	IV
SG	*-e*	-a	-a	-a
PL	-a	-a	-a	-a

The sets of cells distinguished in the full paradigm are collapsed. First, nouns of ICs II (*'fijːu/'fijːi* "son/-s"), III (*'paːʈʈa/'paːʈʈi* "father/-s") and IV (*'vratːʃu/'vratːʃa* "arm/-s") merge into one and the same class II–IV, which has just one form for both singular and plural. Second, this reduced inflectional form has the same -*a*, which was already shared by many inflectional classes in the full paradigm. We can consider this as a default and say that the (gender-)number morphological realization is almost neutralized. The inflectional system can be displayed, for the weak paradigm too, by means of the modified inheritance hierarchy in Figure 2, which accounts for the loss of contrasts with respect to the one in Figure 1.

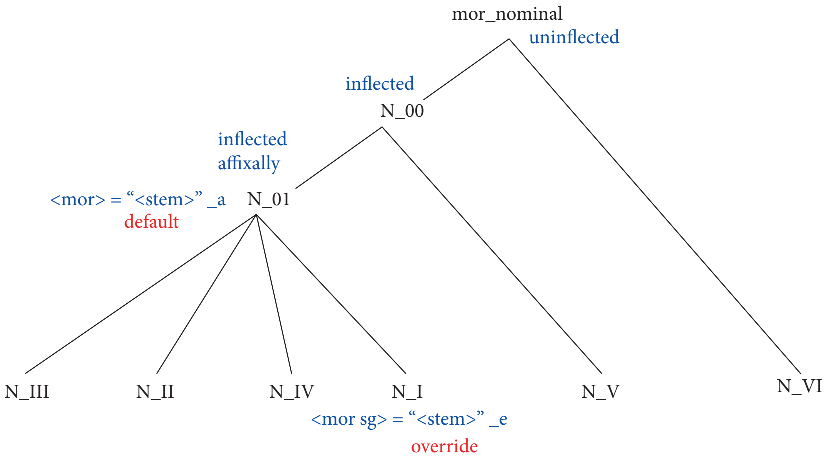


Figure 2. Inheritance hierarchy for Ripano nominals (weak paradigm)

The first two branchings, leading to N_VI and N_V, are unchanged, and so is the node N_01, specifying the default (-*a*) for both singular and plural forms of all affixally inflected nouns. Within this superclass, there is just one override left with respect to the more articulated hierarchy for strong inflection in Figure 1: this specifies, at the node N_I, that class I (feminine) nouns have the exponent -*e* for the singular. All the rest is annexed to the province of the default exponent -*a*.

2.2 The rise of overt gender in Ripano

In Italian, the major IC responsible for inflection not being a reliable gender predictor is the *cane/-i* "dog/-s" class (IC III in D'Achille & Thornton 2003: 213). This contains nouns stemming from Latin 3rd declension, which subdivide about fifty-fifty into masculines (e.g., *padre/-i* "father/-s") and feminines (e.g., *volpe/-i* "fox/-es"). In Ripano, the heir of Latin 3rd declension is IC III (in (2)): this has retained (mass) neuters (IIIc), which merged into the masculine in Italian, but lost all feminines, since these were systematically attracted into IC I (*ˈkaːse/-a* "house/-s"), the heir of Latin 1st declension: *ˈ(v)orbe/-a* "fox/-es", *ˈnɔtːe/-a* "night/-s", etc. As for Latin 1st and 2nd declensions, which were predominantly associated with feminine and masculine respectively, their diachronic successors in Ripano (IC I and II respectively) got rid of all exceptions.[8] One such exception in standard Italian is *mano/-i* "hand/-s", the outcome of Latin *manus/-us*, a 4th declension noun which merged into the productive successor of Latin 2nd declension and represents the only feminine noun in that IC. But in present-day Ripano, Latin *manus* became invariable *ˈma* due to regular sound change or was reassigned analogically to IC I (*mane/mana* in Rossi 1999: 46, 77), thus leaving exclusively masculines in IC II. On the other hand, analogical change redeployed all originally 1st declension masculines, such as those formed with the suffix *-ˈiʃta*, not into a newly created dedicated class as in Italian (*socialista/-i* "socialist/-s") but rather into IC II: e.g., *kumːuˈniʃtu/-i* "communist/-s" (Rossi 2008: 86, 168) (compare weak inflection in *lu/li kumːuˈniʃta* "the communist/-s", *nu spetʃaˈliʃta* "a specialist", Cardarelli 2010: 49).

This series of changes converged to yield the virtually perfect IC-to-gender predictability observed in (2). All these changes, in turn, grafted onto a general reshuffling of inflectional endings – preceded by a stage in which all affixal endings except /a/ tended to merge into /ə/ – which favored the rise of the unusual agreement system addressed in § 3.

2.3 Adjective inflectional classes

While "[w]e typically conceive of inflectional classes as discrete entities associated with one particular part of speech" (Brown & Hippisley 2012: 113), there are languages such as Latin "where it is difficult to identify specifically adjectival inflection, so that adjectival inflection is essentially taken from noun declensions" (Brown & Hippisley 2012: 124). This holds also for Ripano. Therefore, contrary to, say, Russian where the node MOR_NOMINAL dominates two separate nodes MOR_NOUN vs MOR_ADJ, the hierarchies provided exhaust the nominal portion of the lexemic hierarchy: simply, "the node ADJ can inherit via network relations from the inflectional class nodes defining noun morphology and add information about gender accordingly" (Brown & Hippisley 2012: 124). Hence, Ripano adjectives also (except the invariable ones, e.g., *ˈbːlu* "blue") show two sets of inflections, selected in complementary syntactic contexts (see § 3), with one set making more distinctions than the other.

8. Originally 2nd declension neuters were attracted into IC III, which has retained Latin (mass) neuters: *ˈɔːra* "gold" (< Latin AURUM).

Among affixally inflected adjectives, there are two minor classes containing very few items each, mostly conversions from nouns: one, exemplified with the adjective "shabby (of people)" (*ʃa'tːo* M.SG, *ʃa'tːu* M.PL, *ʃa'tːoːne* F.SG, *ʃa'tːoːna* F.PL) combines N_V (masculine) and N_I (feminine);[9] the other is exemplified with the adjective *marki'dːʒaːna* N (N_III), *marki'dːʒa* M.SG/M.PL (N_VI), *marki'dːʒaːne* F.SG, *marki'dːʒaːna* F.PL (N_I) "in/from the Marches".

The major inflectional class combines N_III (neuter), N_II (masculine) and N_I (feminine). It subdivides into two subclasses depending on the presence/absence of stem alternation (cf. Harder 1988: 132–133). (9) and (10) illustrate the two paradigms of the subclass without stem alternation:

(9) Adjective *'bːrutːu* "ugly"

a. Strong paradigm (full inflection)

GENDER	NUMBER	
	SINGULAR	PLURAL
N	'bːrutːa	–
M	'bːrutːu	'bːrutːi
F	'bːrutːe	'bːrutːa

b. Weak paradigm (reduced inflection)

<table>
<tr><th>GENDER</th><th colspan="2">NUMBER</th></tr>
<tr><th></th><th>SINGULAR</th><th>PLURAL</th></tr>
<tr><td>N</td><td rowspan="2">'bːrutːa</td><td></td></tr>
<tr><td>M</td><td rowspan="2">'bːrutːa</td></tr>
<tr><td>F</td><td>'bːrutːe</td></tr>
</table>

(10) Affixal marking of gender and number

a. Strong paradigm

	N	M	F
SG	A + *a*	A + *u*	A + *e*
PL	–	A + *i*	A + *a*

b. Weak paradigm

	N	M	F
SG	A + a	A + a	A + *e*
PL	–	A + a	A + a

9. The neuter form in -*'oːna* is not used for semantic reasons.

(11) and (12) illustrate the complementary subclass, exemplified with the adjective *ˈfwortu* “strong”:

(11) Adjective *ˈfwortu* “strong”

a. Strong paradigm (full inflection)

GENDER	NUMBER	
	SINGULAR	PLURAL
N	ˈfɔʀta	–
M	ˈfwortu	ˈfworti
F	ˈfɔrte	ˈfɔrta

b. Weak paradigm (reduced inflection)

GENDER	NUMBER	
	SINGULAR	PLURAL
N	ˈfɔʀta	–
M	ˈfworta	ˈfworta
F	ˈfɔrte	ˈfɔrta

(12) Affixal and stem-internal marking of gender and number

a. Strong paradigm

	N	M	F
SG	A + a	B + *u*	A + *e*
PL	–	B + *i*	A + a

b. Weak paradigm

	N	M	F
SG	A + a	B + a	A + *e*
PL	–	B + a	A + a

Here stem alternation contributes to the contrast between different cells. The distribution of stem allomorphy is defined in terms of gender. In Network Morphology, this is expressed through ‘gender’ being inserted as a last path extension:

(13) Brown & Hippisley (2012: 61):
“*Number and gender (a constraint on the full morphological model)*
In paths containing gender and number attributes, gender attributes will extend number attributes. (e.g., <mor sg masc>)”

Membership in inflectional subclasses is a lexical property, whereas the selection of the paradigm type (either strong or weak) depends on purely syntactic factors (§ 3).

The Italian counterparts of the adjectives in (9) vs (11), viz. *brutto* vs *forte*, belong to two different ICs (just like their Latin ancestors), which, contrary to Ripano, differ because the former, but not the latter, marks gender contrasts.[10] In Ripano, class-membership has been totally reshuffled, as all adjectives, whatever Latin class they stem from, mark gender affixally, as seen in (14):

(14) Ripano adjectives	from Latin Class 1	from Latin Class 2
+ gender marking via stem alternation (11)	*ˈbːwoːnu* “good”, *ˈfrisku* “cool”	*ˈfwortu* “strong”
– gender marking via stem alternation (9)	*ˈveːru* “true”, *ˈjuʃtu* “right”	*ˈfaːtʃəlu* “easy”, *ˈtriʃtu* “sad”

Finally, the syntax of Ripano offers spectacular confirmation of the conflated MOR_NOMINAL node for nouns and adjectives. Indeed nouns not only select either the strong or weak paradigm depending on syntactic context but, in addition, undergo contextual agreement with the clause subject, in certain syntactic constructions, whatever their inherent gender. The constructions involved are exemplified in (15) with two agreeing nouns (one feminine and one masculine):[11]

(15) a. *n ˈɔːm-a tʃ ˈa ˈseːt-u / ˈtjemb-u*
INDF man(M)-NONF.SG has thirst(F)-M.SG / time(M)-M.SG
“a man is thirsty/has time”

b. *n-e ˈfemːən-e tʃ ˈa ˈseːt-e / ˈtjemb-e*
INDF-F.SG woman(F)-SG has thirst(F)-F.SG / time(M)-F.SG
“a woman is thirsty/has time”

These nouns take the same five-cell paradigm as the adjectives in (9), whenever required by subject agreement:

10. Cf. Thornton (2019) and Bond (2019) for discussion within Canonical Typology of parallel data from Italian and Spanish, addressing the issue whether lack of gender agreement in Class 2 adjectives – which derives from reduced gender agreement in the Latin forerunner IC: *facilis* “easy:NOM.M.SG = :NOM.F.SG” vs *facile* “easy:NOM.N.SG”, as opposed to Class 1 *bonus/-a/-um* “good:NOM.M.SG ≠ :NOM.F.SG ≠ :NOM.N.SG” – is best analyzed in terms of syncretism or featural inconsistency.

11. The DAI data (cf. Paciaroni 2020: § 4.4.1.4) show that noun agreement, exemplified in (15) and described in the literature (cf. Parrino 1967: 161f., 166; Harder 1988: 245; Paciaroni & Loporcaro 2018: 162f.), is not categorical: alternatively, the noun can stay in its gender/number-unmarked form.

(16) Agreement on nouns in Ripano

a. *ˈseːte* "thirst(F)"

GENDER	NUMBER	
	SINGULAR	PLURAL
N	ˈseːta	–
M	ˈseːtu	ˈseːti
F	ˈseːte	ˈseːta

b. *ˈtjembu* "time(M)"

GENDER	NUMBER	
	SINGULAR	PLURAL
N	ˈtjemba	–
M	ˈtjembu	ˈtjembi
F	ˈtjembe	ˈtjemba

Contrary to adjectives, there is no counterpart to (11) for agreeing nouns, since the stem allomorph is determined lexically.[12]

3. Two syntactic defaults in Ripano: Preliminaries on Ripano syntax

In addition to the morphological default established in the inheritance hierarchies in Figures 1 and 2, Ripano has two distinct syntactic defaults: a normal-case and an exceptional-case default (Corbett & Fraser 2000; Evans et al. 2002: 119–123; Brown & Hippisley 2012: 86–106). In order to introduce these, we must provide some background on the syntactic conditions on which selection of the two morphological paradigms (5) vs (7) (for nouns) and (9a)/(11a) vs (9b)/(11b) (for adjectives) depends (Paciaroni & Loporcaro 2018 provide the full picture). The difference concerns only masculine agreement on adjectives and masculine and NAN nouns

12. This difference has a diachronic explanation. In *ˈbːwoːnu* "good:M.SG" vs *ˈbːɔːne* "good:F.SG" stem allomorphy arose from metaphony, which was triggered by final high vowels and hence applied in the Proto-Romance masculine **ˈbɔːnu*, not in feminine **ˈbɔːna*. Conversely, the Proto-Romance forerunners of the nouns of the relevant classes uniformly contained either high (**ˈtɛmpu/*ˈtɛmpi* "time/-s") or non-high vowels (**ˈseːte*), which determined the application vs non-application of metaphony. The phonological process was already completed by the time contextual agreement was extended to nouns. The latter were not reshaped analogically either, which happened, on the contrary, to adjectives that were subjected secondarily to non-phonologically determined extension of metaphony, as seen in (11): in fact, the Proto-Romance ancestor of *ˈfwortu* "strong:M.SG", i.e., **ˈfɔrte*, ending in a non-high final vowel, did not meet the structural description of metaphonic diphthongization.

belonging to classes II–IV. Consider the two sets of forms for masculine nouns, exemplified in context:

(17) a. *ˈisː-u ˈɛ ˈfːijː-u/*-a də ˈbːrave ˈdːʒɛnde*
3M-M.SG be.PRS.3 son(M)-M.SG/-NONF.SG of good people
"he is the son of honest people" (Rossi 1999: 173)

b. *ˈa ˈfatːa/-ə ˈlamb-i/*-a e*
has do:PTP.N.SG lightning(M)-M.PL/-NONF.SG and
ˈtːwoːn-i/-a*
thunder(M)-M.PL/-NONF.SG
"there was thunder and lightning"

(18) a. *l-u ˈfijː-a/*-u ˈbːjelː-u/*-a/*-ə*
DEF-M.SG son(M)-NONF.SG/-M.SG beautiful\M-M.SG/-NONF.SG
"the beautiful son"[13]

b. *l-i ˈlamb-a/*-i e l-i*
DEF-M.PL lightning(M)-NONF.SG/-M.PL and DEF-M.PL
ˈtwoːn-a/-i*
thunder(M)-NONF.SG/-M.PL
"there was thunder and lightning"

In addition to occurring in isolation, strong inflections are selected in a series of syntactic contexts, partly exemplified in (17), including occurrence as bare noun (predicative or argumental),[14] and occurrence after modifiers not agreeing in gender/number, the invariable complementizer *kə* (e.g., *kə ˈtːjemb-i/*-a* "what times(M)-M.PL/-NONF.SG!"), the invariable quantifier *ˈkakːa* "some" (e.g., *ˈkakː ˈwomːən-i/*-a* "some men(M)-M.PL/-NONF.SG"), periphrastic quantifiers such as *ne ˈfreːke də* "a lot of" (e.g., *ne ˈfreːke də ˈtwoːn-i/*-a* "a lot of thunders-M.PL/-NONF.SG") and adnominal numerals (e.g., *do ˈpasː-i/*-a* "two steps(M)-M.PL/-NONF.SG").[15] By contrast, (18a)–(b) exemplify the syntactic contexts in which overt gender marking is neutralized via extension of the default ending *-a*. This happens whenever the noun is preceded by determiners or modifiers marked by gender/number: the list includes articles (definite and indefinite: e.g., *n-u ˈlamb-a/*-u* "a-M.SG lightning(M)-NONF.

---

13. In Paciaroni & Loporcaro (2018: 154), (18a) reads erroneously *lu ˈfijː-u ˈbːjelː-u*, which is ungrammatical (because of *ˈfijː-u*) for all Ripano informants. We take this opportunity to flag this unfortunate error.

14. A special case is that of the strong paradigm of agreeing nouns ((15)–(16)), whose gender/number values are selected contextually via subject agreement.

15. The numeral *one* is homophonous with the indefinite article and, like the latter, agrees in gender and number: consequently, it triggers weak inflection on the noun, contrary to all other adnominal numerals.

sg/-m.sg"), demonstratives (e.g., *ʃt-u 'lamb-a/*-u* "this-m.sg lightning(m)-nonf.sg/-m.sg"), gender-variable quantifiers (e.g., *'tand-u 'tjemb-a/*-u* "so much-m.sg time(m)-nonf.sg/-m.sg"), and adjectives (e.g., *'bːjelː-u ka'valː-a/*-u* "beautiful-m.sg horse(m)-nonf.sg/-m.sg").[16] Note that the condition that the modifier be gender/number-(in)variable refers to the paradigm as a whole, not to the individual form selected: thus, *kə 'ɔːm-u/*-a 'bːrutː-u* "what an ugly-m.sg/-nonf.sg man(m)-m.sg!" contrasts with *l/n 'ɔːm-a/*-u* "the/a man(m)-nonf.sg/-m.sg" because *kə* is invariable while the articles are gender/number-variable as such, though the prevocalic singular form never shows gender agreement overtly (compare feminine *l 'uːv-e* "the grapes(f)").

As highlighted in Paciaroni & Loporcaro (2018: 175) and Loporcaro (2018: 312), the context-dependency of overt gender is a unique peculiarity of Ripano. Compared with, e.g., the different adjective inflectional paradigms of German, one must remark that context-dependency here concerns also agreement controllers, not only agreement targets.[17] The latter too show context-sensitive selection of the weak vs strong paradigms seen in (9)–(12) above. In (18a), in fact, the postnominal adjective *'bːjelː-u* takes strong inflection, and so does the prenominal adjective, if not preceded by a determiner marking gender and number:

(19) a. *'juːtə=ma* *'bːjelː-u/*-a/*-ə* *frə'ki*
help.imp.2sg=1sg.io beautiful\m-m.sg/-nonf.sg child(m)
"help me, beautiful child!"

b. *n-u / l-u* *'bːrutː-ə/*-u* *'ka/'disk-a*
indf-m.sg / def-m.sg bad-nonf.sg/-m.sg dog(m)/disk(m)-nonf.sg
"a/the bad dog/disk"

We now move on to illustrate the two syntactic defaults of Ripano.

---

16. Masculine plural determiners are exceptional in that, unlike determiners from all other paradigm cells, they invariably select strong inflection on the noun (see Paciaroni & Loporcaro 2018: 172–174):

(i) *jː* *'anː-i/*-a//'wort-i/*-a*
def.m.pl year(m)-m.pl/-nonf.sg//garden(m)-m.pl/-nonf.sg
"the years/gardens"

17. Also, contrary to German adjective inflection, selection of the weak paradigm is sensitive to the mere presence of a prenominal determiner, not to its (in)definiteness: thus, on the one hand, indefinite/definite articles all induce weak inflection (e.g., *'twoːn-u* "thunder(m)-m.sg" vs *n-u/l-u 'twoːn-a/*-u* "a-m.sg/the-m.sg thunder(m)-nonf.sg"), on the other hand, definite modifiers such as possessives, being always postnominal, do not induce weak inflection unless the article co-occurs: e.g., *'nɔnː-u 'mje* "my grandfather(m)-m.sg" vs *l-u 'nɔnː-a* "the-m.sg grandfather(m)-nonf.sg", *l-u 'fijː-a 'swo* "[*lit*. the-m.sg] his son(m)-nonf.sg" (Harder 1988: 137, Rossi 2008: 175).

3.1 Normal-case default

The *Criterion* 2 of the canonical typology of agreement, mentioned in (4), would prescribe that all nouns have overt marking of both number and gender. In Ripano we find such an overt marking only if a series of phonological, morphological, morphosyntactic and syntactic conditions are met, whereby purely morphological IC-based conditions and morphosyntactic ones go hand in hand, given the widespread occurrence of overt gender discussed in §§ 2.1–2.2. Firstly, one has to put aside uninflected nouns in class VI, which may belong to any gender: most of these can be singled out because of their phonological shape, since they are the only nouns in the language which end in a stressed vowel and do not show number alternation. Moving on to inflected nouns, all feminines (i.e., all nouns belonging to IC I) and masculines ending in SG /ˈo/ and PL /ˈu/ (belonging to IC V) satisfy Criterion 2 (in (4)) in a context-independent way: this follows from the identity of their inflected forms in both the strong and the weak paradigms (2) and (7). On the other hand, masculine nouns – as well as NAN nouns, due to their singular forms – do not show gender overtly in a context-independent way. Rather, as shown in (17)–(18), their gender (and, partly, number, depending on IC) is manifested overtly only in the strong paradigm. We conclude that the strong paradigm corresponds to the default, in a syntactic sense: more precisely, this is the normal-case default, since it is selected when the noun is used by itself, while selection of the weak paradigm is context-sensitive, subject to syntactic conditions (presence of a prenominal determiner/modifier marked for gender/number). Exemplifying with a class II masculine, the two defaults addressed so far look as follows:

(20)	The two defaults for Ripano nouns	SINGULAR	PLURAL	gloss
a.	morphological default	ˈfijːa	ˈfijːa	“son/-s”
b.	(syntactic) normal-case default	ˈfijːu	ˈfijːi	

Not unlike the default hierarchies in Figures 1–2, the normal-case default can be described in the same terms for adjectives too. In fact, adjectives, too, show strong inflection by (syntactic) default, i.e., when used in isolation and whenever not preceded by a gender/number-marked determiner while the latter context – exemplified in (19b) – is the only one in which the weak paradigm is selected. Again, this contrast – schematized in (21a–b) – only concerns the masculine. However, adjectives – and other agreement targets as well – differ from nouns in that they also show the exceptional-case default (in (21c)), to be addressed in § 3.2:

(21) Three defaults for Ripano adjectives
a. morphological default
b. (syntactic) normal-case default
c. (syntactic) exceptional-case default

SINGULAR	PLURAL	gloss
ˈbːjelːa	ˈbːjelːa/ˈbːjejːa[18]	"beautiful"
ˈbːjelːu	ˈbːjejːi/ˈbːjelːi	
ˈbːɛlː-ə/-a	–	

3.2 Exceptional-case default

Exceptional-case default is relevant only for agreement targets. While the reader is referred to Corbett & Fraser (2000), Evans et al. (2002: 119), Brown & Hippisley (2012: 87) for the definition and formalization, suffice it to mention the kind of contexts in which such a syntactic default has been argued to occur: "The circumstances are, roughly speaking, all those where agreement is not controlled by a prototypical noun phrase (one headed by a noun or pronoun)" (Corbett & Fraser

---

18. The DAI data show a tendency to reduce stem allomorphy in the paradigm. Thus, in the M.PL there is variation between the expected stem *ˈbːjejː-* (with the palatalizing effect of *-i* on /l/) which is described in the literature (Harder 1988: 133) and attested in folk literature (cf., e.g., *'sti grisendèma è bbjéji prassà* "these chrysanthemums are very beautiful", Rossi 1999: 80; *li bbjeja jestra suó* "his beautiful gestures", Rossi 2007: 19), and the stem *ˈbːjelː-* without palatalization due to analogical extension of the M.SG stem:

(i) a. *kə ˈwocː-i ˈbːjelː-i*
what eye(M)-M.PL beautiful\M -M.PL
"what beautiful eyes!"

b. *l-i ˈbːjelː-ə*
DEF-M.PL beautiful\M-NONF.SG
"the beautiful …"

The youngest recorded speaker (born in 1996) in addition to the loss of palatalization (iia), also showed the elimination of the metaphonic diphthong [je] (iib–c) distinguishing the masculine from the feminine and (mass) neuter stem:

(ii) a. *l-i ˈbːjelː-ə kaˈvalː-ə*
DEF-M.PL beautiful\M-NONF.SG horse(M)-NONF.SG
"the beautiful horses"

b. *l-u ˈbːɛlː-ə kaˈvalː-ə*
DEF-M.SG beautiful-NONF.SG horse(M)-NONF.SG
"the beautiful horse"

c. *l-i ˈbːelː/ˈbːejː ˈoːmən-i*
DEF-M.PL beautiful man(M)-PL
"the beautiful men"

2000: 71).[19] These include agreement with non-nominal controllers such as clauses, or non-agreement (e.g., in an impersonal clause). In addition, "[e]ven if the agreement controller is a noun phrase headed by a noun or pronoun, there may still be problems involving gender agreement, caused by reference difficulties" (Corbett & Fraser 2000: 82). The two cases are exemplified with Russian in (22):

(22) a. *byl-o xolodn-o* Russian
was-N.SG cold-N.SG
"it was cold"
b. *kto èt-o zdela-l*
who this-N.SG did-M.SG
"who did this?"

Lacking any eligible agreement controller, as in impersonals ((22a)), Russian agreement targets take neuter endings by default, while masculine is selected in (22b), where the interrogative pronoun has unspecified human reference: the appropriate gender is unknown, as the person involved may be either male or female, but the referent – given the properties of *kto* – is known to be human, which rules out neuter agreement. Compare the Ripano counterparts:

(23) a. *ʃt iˈmːɛrn-a nn a pjəˈvuːt-a pə ˈɲːɛnda*
this winter(M)-nonF.SG NEG has rained-N.SG at all
"this (last) winter it did not rain at all"
b. *sə dərˈm-jeːv-a ˈbːɛ su ˈlːu ˈljetːa*
UHS sleep-IMPF-3N.SG well on that bed
"did one sleep well on the bed?" (Harder 1988: 197; our glosses, ML&TP)

(24) a. *ki ˈmaɲː-a / ˈɛ rːiˈvaːt-a / ˈɛ ˈpːjand-a*
who eat.PRS-3N.SG / is come-N.SG / is cried-N.SG
"who is eating/has arrived/has cried?" (Harder 1988: 160, 237–239)
b. *ki s=a ˈmisː-a a ˈrːid-a*[20]
who REFL=has put\M-N.SG to laugh-INF.N.SG
"who started laughing?"

Both the finite verb, which consistently shows gender, in addition to person/number, agreement, and the past participle in perfective periphrastics in (23)–(24) (see

19. In a different terminological framework, for Aronoff (2013: 92) verb agreement when the subject is unspecified or otherwise agreement with a controller lacking relevant specifications falls under the label "orphan default".

20. Note in passing that root-stressed infinitives can show subject agreement in number and gender. Since *ˈrid-a* in (24b) is the neuter form, it is impossible to decide whether there is agreement with the subject or the default form. That even infinitives (albeit only from one IC) can agree in gender is, again, an unusual property of Ripano (cf. Paciaroni & Loporcaro 2018: 158).

also the impersonal clause in (17b)) appear in the neuter, taking the same inflection -*a* as the morphological default (§ 2). The same happens to adjectives (see (21c)) whenever the agreement controller does not carry any gender specification:

(25) *ˈkomːa ˈɛ ˈbːɛlː-a ˈji a l-u ˈmaːr-a*
how be.PRS.3 nice-N.SG go.INF to DEF-M.SG sea(M)-NONF.SG
"how nice it is going to the sea!"

Use of the neuter in the unspecified human subject clause in (23b) can be contrasted with (26), where the verb agrees with the pronominal subject (either overt or silent: like Italian, Ripano is a pro-drop language):

(26) *(ˈia) nən ˈdɔrm-u ˈmaŋg-u l-a ˈnɔtː-u*
1SG NEG sleep.PRS-[1]M.SG not_even-M.SG DEF-F.SG night(F)-M.SG
"I don't sleep even at night" [Male referent] (DAI data; speaker AnIa)[21]

The fact that (exceptional) default agreement with non-canonical controllers takes this particular shape is due to diachronic persistence, as the ending -*a*/-*ə* is the exponent of the neuter, viz. the same gender value which occurred for exceptional-case default in Latin (cf., e.g., Loporcaro 2018: 22–26 on Latin and 233–235 on the diachronic continuity between form and function of the Latin and Romance neuters). While use of the neuter in impersonals such as (17b), (23a) and (25) seems to be stable in present-day Ripano, its occurrence with the interrogative pronoun *ˈki* "who", described as regular by Harder (1988: 160, 237–239) and Mancini (1993: 119), is vacillating: in our informants' answers alongside the conservative option (the neuter) we have recorded also use of the masculine, most probably due to interference from standard Italian which uses masculine in this context. This is exemplified in (27a–b), both stemming from one and the same informant (FiPi, born in 1953):

(27) a. *ki s=ɛ ˈmːɔrt-ə*
who REFL=is died\N-N.SG
"who has died?"
b. *ki saˈrː-a ˈmwort-u*
who be.FUT-3 died\M-M.SG
"who will have died?"

---

21. In *la ˈnɔtːu*, which has the form of a the noun but means "at night", noun phrase takes on the -*u* ending via subject agreement. Outside this construction, the noun is a class I feminine: *l-a ˈnɔtː-e*. Just like agreement in locative adjuncts (cf. Paciaroni & Loporcaro 2018: 165; Paciaroni 2020: §§ 4.4.1.3–4; Loporcaro, to appear), agreement in temporal adjuncts is also vanishing and exceedingly rare in today's Ripano. In our fieldwork, we were able to observe subject agreement on nouns in both contexts in just one speaker, AnIa, born in 1937 in the hamlet of San Rustico, whose dialect differs minimally from urban Ripano, e.g., in the inflection of feminine determiners and class I nouns.

4. Conclusion

We have isolated three types of default in Ripano, one morphological and two syntactic. The normal-case default contrasts with the other two in that it is maximally specified, while the syntactic exceptional-case and the morphological default converge to the same realization. For the syntactic exceptional-case default, this corresponds to the neuter ending, which has inherited this function from Latin. The morphological default, on the other hand, which can be synchronically modeled by means of the inheritance hierarchies in Figures 1–2, has its diachronic roots in a stage in which all final unstressed vowels but /a/ merged into /ə/ – as is the case in the dialects spoken immediately to the south – to then be differentiated again through analogical change giving rise to the intricate agreement system of this fascinating language.

Acknowledgements

Thanks to the editors for involving us in this homage to a great linguist. Part of this research was presented at the workshop 'Inheritance Hierarchies in Morphology', University of Zurich, 10–11 November 2015: we thank Geert Booij, Dunstan Brown, Sebastian Fedden, Anja Hasse, Anna Thornton, Rik van Gijn and an anonymous reviewer for valuable comments. We are grateful to our Ripano consultants, especially Luigino Cardarelli, for help with the organization of fieldwork and for responding to our many questions.

For academic purposes § 2.2, § 2.3, § 3, § 3.2 and § 4 must be ascribed to ML, § 1, § 2, § 2.1 and § 3.1 to TP. Besides those featuring in the Leipzig Glossing Rules, we use the following abbreviations: IMPF: imperfect tense, IC: inflectional class, ISC: inflectional subclass, NAN: non-autonomous neuter, UHS: unspecified human subject. Moreover, we use non- instead of N- for 'non' in nonF 'non-feminine' to avoid confusion with N(EUTER).

Funding

We gratefully acknowledge the financial support of the Swiss National Science Foundation, which partly funded this research within the project 'The Zurich database of agreement in Italo-Romance' [SNF 100012–156530].

References

Aronoff, Mark. 1994. *Morphology by itself: Stems and inflectional classes*. Cambridge, MA: MIT Press.

Aronoff, Mark. 2013. Varieties of morphological defaults and exceptions. Special issue of *ReVEL* 7. 84–97, http://www.revel.inf.br/files/8596b454a730ddbf4903e41aee04a0b8.pdf (10 August, 2019).

Baerman, Matthew, Dunstan Brown & Greville G. Corbett. 2005. *The syntax-morphology interface. A study of syncretism*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486234

Bond, Oliver. 2019. Canonical typology. In Jenny Audring & Francesca Masini (eds.), *The Oxford handbook of morphological theory*, 409–431. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199668984.013.26

Brown, Dunstan. 2016. Defaults and overrides in morphological description. In Andrew Hippisley & Gregory Stump (eds.), *The Cambridge handbook of morphology*, 272–296. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781139814720.011

Brown, Dunstan & Andrew Hippisley. 2012. *Network Morphology: A defaults-based theory of word structure*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781139814720.011

Cardarelli, Luigino. 2010. *Le parole ritrovate*. [s.l.] [s.n.].

Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139166119

Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.

Corbett, Greville G. 2009. Canonical inflectional classes. In Fabio Montermini, Gilles Boyé & Jesse Tseng (eds.), *Selected Proceedings of the 6th Décembrettes: Morphology in Bordeaux*, 1–11. Somerville, MA: Cascadilla Proceedings Project. http://www.lingref.com/cpp/decemb/6/paper2231.html

Corbett, Greville G. 2012. *Features*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139206983

Corbett, Greville G. & Norman M. Fraser. 1993. Network Morphology: A DATR account of Russian nominal inflection. *Journal of Linguistics* 29(1). 113–142. https://doi.org/10.1017/S0022226700000074

Corbett, Greville G. & Norman M. Fraser. 2000. Default genders. In Barbara Unterbeck, Matti Rissanen, Tertu Nevalainen & Mirja Saari (eds.), *Genders in grammar and cognition*, 55–97. Berlin & New York: Mouton de Gruyter.

D'Achille, Paolo, & Anna M. Thornton. 2003. La flessione del nome dall'italiano antico all'italiano contemporaneo. In Nicoletta Maraschio & Teresa Poggi Salani (eds.), *Italia linguistica anno Mille, Italia linguistica anno Duemila*, 211–230. Rome: Bulzoni.

DAI = Loporcaro, Michele, Tania Paciaroni, Diego Pescarini, Alice Idone, Serena Romagnoli, Chiara Zanini & Taras Zakharko. 2020. *The Zurich Database of Agreement in Italo-Romance*. University of Zurich, http://www.dai.uzh.ch

D'Alessandro, Roberta. 2017. When you have too many features: Auxiliaries, agreement and clitics in Italian varieties. *Glossa. A Journal of General Linguistics* 2(1). 50. https://doi.org/10.5334/gjgl.102

D'Alessandro, Roberta. 2020. Agreement across the board: Topic agreement in Ripano. In Peter W. Smith, Katharina Hartmann & Johannes Mursell (eds.), *Agree to agree: Agreement in the minimalist programme*, 235–270. Berlin: Language Science Press. https://doi.org/10.5281/zenodo.3541757

Evans, Nicholas, Dunstan Brown & Greville G. Corbett. 2002. The semantics of gender in Mayali: Partially parallel systems and formal implementation. *Language* 78. 111–155. https://doi.org/10.1353/lan.2002.0014

Fedden, Sebastian & Greville G. Corbett. 2017. Gender and classifiers in concurrent systems: Refining the typology of nominal classification. *Glossa. A Journal of General Linguistics* 2(1). 34. https://doi.org/10.5334/gjgl.177

Gisborne, Nikolas & Andrew Hippisley. 2017a. Defaults in linguistics. In Nikolas Gisborne & Andrew Hippisley (eds.), *Defaults in morphological theory*, 1–17. Oxford: Oxford University Press.

Gisborne, Nikolas & Andrew Hippisley (eds.). 2017b. *Defaults in morphological theory*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198712329.001.0001

Harder, Andreas. 1988. Laut- und Formenlehre der Mundart von Ripatransone. Kiel, Germany: University of Kiel dissertation.

Lambertelli, Giulio. 2003. *Elementi di fonetica, ortografia e grammatica della parlata ripana con doppia versione di "J'sè d lu Rr'tì"*. Acquaviva Picena (AP): Edizioni Lambertelli.

Ledgeway, Adam. 2012. *From Latin to Romance. Morphosyntactic typology and change*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199584376.001.0001

Loporcaro, Michele. 2018. *Gender from Latin to Romance: history, geography, typology*. Oxford: Oxford University Press.

Loporcaro, Michele. To appear. Nouns as agreement targets in Romance: Inflection and derivation. In Martina Werner & Wolfgang U. Dressler (eds.), *Between Derivation and Inflection*. Vienna: ÖAW-Verlag.

Loporcaro, Michele & Tania Paciaroni. 2011. Four-gender systems in Indo-European. *Folia Linguistica* 45(2). 389–434. https://doi.org/10.1515/flin.2011.015

Lüdtke, Helmut. 1976. La declinazione dei verbi in un dialetto di transizione marchigiano-abruzzese. *Abruzzo* 16(3). 79–84.

Mancini, Anna Maria. 1993. Le caratteristiche morfosintattiche del dialetto di Ripatransone (AP), alla luce di nuove ricerche. In Sanzio Balducci (ed.), *I dialetti delle Marche meridionali*, 111–136. Alessandria: Edizioni dell'Orso.

Paciaroni, Tania. 2020. *Ripano*. University of Zurich, http://www.dai.uzh.ch/ripatransone.html. (12 February, 2020.)

Paciaroni, Tania & Michele Loporcaro. 2018. Overt gender marking depending on syntactic context in Ripano. In Sebastian Fedden, Jenny Audring & Greville G. Corbett (eds.), *Non-canonical gender systems*, 147–175. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198795438.003.0007

Parrino, Flavio. 1967. Su alcune particolarità della coniugazione nel dialetto di Ripatransone. *L'Italia dialettale* 30. 156–166.

Rossi, Alfredo. 1999. *'Ne 'mbize dë perola*. Poesie e sonetti in lingua dialettale di Ripatransone 2ª raccolta. Ripatransone: Maroni.

Rossi, Alfredo. 2007. *Une tire 'n-atre… comma lë cëriescia!* Poesie, sonetti ed altro in lingua dialettale di Ripatransone. 4ª raccolta. Ripatransone: Maroni.

Rossi, Alfredo. 2008. *Dizionario del dialetto Ripano*. Ripatransone: [s.n.] (Centobuchi di Monteprandone: Linea grafica).

Thornton, Anna M. 2019. How (non-)canonical is Italian morphology? In Matthew Baerman, Oliver Bond & Andrew Hippisley (eds.), *Morphological perspectives. Papers in honour of Greville G. Corbett*, 65–99. Edinburgh: Edinburgh University Press.

PART IV

Morphomes

CHAPTER 13

Morphomes all the way down!

Andrew Spencer
University of Essex

Taking Stump's (2016) PFM2 C(ontent)/F(orm)/R(ealized) paradigm distinction I argue that the F/R-paradigm features are conceptually different from C-paradigm features. C-paradigm features interface with syntax/semantics, hence are 'interpretable'. F-paradigm features, by contrast, induce purely formal (morphomic) partitionings (cf. Boyé & Schalchli 2019), even for canonical systems (one:one Content-Form correspondence), a reflection of the true autonomy of inflectional morphology, 'morphology-by-itself'. The C-paradigm features are a subset of Sadler & Spencer's (2001) 's(yntactic)-features'. Canonical Content-Form correspondence is achieved by typing features as m- and s-features.

Keywords: content paradigm, form paradigm, inflectional paradigm, m-feature, morphology-by-itself, morphome, morphomic feature, Paradigm Function Morphology, s-feature

1. Introduction

In the days of the morpheme it was a very simple job to describe the form of an expression such as Finnish *taloissa*: three morphemes, *talo* "house", *i* "plural" and *ssa* "in", were compounded together and the meanings were combined to give the interpretation "in houses". In particular, there was no need to worry about inflectional paradigms: paradigms didn't exist and the notion of 'inflection' wasn't coherent.

With the development of realizational, lexeme-based models of morphology, such as that adumbrated in Aronoff (1994) and developed in the Paradigm Function Morphology (PFM) model of Stump (2001) matters have become more complex. A complete description of a language's inflectional system has to define sets of morphosyntactic features and their permissible values and combinations ([NUMBER:{SG, PL,…}], [CASE:{NOM, GEN, INESS,…}]) for given classes of lexemes, such that the totality of those feature combinations defines each lexeme's inflectional paradigm, while a separate component of grammar specifies the precise morphological

https://doi.org/10.1075/cilt.353.13spe

form of each word. The PFM model also includes separate machinery for defining patterns within the paradigms, such as syncretism, as well as additional machinery for defining the morphophonological shape of various stems the lexeme might deploy, and a set of selection rules defining which stem is appropriate for which set of morphosyntactic properties. The stems themselves are 'pure forms', expressing no meaning other than that of the lexeme itself, in other words, the morphomic stems, first proposed in Aronoff (1994). Stump further demonstrates that the morphophonologically defined stem sets are not necessarily congruent with the morphomic stems (his Indexing Autonomy Hypothesis, Stump 2001: 190–199), so that a complex set of mappings may be required to define the set of stems that are selected by the inflectional rules.

In later work (Stump 2002, 2006, and especially Stump 2016), Stump observes that even the full complexity of the PFM model is insufficient to do formal justice to the host of form-function or form-content mismatches which abound in inflectional systems. Stump therefore proposes a modified model, PFM2 (Bonami & Stump, 2016), in which a distinction is drawn between two types of morphological paradigm. The Content paradigm represents the interface between word structure and the compositionally defined syntactic constructions and semantically interpreted functional categories. It therefore has to be maximally regular, whatever the peculiarities of the actual word forms instantiating these categories. The Form paradigm, derived from the Content paradigm, neutralizes some of the featural distinctions in the Content paradigm, and rearranges some of the other distinctions so as to capture the form-content mismatches.

In this chapter I explore the implications of Stump's proposals for the notion of morphomic organization as proposed by Aronoff (1994). Following Round (2015) I take a morphome to be any morphological object, expression, relation, or pattern that is independent of phonology, syntax, semantics, or any other non-morphological domain. I will focus on the notion of 'morphomic morphological feature'. In most of Stump's discussion, the feature sets which define the Form paradigms are either identical to, or a proper subset of, the (by definition non-morphomic) Content paradigm features. However, he provides evidence that in some inflectional systems it is necessary to posit features which do nothing more than regulate the form of inflected words, without appeal to any principles or properties outside of morphology, that is morphomic features. In fact, such morphomic feature systems are quite common, even though they have generally gone unrecognized as such (an important exception being the work of Boyé & Schalchli 2019). I will argue that in fact all features in Form paradigms are morphomic, even if they appear to be directly related to non-morphomic Content paradigm features. The convention of giving Form paradigm features the same labels as the interpretable Content paradigm features to which they (usually) correspond is thus misleading.

But it would actually make the architecture of PFM2 difficult to understand (and possibly even incoherent) if we were to regard the Form paradigm features as anything but morphomic. Making all Form paradigm features explicitly morphomic therefore constitutes a (very modest) simplification of PFM2.

In § 2 I present the Content~Form paradigm distinction in more detail. In § 3 I illustrate cases which show the need for irreducibly morphomic features in Form paradigms. I extend this argumentation in § 4, where I argue that we should make morphomicity a definitional property of Form paradigm features, extending recent proposals by Boyé & Schalchli (2019). In § 5 I propose we adopt Sadler & Spencer's (2001) m-/s-feature distinction (broadly speaking, 'morphological' vs 'syntactic' features): the m-features are precisely the morphomic Form paradigm features, while s-features are a generalization of the interface features of PFM2 which define Content paradigms.

2. Two types of paradigm

Inflected morphological forms (i) express meanings of certain kinds (dual number, inessive case, jussive mood, etc., cf. Booij's (1996) 'inherent inflection') or (ii) grammatical relationships (adjective – noun agreement, ezafe, different-subject clause, etc., cf. Booij's (1996) 'contextual inflection'). In each case, the inflected form is associated with (realizes) some morphosyntactic property. However, as extensively illustrated by Stump (2016), the inflectional paradigms defined by those morphosyntactic properties do not always map neatly onto the paradigms defined by the inflected forms of individual lexemes or classes of lexemes themselves. Pretty well any language with inflectional morphology will have lexemes which exhibit such mismatches as syncretism, deponency, heteroclisis, or other types. Stump therefore develops a modification of his Paradigm Function Morphology model, PFM2, drawing a distinction between the Content Paradigm (which I shall abbreviate as Π_C) and the Form Paradigm Π_F. The Π_C is a completely regular set of oppositions defined by those morphosyntactic properties which have a direct interpretation in syntax/semantics. By contrast, the Π_F reflects just those sets of oppositions which are reflected in the set of distinguishable word forms. (The set of actual word forms and their associated morphosyntactic properties is the Realized Paradigm, Π_R.)

For instance, in Indo-European languages such as Latin or Russian, neuter gender nouns of whatever inflectional class syncretise their NOM case forms with ACC case in all numbers. For such nouns the Π_F therefore has a different structure from the Π_C. Stump handles cases such as these by defining a mapping, ***Corr(espondence)***, between Π_C and Π_F, which maps the property set defining the Π_C to a (possibly slightly different) property set, reflecting the (slightly different) organization of the Π_F, a relation he calls 'Paradigm Linkage'. The ***Corr*** mapping is defined in terms of a property mapping function, ***pm***. For instance, we might set up a ***pm*** function of the kind shown in (1).

(1) For any neuter gender noun lexeme the following mapping holds:
pm(NOM) = ACC

These special mappings override the canonical, default mapping, which simply copies the feature structure of the Π_C to the Π_F, as shown in (2).

(2) The following mapping holds for any lexeme of any class, for any collection of morphosyntactic property specifications $\sigma \subset \Phi(\Pi_C)$ (= the set of feature specifications which define the Content paradigm, Π_C):
pm$(\sigma) = \sigma$

The property set which defines the Form paradigm, $\Phi(\Pi_F)$, is, in the most general case, distinct from σ. This means that, formally speaking, (2), like (1), is actually a special case of the general mapping, (3).

(3) ***pm***$(\sigma) = \tau$

In practice, τ is often a proper subset of σ, but in § 4 we will see cases in which PFM2 would treat the sets σ, τ as overlapping. Arguably (see § 3), for a highly aberrant system such as English verb inflection we may require the two feature sets to be completely disjoint.

3. Morphomic features in form paradigms

In most discussion of morphosyntactic properties/features that draw any distinction at all between their morphological role and their syntactico-semantic role it is (tacitly) assumed that the two roles or functions are fulfilled by a single feature, rather than, say, by two distinct but (possibly) related features. In other words, the attribute-value pairing [NUMBER:plural] in the Π_C, which permits syntax/semantics to treat a noun phrase as having a plural denotation, triggering plural agreement and so on, is identical to the attribute-value pairing which determines the way that a noun is inflected for plural number (say, by suffixation of *-es*), in the morphology (Π_F/Π_R). There are two important observations to make here.

First, no explicit conceptual justification has ever been offered for identifying the two types of feature, despite the fact that they fulfil such disparate functions. Second, it is universally acknowledged that there are morphomic Π_F features, so the question arises as to how these can properly be integrated into a feature set which includes non-morphomic (syntactically-semantically interpreted) features.

For instance, in his discussion of the Nepali long negative present tense Stump (2016: 139–145) adopts the important proposals of Bonami & Boyé (2008, 2010). They argue very persuasively that none of the normal machinery for delivering the various types of syncretism will work for the Nepali case. They therefore rearrange the paradigm adding an extra row and defining the paradigm (what in PFM2 is the Π_F) in terms of two completely morphomic features, which they label [COL: {a, b, …}] and [ROW:{α, β, …}], underlining the fact that these are Π_F features with no Π_C correspondents whatsoever. Independently of PFM, Round (2015) shows how the apparently agglutinative affixal morphology of Kayardild exhibits pervasive morphomic organization.

The problem posed by morphomic feature oppositions is, in fact, ubiquitous, but it is underestimated because most discussion of morphomes is focussed, understandably, on those scenarios in which a morphomic analysis is the only possible one, and there is no plausible reanalysis of the morphome in terms of syntactico-semantic properties. However, it is very common to find morphomic feature sets as a consequence of the severe attrition of any erstwhile rich inflectional system.

Depleted case systems provide some good examples (Arkadjev, 2009). Old French nouns distinguished [NUMBER:{SG, PL}] and [CASE:{NOM, ACC}] but a noun such as MUR "wall" had just two forms: ⟨murs, {NOM SG, ACC PL}⟩ and ⟨mur, {ACC SG, NOM PL}⟩. Clearly, the Π_F here can only be defined in terms of a purely morphomic, two-valued feature such as [NFORM:{s, null}].

A more complex example is that of German noun inflection, as pointed out in Spencer (2009) (similar arguments are adduced by Boyé & Schalchli 2019). The German determiner system requires the feature oppositions [NUMBER:{SG, PL}], [CASE:{NOM, ACC, GEN, DAT}], and [GENDER:{m, f, n}]. However, a noun such as HAUS "house" systematically distinguishes at most four forms (three in the case of feminine nouns): ⟨Häusern, {DAT PL}⟩, ⟨Häuser, {¬DAT PL}⟩, ⟨Hauses, {GEN SG}⟩, ⟨Haus, {¬GEN SG}⟩. A 'weak' noun such as STUDENT "student" on the other hand distinguishes just the forms ⟨Student, {NOM SG}⟩, ⟨Studenten, {elsewhere}⟩. Thus, no noun ever distinguishes [CASE:ACC], and no noun factorizes case and number as independent properties. This makes the system entirely morphomic.[1]

The English verb system likewise demonstrates a radical mismatch between functional features and purely formal oppositions. A regular transitive verb such as PAINT has four forms {*paint, paints, painting, painted*}, of which *painting*

1. There are syntactic nouns converted from adjectives which retain all the morphosyntax of adjectives, including gender and the strong~weak~mixed inflectional classes system, the 'Angestellte(r)nouns' of Spencer (2013). These further illustrate the need for radical separation between Π_C features and Π_F/Π_R features for nouns, since morphologically the Π_C and the Π_F don't even share the same major lexico-syntactic category label.

corresponds to no less than three lexico-syntactic categories (parts of speech): verb (*Harriet is painting a picture*), adjective (*the artist painting this picture*), noun (*painting is a relaxing pastime*). The form *painted* realizes (finite) past tense, as well as the perfective (active) and passive participle(s), and the bare form *paint* expresses every other category except 3SG non-past. Arguably, therefore, the best way of describing the verb Form paradigm is in terms of a single four-valued feature m-[VFORM:{s, ing, ed, bare}], making the Π_F feature set completely disjoint from the Π_C feature set (Spencer, 2020).

The importance of such observations has recently been underscored by Boyé & Schalchli (2019). They present the concept of the Form paradigm from the perspective of what they call the Paradigm Cell Finding problem: given a set of inflected word forms, how do we compute the set of featural oppositions expressed by a language's morphology? They point out that we can treat Π_F oppositions as that set of equivalence classes of word forms which syncretise corresponding Π_C oppositions. This leads them to define an Optimal Morphomic Paradigm (OMP) " … a set S of morphosyntactic property sets such that members of S are *identical* in their realization across *lexemes*." (Boyé & Schalchli 2019: 213, emphasis original). In the OMP, " … every cell represents a systematic syncretism". In the implementation of Boyé and Schalchli the OMP represents the maximum set of distinctions made by any particular lexeme, including overdifferentiating lexemes such as English BE. This entails that all verbs in English have, for instance, a 1SG present indicative form (because BE has *am*), even though that form syncretises with the base form for every other verb in the language. However, we can factor out cases of overdifferentiation and still define the Π_F/Π_R in terms of purely morphomic features.[2]

4. All form paradigm features are morphomic

Despite the clear evidence for morphomic Π_F features standard practice in morphology is to assume a single vocabulary of features applicable to both morphology and syntax. This may seem reasonable enough for highly agglutinating systems such as Turkish noun morphology, illustrated by Stump (2016: 106–108) and shown in abbreviated form in Table 1 (cf. Stump 2016: 112, Table 7.10). Indeed, some might find it perverse to provide distinct labels in such cases. But tabular representations of this sort are misleading, as argued extensively for the similar case system of

2. In Boyé & Schalchli's definition it is assumed that we have already identified the lexical classes of the language. We don't want to say that every noun has a past tense form, for instance. However, it must be recognized that in many languages it is a far from trivial matter to individuate all the morpholexical categories.

Hungarian, and for case systems generally in Spencer (2008). In purely formal terms, we could, for instance, relabel the cells in the Form paradigm by reversing the 'SG' and 'PL' specifications and nothing at all would be lost (other than superficial transparency). Tables such as Table 1 therefore fail to reflect the fact that a given feature specification, say, 'SG', is constant across the paradigms for each corresponding cell.

Let us ask what it would mean if the Π_F/Π_R cells were actually defined in terms of the same property sets or feature-value specifications as those of the Content paradigm. First, consider the structure of morphological morphosyntactics. Given a set of features F_i, with values $v_{i,j}$ a paradigm is a set of orthogonal arrays. The property set σ = {[CASE:{NOM, GEN}], [NUMBER:{SG, PL}]} defines the 4×1 Π_C submatrix of Table 1. The corresponding Π_F/Π_R submatrices are isomorphic to Π_C, except that the cells of Π_C take the form ⟨Lexemic Index, σ_i⟩, while the cells of the Π_F/Π_R are occupied by pairs ⟨lexical root, σ_i⟩, ⟨inflected word form, σ_i⟩, respectively. However, the feature set σ is supposed to be common to all three paradigms, so a better way of diagramming the situation is as in Table 2, in which the feature labels are abstracted away from each of the cells and generalized across the three paradigms. We can think of such a diagram as defining an address for each cell in each paradigm by computing the product of labels such as ADAM, *adam, adamların* and a set of feature specifications such as {[CASE:GEN], [NUMBER:PL]}.[3]

Table 1. Abbreviated case/number paradigm for Turkish noun ADAM "man"

Content paradigm	Form paradigm	Realized paradigm
⟨ADAM, {NOM SG}⟩	⟨*adam*, {NOM SG}⟩	⟨*adam*, {NOM SG}⟩
⟨ADAM, {GEN SG}⟩	⟨*adam*, {GEN SG}⟩	⟨*adamın*, {GEN SG}⟩
⟨ADAM, {NOM PL}⟩	⟨*adam*, {NOM PL}⟩	⟨*adamlar*, {NOM PL}⟩
⟨ADAM, {GEN PL}⟩	⟨*adam*, {GEN PL}⟩	⟨*adamların*, {GEN PL}⟩

Table 2. Turkish noun paradigm with abstracted features

Property set	Content paradigm	Form paradigm	Realized paradigm
CASE:NOM, NUMBER:SG	ADAM	adam	adam
CASE:GEN, NUMBER:SG	ADAM	adam	adamın
CASE:NOM, NUMBER:PL	ADAM	adam	adamlar
CASE:GEN, NUMBER:PL	ADAM	adam	adamların

3. Better still would be a multidimensional diagram in which the three paradigms each occupied a separate dimension and the column of property sets, σ, formed a fourth dimension whose elements are linked to the cells in each of the three paradigms' dimensions.

Table 2 characterizes the PFM2 analysis of canonical paradigms. For the majority of the mismatches discussed by Stump the ***pm*/*Corr*** functions map elements of the Π_C feature set, σ, to $\tau \subset \sigma$, effectively directing the morphology to a set of Π_F cells[4] corresponding to some other Π_C cell. However, where the Π_F is defined by a disjoint morphomic feature set it is impossible to abstract out a single overarching feature set in the manner of Table 2 and a type of organization is required such as that shown schematically for Old French nouns in Table 3.

Table 3. Old French noun paradigm

Content paradigm		Form/realized paradigm	
CASE:NOM NUMBER:SG	MUR	[NFORM:S]	*/mur/, /murs/*
CASE:ACC, NUMBER:SG	MUR	[NFORM:null]	*/mur/, /mur/*
CASE:NOM, NUMBER:PL	MUR		
CASE:ACC, NUMBER:PL	MUR		

Table 4. German noun paradigm (first pass)

Content paradigm		Form/realized paradigm	
CASE:NOM NUMBER:SG	HAUS	[NUMBER:SG]	*/haus/*
CASE:ACC, NUMBER:SG	HAUS	[NFORM:S]	*/hauses/*
CASE:GEN, NUMBER:SG	HAUS	[NUM:PL]	*/häuser/*
CASE:DAT, NUMBER:SG	HAUS	[NFORM:n]	*/häusern/*
CASE:NOM NUMBER:PL	HAUS		
CASE:ACC, NUMBER:PL	HAUS		
CASE:GEN, NUMBER:PL	HAUS		
CASE:DAT, NUMBER:PL	HAUS		

Where we have a mixed system, in which not all of the Π_F features are strictly morphomic, matters get more complex. Consider the analysis of German strong nouns proposed in Spencer (2009). On PFM2 assumptions these have a straightforward semantically interpretable singular/plural number opposition, but the case feature is morphomic. This requires a set of paradigm matrices along the lines of Table 4 (with Π_R as proxy for Π_F). As a symptom of the conceptual problem at hand observe that the choice of Π_F/Π_R feature labels is now largely arbitrary. The ***Corr*** function will map non-genitive case singular forms to a cell labelled (non-morphomically) [NUMBER:SG], but will map the genitive singular and dative plural cells to the purely morphomic [NFORM:s/n] cells.

4. The structure of the Π_R is by definition isomorphic to that of the Π_F, so I will conflate the two from now on.

The solution I propose is to adopt the Morphomicity Principle (4), effectively generalizing Boyé & Schalchli's (2019) OMP to a strict architectural principle[5]

(4) Morphomicity Principle
A feature is a member of the feature array of a Π_F/Π_R iff it is morphomic

If the features defining Form paradigm cells are a different kind of object from those defining Π_C cells then this will necessitate a minor reconfiguring of the PFM2 ***Corr*** and especially the ***pm*** functions. In particular, the identity mapping $\boldsymbol{pm}(\sigma) = \sigma$, is no longer available to us as a (partial) definition of the canonical inflectional system, or, more importantly, as a default mapping where there are no $\Pi_C{\sim}\Pi_F$ mismatches. What we need to do is to rewrite the ***pm*** function so that by default it defines a kind of 'alphabetic variant' of the Π_C, preserving all the feature structure but systematically relabelling the features. We already have the necessary machinery available to any model which countenances feature-typing. We first assume that all morphosyntactic features are typed as 's-' or 'm-' features: $\langle$s, $\sigma_i\rangle$, $\langle$m, $\sigma_j\rangle$.[6] We can then rewrite the canonical identity ***pm*** as a function which just re-types s-features as m-features, as in (5).

(5) Given σ_C = the set of all feature specifications in Π_C, $\forall$i, if $\langle$s, $\sigma_i\rangle \in \sigma_C$, then $\boldsymbol{pm}(\langle \text{s}, \sigma_i\rangle) = \langle \text{m}, \sigma_i\rangle$

Let us call the set of all the (morphomic) Π_F features τ_F. By definition, these features have the form $\langle$m, $\tau\rangle$. Where we have mismatches defined in terms of a subset, τ, of σ in PFM2 we simply redefine the ***pm*** function over a subset of τ_F. In fact, the feature-typing device is already implicit in PFM2, as can be seen from Table 2. There we can see that in a canonical inflectional system the cells of Π_C, Π_F, and Π_R are essentially the same cells except that their feature specifications are indexed or

5. In PFM1/PFM2 rules of stem selection are formally distinct from rules of exponence. Spencer (2016) points out that this entails that all stems in PFM are technically morphomic, even if they are traditionally given labels such as 'plural stem' or 'past tense stem'. Thus, all components of the Π_F/Π_R will by definition be morphomic under my proposals. In Information-based Morphology (IbM) (Crysmann & Bonami 2016) the PFM rules of exponence are redefined as principles governing the placement of morphs, and stem selection is reduced to a species of exponence. On the assumption that there exist non-morphomic Π_F/Π_R features, the IbM model therefore undermines the arguments of Spencer (2016), but if we accept the Morphomicity Principle, (4), in IbM too, then IbM is entirely compatible with the arguments of Spencer (2016).

6. Such a typing would be very straightforward in Crysmann & Bonami's (2016) IbM model. The same effect is achieved in earlier work by Bonami and colleagues by making s-features values of the HEAD attribute and m-features values of an INFL attribute, and then stating that by default the INFL features are re-entrant with the HEAD features (Bonami et al. 2016: 113). Space unfortunately doesn't permit a detailed comparison of these rather different technical devices.

typed with the Lexemic Index (Π_C), root/stem (Π_F) or inflectional word form (Π_R) of the inflected word.

A corollary of the Morphomicity Principle is that Π_C feature labels are always distinct from Π_F labels. For Old French this means we can just adopt the morphosyntactic organization of Table 3. For German strong nouns the Π_F should look something like Table 5.[7]

Table 5. German noun paradigm, morphomic analysis

Content paradigm		Form/realized paradigm	
CASE:NOM NUMBER:SG	HAUS	[NFORM:'SG']	*/haus/*
CASE:ACC, NUMBER:SG	HAUS	[NFORM:S]	*/hauses/*
CASE:GEN, NUMBER:SG	HAUS	[NFORM:'PL']	*/häuser/*
CASE:DAT, NUMBER:SG	HAUS	[NFORM:n]	*/häusern/*
CASE:NOM NUMBER:PL	HAUS		
CASE:ACC, NUMBER:PL	HAUS		
CASE:GEN, NUMBER:PL	HAUS		
CASE:DAT, NUMBER:PL	HAUS		

An interesting consequence of the pan-morphomic treatment of Form paradigms is that it entails that the syntax/semantics interface has no access to the structure of the Π_F/Π_R. The Π_C is defined in terms of s-features, and it is only these that syntax/semantics has access to in PFM2.[8] But if all Π_F cells are defined in terms of morphomic/m-features then by definition the interface will have no access to those Π_F (or Π_R) cells. This entailment effectively defines the doctrine of morphological autonomy or 'morphology-by-itself' for inflection. In fact, we can think of the morphomic relabelling of morphosyntactic properties in Form/Realized paradigms as an index or emblem of morphologization.

7. NB: [NFORM:{'SG', 'PL'}] is an entirely different feature from [NUMBER:{SG, PL}]).

8. A very interesting potential exception to pure morphomicity might be the 'semi-autonomous' feature patterning described for Hualapai by Baerman (2019). This very complex system is reminiscent of the morphomic stem patterns of Sanskrit which motivate Stump's Indexing Autonomy Hypothesis. However, only detailed further analysis will confirm whether the analysis presented by Baerman counterexemplifies the Morphomicity Principle.

5. Two types of feature[9]

A distinction between morphological features (m-features) and syntactic features (s-features) has already been drawn by Sadler & Spencer (2001). Their m-features are essentially the morphomic Π_F features introduced already. However, for Sadler & Spencer the s-features include properties of purely syntactic constructions such as subject-auxiliary inversion or definite article insertion, as well as inflectional properties such as noun number. In other words, the set of s-features as a whole properly contains the set of syntactically or semantically interpretable interface features that define Content paradigms.

The rationale behind Sadler & Spencer's (2001) distinction comes from inflectional property mismatches seen in periphrastic exponence, as when past tense is expressed by a periphrase consisting of a non-finite converb combined with a *present* tense auxiliary (as, for instance, Spoken French, and many of the Slavic languages).

The problem of periphrastic constructions and the morphosyntactic architecture is a rather complex one, and there is as yet no consensus on how best to handle the feature mismatches that periphrases give rise to. I will therefore illustrate the expanded m-/s-feature distinction by considering the expression of definiteness in Danish.

Nouns in Danish inflect for singular/plural number and in the singular belong to either the common or the neuter gender. Attributive adjectives have an indefinite and a definite declension. The definite declension forms are found when the NP includes a definite article, demonstrative, possessive, or similar. The indefinite forms are found elsewhere, and especially with the indefinite article *en, et*, with count nouns, and also in predicative constructions (Allan et al. 2000: 35). As seen in Examples (6), (8) in indefinite contexts and in the predicative construction the adjective takes agreement inflections. The definite noun forms *bilen/bilerne, huset/huserne* will be discussed presently. In definite contexts, by contrast, the adjective has an invariant form in *-e* (homophonous with the plural inflected form).

(6) Danish indefinite NPs

a. *en* *stor* *bil*
a.COMMON big-COMMON.INDEF car
"a big car"

b. *et* *stor-t* *hus*
a.NEUTER big-NEUTER.INDEF house
"a big house"

9. This discussion expands on the somewhat programmatic remarks in Sadler & Spencer (2001).

c. *stor-e bil-er/hus-e*
big-PL car-PL/house-PL
"big cars/houses"

(7) Danish definite NPs
a. *den/min stor-e bil*
the.COMMON/my.COMMON big-DEF car
"the/my big car"
b. *det/mit stor-e hus*
the.NEUTER/my.NEUTER big-DEF house
"the/my big house"
c. *de/mine stor-e bil-er/hus-e*
the.PL/my.PL big-DEF car-PL/house-PL
"the/my big cars/houses"

(8) Danish predicative adjectives
a. *bil-en er stor*
car-DEF.ART.COMMON is big.SG.INDEF
"The car is big"
b. *hus-et er stor*
house-DEF.ART.NEUTER is big.SG.INDEF
"The house is big"
c. *bil-erne/hus-erne er stor-e*
car-DEF.ART.PL/house-DEF.ART.PL are big.PL
"The cars/houses are big"

Crucially, the definite article *den, det, de* is only found when the noun is in construction with an attributive modifier, for instance, an adjective. We cannot say, for instance, **Den bil er stor* (*e*) for "The car is big". A simple unmodified noun is made definite morphologically, by suffixation. The main allomorphs are *-en*, common gender, singular, *-et*, neuter gender, singular, and *-ne*, plural.

Clearly, at some level of description both the *-en* suffix of *bilen* and the definite article *den* realize the same property of definiteness. We must therefore assume for Danish an s-feature s-[DEF:{yes, no}], such that the specification s-[DEF:yes] determines selection of the definite form of the noun (Examples (8)), and selection of the definite article (Examples (7)).[10] Thus, in at least one of its roles (definite article selection) it is a purely syntactic feature, with no reflex in the determination of morphological form. Note that we need to be careful about what morphosyntactic description to assign to *bil* in (7). It is not, despite appearances, [DEF:no], since that would give rise to a fatal clash of specifications for [DEF].

10. It also selects the special 'weak declension' inflectional paradigm on the adjective.

The problem posed by constructions such as Danish definiteness marking for any inferential/realizational model and in particular for PFM2 is this: The Π_C of PFM2 represents the interface between morphological form and syntax/semantics. But selection of the definite article is not conditioned by any feature in the array which defines the Π_C for any lexeme: definiteness is realized by (selection of) the definite article lexeme as a whole, not by specific inflected forms of that lexeme. Thus, there seems to be no way of saying in PFM2 that selection of a free-standing definite article function word and the definition of the [DEF:yes] cells in a noun's morphological paradigm are governed by the same feature (marking), and have exactly the same syntactic/semantic interpretation. In the expanded conception of morphosyntactic feature presented here both the analytic (synthetic) variant of the construction, and the synthetic variant are defined by the same semantically-interpretable s-[DEF:{yes, no}] feature, mapped in the synthetic construction to the m-[DEF:{yes, no}] feature in the Π_F of the head noun.

6. Conclusions

Taking Stump's (2016) PFM2 model as a starting point I have argued against the widespread, if generally tacit, assumption that the morphological properties that define inflected word shapes (Form paradigm features) are exactly the same formal objects as the syntactically/semantically interpretable interface properties that define Content paradigms.

In the 'morphology-by-itself' program of autonomous morphology launched in Aronoff (1994), the strongest assumption is that all such Form paradigm-based properties are 'informationally encapsulated' from syntax/semantics, i.e., Form paradigms are defined in terms of purely morphomic features. This considerably streamlines the relationship between Content and Form paradigms (Paradigm Linkage) for systems in which some, at least, of the Form features are irreducibly morphomic. This conclusion is strengthened if we accept Spencer's (2016) arguments for the morphomicity of stems in PFM2. On the empirical side, this opens the way to a straightforward analysis of recalcitrant inflectional systems such as English conjugation or German noun declension, which would otherwise remain hopelessly confused.

The connection between Content and Form paradigm features in the absence of mismatches is captured by typing otherwise identical features with the 's-' and 'm-' type indices. The s-features that define Content paradigms can then be taken to be a subset of the syntactic features of Sadler & Spencer (2001), as exemplified by the dual definiteness constructions of Danish, in which definiteness is expressed either by the free-standing definite article or by suffixed forms of the noun.

In sum, the cells of the Form paradigm are defined by morphomic address labels, and morphological realization rules are defined over purely morphomic stems which undergo morphomic exponence operations of various kinds, in particular, combining with purely morphomic affixes. In an autonomous morphology, everything is morphomic.

Acknowledgements

I am grateful to Greg Stump for extremely detailed comments on an earlier draft. He is not responsible for any remaining errors or infelicities. It is my pleasure to record here my gratitude to Mark Aronoff for doing more than any other linguist to bring the serious study of morphology into mainstream theoretical/generative linguistics. Like much that I have written over the past 25 years this chapter should be viewed as a very modest contribution to the program initiated in Aronoff (1994).

Abbreviations

ACC	accusative
Corr	Correspondence (function) (PFM2)
DAT	dative
DEF	definite
GEN	genitive
INESS	inessive
NOM	nominative
OMP	Optimal Morphomic Paradigm
PFM	Paradigm Function Morphology
PL	plural
pm	property mapping (PFM2)
SG	singular
Π_C	Content paradigm
Π_F	Form paradigm
Π_R	Realized paradigm

References

Allan, Robin, Philip Holmes & Tom Lundskær-Nielsen. 2000. *Danish: An essential grammar*. London: Routledge.

Arkadjev, Peter M. 2009. Poor (two-term) case systems: Limits of neutralization. In Andrej Malchukov & Andrew Spencer (eds.), *The Oxford handbook of case*, 686–699. Oxford: Oxford University Press.

Aronoff, Mark. 1994. *Morphology by itself: Stems and inflectional classes*. Cambridge, MA: MIT Press.

Baerman, Matthew. 2019. Feature duality. In Matthew Baerman, Oliver Bond & Andrew Hippisley (eds.), *Morphological perspectives. Papers in honour of Greville G. Corbett*, 124–137. Edinburgh: Edinburgh University Press.

Bonami, Olivier, Robert D. Borsley & Maggie Tallerman. 2016. On pseudo-non-finite clauses in Welsh. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar*, 104–124. Stanford, CA: CSLI Publications.

Bonami, Olivier & Gilles Boyé. 2008. Paradigm shape is morphomic in Nepali. Paper presented at the 13th Internation Morphology Meeting, Vienna.

Bonami, Olivier & Gilles Boyé. 2010. Opaque paradigms, transparent forms in Nepali conjugation. Paper presented at the Workshop on Theoretical Morphology 5, Lutherstadt, Wittenberg.

Bonami, Olivier & Gregory Stump. 2016. Paradigm Function Morphology. In Andrew Hippisley & Gregory Stump (eds.), *The Cambridge handbook of morphology*, 449–481. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781139814720.017

Booij, Geert. 1996. Inherent versus contextual inflection and the split morphology hypothesis. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 1995*, 1–16. Dordrecht: Kluwer Academic Publishers. https://doi.org/10.1007/978-94-017-3716-6_1

Boyé, Gilles & Gauvain Schalchli. 2019. Realistic data and paradigms: The paradigm cell finding problem. *Morphology* 19. 199–248. https://doi.org/10.1007/s11525-018-9335-1

Crysmann, Berthold & Olivier Bonami. 2016. Variable morphotactics in Information-based Morphology. *Journal of Linguistics* 52. 311–374. https://doi.org/10.1017/S0022226715000018

Round, Erich. 2015. Rhizomorphomes, meromorphomes and metamorphomes. In Matthew Baerman, Dunstan Brown & Greville G. Corbett (eds.), *Understanding and measuring morphological complexity*, 29–52. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198723769.003.0003

Sadler, Louisa & Andrew Spencer. 2001. Syntax as an exponent of morphological features. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 2000*, 71–96. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-017-3724-1_4

Spencer, Andrew. 2008. Does Hungarian have a case system? In Greville G. Corbett & Michael Noonan (eds.), *Case and grammatical relations*, 322–340. Amsterdam: John Benjamins Publishing Co. https://doi.org/10.1075/tsl.81.02doe

Spencer, Andrew. 2009. Realization-based morphosyntax: The German genitive. In Patrick O. Steinkrüger & Manfred Krifka (eds.), *On inflection*, 173–218. Berlin: Mouton de Gruyter.

Spencer, Andrew. 2013. *Lexical relatedness: A paradigm-based model*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199679928.001.0001

Spencer, Andrew. 2016. Stems, the morphome, and meaning-bearing inflection. In Ana R. Luís & Ricardo Bermúdez-Otero (eds.), *The morphome debate: Diagnosing and analysing morphomic patterns*, 207–227. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198702108.003.0008

Spencer, Andrew. 2020. Inflection and derivation. In Bas Aarts, Jill Bowie & Gergana Popova (eds.), *Oxford Handbook of English Grammar*, 243–261. Oxford: Oxford University Press.

Stump, Gregory. 2001. *Inflectional morphology: A theory of paradigm structure*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486333

Stump, Gregory. 2002. Morphological and syntactic paradigms: Arguments for a theory of paradigm linkage. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 2001*, 147–180. Dordrecht: Kluwer Academic. https://doi.org/10.1007/978-94-017-3726-5_5

Stump, Gregory. 2006. A theory of heteroclite inflectional paradigms. *Language* 82. 279–322. https://doi.org/10.1353/lan.2006.0110

Stump, Gregory. 2016. *Inflectional paradigms. Content and form at the syntax-morphology interface*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781316105290

CHAPTER 14

Conditional exponence

Gregory Stump
Emeritus, University of Kentucky

In inferential-realizational theories of morphology, the content realized by the application of a rule of exponence is customarily assumed to be invariant across the range of contexts in which that rule applies. Yet, there are morphomic exponents whose content is sensitive to context; Breton verb inflection presents striking examples of exponents of this sort. I argue that the patterns of context dependency presented by such exponents are an effect of the phenomenon of rule combination, by which simple rules of morphology may combine to form more complex rules. Some rule combinations involve ordinary rule composition; the properties of a combination of this sort are deducible from those of its component rules. Other rule combinations are supplementational; a combination of this sort possesses one or more properties that are not deducible from those of its component rules. As I show, the Breton examples of conditional exponence involve supplementational rule combination. I elucidate this claim formally and discuss its wider implications for morphological theory.

Keywords: conditional exponence, morphome, rule combination, rule composition, rule of exponence

1. Introduction: Exponence can be conditional

In an inferential-realizational theory of inflectional morphology,[1] it is usual to assume that a rule of exponence applies to a stem Z paired with a particular set σ of

1. Stump (2001: Chapter 1) proposes a four-way classification of theories of inflectional morphology based on two cross-cutting criteria. In lexical theories (≠ lexicalist theories!), bound morphs are assumed to be lexically listed, hence inserted into larger structures from the lexicon in much the same way as free forms. By contrast, inferential theories hold that bound morphs are introduced by rules that apply to stems to form more complex morphological combinations; such rules are of the same type as rules that effect nonconcatenative morphological operations such as ablaut or consonant gradation and rules of suprasegmental morphology. Incremental theories construct the content of a complex word concomitantly with its form. Realizational theories, by contrast, hold that a word's content determines its form. The lexical/inferential distinction

https://doi.org/10.1075/cilt.353.14stu

properties to realize a subset of σ by marking Z with a particular inflectional exponent; for instance, the rule of exponence in (1a) applies to ⟨stem, property set⟩ pairings such as those in (1b) to realize the subset {AGR:{1PL}} by suffixing *-mus* to each stem, yielding the respective pairings in (1c).

(1) a. [{AGR:{1PL}} : *-mus*]
b. ⟨ *amā-*, {AGR:{1PL}, TNS:PRS, PERF:–, VCE:act, MOOD:IND}⟩
⟨ *amābā-*, {AGR:{1PL}, TNS:PST, PERF:–, VCE:act, MOOD:IND}⟩
⟨ *amāvi-*, {AGR:{1PL}, TNS:PRS, PERF:+, VCE:act, MOOD:IND}⟩
c. ⟨ *amāmus*, {AGR:{1PL}, TNS:PRS, PERF:–, VCE:act, MOOD:IND}⟩
⟨ *amābāmus*, {AGR:{1PL}, TNS:PST, PERF:–, VCE:act, MOOD:IND}⟩
⟨ *amāvimus*, {AGR:{1PL}, TNS:PRS, PERF:+, VCE:act, MOOD:IND}⟩

The content realized by the application of a rule of exponence is customarily assumed to be invariant across the range of contexts in which that rule applies; rule (1a), for example, is assumed to realize {AGR:{1PL}} and nothing else, no matter what the context of its application. This fact is in conformity with the classic, sign-based conception of the morpheme as the minimal pairing of a form with a content. Yet, there are clear cases in which the content realized by a rule of exponence is conditional; the exponent introduced by such a rule might be characterized as a morphome (Aronoff 1994), in the sense that it lacks any unitary syntactico-semantic function, depending instead upon its context for the precise identification of its function. Dismissing the idea that the content expressed by a rule of exponence might ever be conditional, one might argue that apparent cases of this sort actually involve distinct but homophonous exponents. This argument, however, fails to account for the fact that an exponent whose content is seemingly conditional may be obviously unitary in other respects.

Consider, for instance, the finite paradigm of Breton SKRIVAÑ “write” in Table 1. Analyzed purely in terms of their form, most of the word forms in Table 1 have suffix sequences consisting of up to three inflectional suffixes; in Table 2, these suffixes are abstracted from the forms in Table 1. The suffixes in these sequences can be sorted into the three groups in (2a–c): (a) initial suffixes (= those that precede a vocalic suffix), (b) vocalic suffixes, and (c) terminations (= suffixes that follow vocalic suffixes). Suffixes in the fourth group (2d) are portmanteaus, comprising both a vocalic element comparable to the suffixes in (2b) and a consonantal element comparable to the terminations in (2c).

---

cross-cuts the incremental/realizational distinction, leading to four logically distinct kinds of theories of inflectional morphology: lexical-incremental, lexical-realizational, inferential-incremental, and inferential-realizational. Theories of all four types have their proponents. A variety of factors favor the inferential-realizational type over the alternatives, as has been argued in a number of contexts (Anderson 1992; Matthews 1972; Stump 2001; Zwicky 1986).

Table 1. Finite forms of Breton SKRIVAÑ "write"

	Indicative				Irrealis		Imperative
	Present	Imperfect	Future	Past	Present	Past	
1SG	*skrivan*	*skriven*	*skrivin*	*skrivis*	*skrivfen*	*skrivjen*	
2SG	*skrivez*	*skrives*	*skrivi*	*skrivjout*	*skrivfes*	*skrivjes*	*skriv*
3SG	*skriv*	*skrive*	*skrivo*	*skrivas*	*skrivfe*	*skrivje*	*skrivet*
1PL	*skrivom*	*skrivem*	*skrivim*	*skrivjom*	*skrivfem*	*skrivjem*	*skrivom*
2PL	*skrivit*	*skrivec'h*	*skrivot*	*skrivjoc'h*	*skrivfec'h*	*skrivjec'h*	*skrivit*
3PL	*skrivont*	*skrivent*	*skrivint*	*skrivjont*	*skrivfent*	*skrivjent*	*skrivent*
IMPS	*skriver*	*skrived*	*skrivor*	*skrivjod*	*skrivfed*	*skrivjed*	

Sources:
Trépos (n.d.: §§ 479, 499, 510, 520, 533, 543, 563);
Kervella (1976: § 206)

Table 2. Suffixes sequences in the inflection of Breton SKRIVAÑ "write"

	Indicative				Irrealis		Imperative
	Present	Imperfect	Future	Past	Present	Past	
1SG	*-an*	*-e-n*	*-i-n*	*-is*	*-f-e-n*	*-j-e-n*	
2SG	*-ez*	*-e-s*	*-i*	*-j-out*	*-f-e-s*	*-j-e-s*	–
3SG	–	*-e*	*-o*	*-as*	*-f-e*	*-j-e*	*-e-t*
1PL	*-o-m*	*-e-m*	*-i-m*	*-j-o-m*	*-f-e-m*	*-j-e-m*	*-o-m*
2PL	*-i-t*	*-e-c'h*	*-o-t*	*-j-o-c'h*	*-f-e-c'h*	*-j-e-c'h*	*-i-t*
3PL	*-o-nt*	*-e-nt*	*-i-nt*	*-j-o-nt*	*-f-e-nt*	*-j-e-nt*	*-e-nt*
IMPS	*-e-r*	*-e-d*	*-o-r*	*-j-o-d*	*-f-e-d*	*-j-e-d*	

(2) Breton conjugational suffixes
  a. Initial suffixes: *-j, -f*
  b. Vocalic suffixes: *-e, -i, -o*
  c. Terminations: *-n, -s, -m, -t, -c'h, -nt, -r, -d*
  d. (b)/(c) portmanteaus: *-out, -an, -ez, -is, -as*

Inspection of the suffix sequences in Table 2 reveals that certain suffixes consistently express the same content and that certain content is consistently expressed by the same suffixal morphology; these consistencies are listed in (3). The initial suffix *-f* always expresses the present irrealis, which never has any other exponent. The initial suffix *-j* always expresses the past tense; there are, however, two portmanteau suffixes (*-is* and *-as*) that express the past tense without *-j*. The terminations *-m* and *-nt* uniformly express 1PL and 3PL subject agreement (respectively). The terminations *-n* and *-s* always express 1SG and 2SG subject agreement (respectively), though this content is also expressed without *-n* or *-s* by certain portmanteau suffixes (by *-is* in the first-person singular past indicative and by *-ez* and *-out* in the second-person singular indicative of the present and past). The termination *-t* is ambiguous, expressing 2PL subject agreement in most cases but 3SG in one case; and the termination *-c'h* is an alternative expression of 2PL subject agreement.[2] Impersonal subject agreement ('one writes, on écrit') is expressed by the termination *-r* in some cases and by the termination *-d* in others. Each of the portmanteau suffixes in (3c) is the sole expression of its content.

(3) Relations of form to content among initial, terminal and portmanteau suffixes in Breton conjugation.
[N.B.: *a* → *C* : affix *a* always expresses content *C*.
a ↔ *C* : *a* always expresses *C* and *C* is always expressed by *a*.
a ← *C* : C is always expressed by *a*.]

a.	Initial suffixes:	*-j*	→	PST
		-f	↔	IRR PRS
b.	Terminations:	*-n*	→	1SG
		-s	→	2SG
		-m	↔	1PL
		-t	→	2PL or 3SG
		-c'h	→	2PL
		-nt	↔	3PL
		-r	→	IMPS
		-d	→	IMPS
c.	Portmanteau suffixes:	*-out*	↔	IND PST 2SG
		-an	↔	IND PRS 1SG
		-ez	↔	IND PRS 2SG
		-is	↔	IND PST 1SG
		-as	↔	IND PST 3SG

Although the association of form with content is reasonably straightforward for the initial suffixes, the terminations, and the portmanteau suffixes in (2a, c, d), the same cannot be said for the vocalic suffixes in (2b). None of the vocalic suffixes expresses a constant meaning in all of its uses. The suffix *-i* appears sporadically in both the present and the future of the indicative; the suffix *-o* appears sporadically in those same tense/mood combinations as well as in the past indicative. (Both *-i* and *-o* also appear in the imperative, though this is presumably an effect of syncretism with the present indicative.) The suffix *-e* appears in all forms of the imperfect and in both irrealis tenses, and also sometimes appears in the present indicative and the

2. In western areas of Brittany, *-c'h* innovatively replaces *-t* in the future tense, yielding forms such as *skrivoc'h* "you (PL) will write" (Favereau 1997: 196; Trépos n.d.: § 509); here, my analysis focuses on the more conservative variety described by Trépos (n.d.) and Kervella (1976).

imperative (though the third-person plural imperative form presumably exhibits syncretism with the imperfect). As a consequence of these facts, the morphosyntactic content expressed by words containing these vocalic suffixes can't be seen as a simple summing up of content that is attributable to their suffixes in isolation from one another; instead, their content must be partially attributed to conditional exponence. For example, in view of the versatility reflected in (4), the suffix *-o* has no fixed content, but is nevertheless an unambiguous exponent of future tense when it appears in the combinations *-o-t* and *-o-r*. That is, the future-tense exponence of *-o* is conditional in *skriv-o-t* "you (PL) will write" and *skriv-o-r* "one will write"; in other contexts, it is a conditional exponent of other properties. Across the paradigm in Table 1 taken as a whole, the semantic function of *-o* might therefore be said to be discriminative rather than strictly compositional.

(4) Vocalic suffixes:
-i IND PRS (with 2PL syncretized in the imperative)
IND FUT
-o IND PRS (with 1PL syncretized in the imperative)
IND FUT
IND PST
-e ← IND IMPF (with 3PL syncretized in the imperative)
← IRR (PRS and PST)
← IMPS IND PRS
← 3SG IMP

Here, I argue that the vocalic suffixes in (4) (as well as the termination *-t* in (3b)) are genuine instances of conditional exponence, whose incidence is an effect of the phenomenon of rule combination (Stump 2017a–c; 2019a, b, 2020; to appear).

2. Conditional exponence in Breton verb inflection

The inflection of Breton verbs presents several clear examples of conditional exponence. As is indicated in (4), the suffix *-i* seems to act as an exponent of two different property sets. It expresses the property set {IND PRS} when it appears with the second-person plural suffix *-t*, but it expresses {IND FUT} when it appears with the first-person singular suffix *-n*, the first-person plural suffix *-m*, and the third-person plural suffix *-nt*; it also expresses the future tense in the second-person singular, even though there is no accompanying personal marker. (It is reasonable to assume that in Breton, a regular verb's plural imperative forms are systematically syncretic, so that *-i* serves to signal the imperative in *skriv-i-t* "you (PL) write!" purely as an effect of syncretism with the indicative present in the second person plural of the imperative.)

The suffix *-o* seems to act as an exponent of three different property sets: {IND PRS}, {IND FUT}, and {IND PST}. In any verb form whose maximal suffix sequence is *-o-m* or *-o-nt*, it functions as an exponent of the indicative present, and in any form whose maximal suffix sequence is *-o, -o-t*, or *-o-r*, it functions as an exponent of future tense. In any form in whose sequence it is preceded by *-j*, it functions as an exponent of the indicative past; this includes forms having the suffix sequence *-j-o-m, -j-o-c'h, -j-o-nt*, or *-j-o-d*. (On the assumption that a regular verb's plural imperative forms are systematically syncretic, *-o* serves to signal the imperative in *skriv-o-m* "let's write!" purely as an effect of syncretism with the present indicative in the first person plural of the imperative.)

The suffix *-e* seems to act as the default vocalic suffix in three tenses (the imperfect and the two irrealis tenses), but it also expresses {IND PRS} in combination with the impersonal suffix *-r* and {imperative} in the third-person singular form in *-t*. (On the assumption that a regular verb's plural imperative forms are systematically syncretic, *-e* expresses the imperative in *skriv-e-nt* "let them write!" purely as an effect of syncretism with the imperfect in the third person plural of the imperative.)

One could, of course, question the apparent variability of the exponence relations in which *-i, -o* and *-e* participate. In particular, one could argue that the inflection of Breton SKRIVAÑ "write" involves

- two different *-i* suffixes (one for the indicative present, the other for the future),
- three different *-o* suffixes (for the indicative present, future, and indicative past), and
- four different *-e* suffixes (for the imperfect, the irrealis, the indicative present impersonal, and the nonsyncretic 3SG imperative).

Under this analysis, however, each set of homophonous suffixes would be distributionally mysterious. In order to see this mystery, consider first some uncontroversial cases of affix homophony observable in Swahili.

The Swahili sentences in (5) exemplify two homophonous prefixes having the form *li-*, one of which expresses noun-class concord with an argument of class 5 and the other of which expresses past tense. As their joint occurrence in (5a) shows, they occupy distinct prefix positions; in addition, the past-tense prefix *li-* is overridden by *ku-* in negative contexts, as in (5b), while the class 5 concord participates in no comparable competition.

(5) a. *Ji-no* *li-li-toka.*
CLASS.5-tooth SBJ.CLASS.5-PST-come.out
"The tooth came out."

b. *Ji-no* *ha-li-ku-toka.*
CLASS.5-tooth NEG-SBJ.CLASS.5-NEG.PST-come.out
"The tooth didn't come out."

The Swahili sentences in (6) exemplify two homophonous prefixes having the form *u*-, one of which expresses noun-class concord with an argument of class 3 (as in (6a)), and the other of which expresses concord with a second-person singular subject. These prefixes may appear in exactly the same morphological context, in which case they contrast. Sentence (6b) is therefore ambiguous; it can mean either "you (SG.) fell down" or "it fell down" (where *it* denotes the referent of a class 3 noun such as *mti* "tree"). Could it be that this is not actually an ambiguity at all – that second-person singular arguments in fact belong to class 3 in Swahili? The answer is no: class 3 object concord is also expressible by *u*- as in (7a), while second-person singular object concord is expressed by a distinct prefix *ku*-, as in (7b).

(6) a. *M-ti* *u-li-anguka*.
CLASS.3-tree SBJ.CLASS.3-PST-fall.down
"The tree fell down."
b. *U-li-anguka*.
SBJ.2SG-PST-fall.down / SBJ.CLASS.3-PST-fall.down
"You (SG) fell down." / "It fell down." (it = the tree)

(7) a. *Ni-li-u-ona*.
SBJ.1SG-PST-OBJ.CLASS.3-see
"I saw it (= tree)."
b. *Ni-li-ku-ona*.
SBJ.1SG-PST-OBJ.2SG-see
"I saw you (SG)."

The Swahili sentences in (5b) and (7b) further exemplify two homophonous prefixes having the form *ku*-; these occupy distinct prefix positions, as their joint occurrence in (8) shows.

(8) *Ha-wa-ku-ku-ona*.
NEG-SBJ.3PL-NEG.PST-OBJ.2SG-see
"They didn't see you (SG)."

These Swahili examples highlight the fact that two homophonous affixes Af_1, Af_2 are distinguishable if Af_1 and Af_2 express distinct content and appear in distinct affix positions (like *li*- in (5a) and *ku*- in (8)); they are also distinguishable if they contrast in exactly the same morphological context (like *u*- in (6b)). But if Breton *-i, -o* and *-e* in fact reflect three sets of homophonous suffixes – an *-i* set, an *-o* set and an *-e* set – then the members of these sets are unlike Swahili *li*- and *ku*-, since they cannot be associated with distinct affix positions: all are members of the syntagmatically intermediate class of vocalic suffixes. In addition, the members of each set are also unlike Swahili *u*-: although they appear in the same affix position, they never contrast in the same morphological context. For instance, although the members of the set of homophonous -*o* suffixes would be restricted to the suffixal contexts in Table 3, no two of them would ever contrast in any of those contexts.

Table 3. Morphological contexts of -*o* in Breton regular verb inflection

	Indicative		
	Present	**Future**	**Past**
3SG		Stem-__*#*	
1PL	Stem-__-*m*		Stem-*j*-__-*m*
2PL		Stem-__-*t*	Stem-*j*-__-*c'h*
3PL	Stem-__-*nt*		Stem-*j*-__-*nt*
IMPS		Stem-__-*r*	Stem-*j*-__-*d*

I reject this appeal to multiple homophony and conclude that the content expressed by the vocalic suffixes is conditioned by the initial suffixes in (2a) and the terminations in (2c). I argue that these dependencies are an effect of the phenomenon of rule combination.[3] Before proceeding with this argument, I elaborate on the notion of rule combination itself.

3. What is rule combination?

Stump (2017a–c; 2019a, b; 2020; to appear) argues that rules of exponence that are successive in their application sometimes combine to form a complex rule of exponence. The simplest manner of combination is that of rule composition. In order to streamline the following discussion, I use the notation [σ : -*x*] to represent a rule of exponence such that for any stem Z and any property set ρ,

i. [σ : -*x*] is applicable to ⟨Z, ρ⟩ only if σ is a subset of ρ;
ii. the application of [σ : -*x*] to ⟨Z, ρ⟩ realizes the property set σ; and
iii. the result of applying [σ : -*x*] to ⟨Z, ρ⟩ is the pairing ⟨Z-*x*, ρ⟩.

3. Although my focus here is on the implications of conditional exponence for inferential-realizational approaches to inflectional morphology, the facts on which my discussion is based have parallel implications for morpheme-based (i.e., lexical-incremental and lexical-realizational) approaches to inflectional morphology. Specifically, these facts show that a combination Morpheme$_1$ + Morpheme$_2$ may express content that is not attributable to either Morpheme$_1$ or Morpheme$_2$ on its own. A morpheme-based theory would therefore need to be enhanced by a notion of supplementational morpheme combination comparable to the notion of supplementational rule combination for which I argue below. But this enhancement would not, in itself, remedy the other conceptual difficulties faced by morpheme-based approaches; for discussion of these difficulties, see, e.g., Matthews 1972, Anderson 1992, Aronoff 1994, Stump 2001, 2016.

Example:

i. [{PL} : -*s*] is applicable to any pairing ⟨Z, ρ⟩ such that {PL} is a subset of ρ;
ii. the application of [{PL} : -*s*] to ⟨*dog*, {PL}⟩ realizes {PL};
iii. the result of applying [{PL} : -*s*] to ⟨*dog*, {PL}⟩ is ⟨*dog-s*, {PL}⟩.

With this notation, rule composition may be defined as in (9), where '$\sigma \cup \tau$' represents the union of the sets σ and τ.

(9) Given two rules [σ : -*x*], [τ : -*y*], the composition ([τ : -*y*] ∘ [σ : -*x*]) is the complex rule [$\sigma \cup \tau$: -*x-y*].

The evidence in favor of incorporating the notion of rule composition into morphological theory is quite varied; consider three kinds of evidence of this sort.

Asymmetrical rule oppositions

Ordinarily, morphological systems conform to a 'symmetry principle' according to which relations of paradigmatic opposition among rules of exponence are relations between individual rules. Yet, there are instances in which the application of an individual rule seems to be paradigmatically opposed to the successive application of two rules: in the definition of the Latin verb forms in (10), the application of the rule [{2PL PASS} : -*minī*] is paradigmatically opposed to the successive application of the rule [{3PL} : -*nt*] and the rule [{PASS} : -*ur*]. This apparent asymmetry is reconcilable with the symmetry principle if one assumes that [{PASS} : -*ur*] composes with [{3PL} : -*nt*] to form a complex rule [{3PL PASS} : -*nt-ur*] and that it is the application of this composed rule that is opposed to that of the simple [{2PL PASS} : -*minī*] rule.

(10) a. *amā-minī*
love-2PL.PASSIVE
"you (PL) are loved"
b. *ama-nt-ur*
love-3PL-PASSIVE
"they are loved"

Dependent rules

Ordinarily, the application of one rule of exponence is not directly conditioned by that of another. Yet, there are instances in which a rule's application seems to depend directly on that of a 'carrier' rule. (See Harris 2017 for detailed discussion of the properties of dependent and carrier affixes.) In Limbu [Kiranti; Nepal], the rule [{[1SG AGT]} : -*ŋ*] is dependent in this way: it must always piggy-back on an appropriate carrier rule. The carrier rules are those filling suffix positions 4 and 8 in a verb's inflectional morphology; in the absence of a carrier rule, the 1SG agent property remains unrealized by [{[1SG AGT]} : -*ŋ*]. The forms of the Limbu verb HUʔMAʔ "teach" in Table 4 illustrate. In the definition of the form in row (a), the carrier rule on which the -*ŋ* rule is dependent fills affix position 4; in the definition of the forms in rows (b) and (c), the carrier rules fill position 8; in row (d), there are two carrier rules (one filling position 4 and another filling position 8) hence two applications of the -*ŋ* rule; and in row (e), there is no carrier rule, hence the -*ŋ* rule fails to apply at all. The relation between a dependent rule and its carrier can be formally represented as a relation of rule composition; on this view, a dependent rule such as [{[1SG AGT]} : -*ŋ*] is a rule that never applies except in composition with a carrier rule, e.g., those in Table 5.

Table 4. Positive nonpreterite forms of Limbu HUʔMAʔ with 1SG agents

	AGT → PAT	Stem	Suffix position					
			1	4	5	8	9	
a.	1SG → 3SG	*huʔr*		*-u*	*-ŋ*			"I teach her/him"
b.	1SG → 2du	*huʔ*	*-nɛ*			*-ci*[1]	*-ŋ*	"I teach you (DU)"
c.	1SG → 2PL	*huʔ*	*-n(ɛ)*			*-i*	*-ŋ*	"I teach you (PL)"
d.	1SG → 3 -SG	*huʔr*		*-u*	*-ŋ*	*-si*	*-ŋ*	"I teach them"
e.	1SG → 2SG	*huʔ*	*-nɛ*					"I teach you (SG)"

1. alternant of -*si* (van Driem 1987: 77)

Source:
van Driem (1987: 368–369)

Table 5. Compositions of the dependent rule [{[1SG AGT]} : -*ŋ*] in Limbu

Rule block	Carrier rule	Composition of [{[1SG AGT]} : -*ŋ*] with carrier rule
4	[{[3 PAT]} : -*u*]	[{[3 PAT],[1SG AGT]} : -*u*-*ŋ*]
8	[{[−3 -INCL PL PAT]} : -*i*]	[{[−3 -INCL PL PAT],[1SG AGT]} : -*i*-*ŋ*]
8	[{[-SG PAT]} : -*si*]	[{[-SG PAT],[1SG AGT]} : -*si*-*ŋ*]

Processing frequent affix sequences

The null hypothesis is that affix sequences of equal length require the same processing time. Yet, Bilgin (2016) shows that, controlling for frequency differences among individual affixes, stems, and affix-stem combinations, affix sequences that are frequent are processed more rapidly than those that are less frequent. This suggests that if the joint application of successive rules is frequent enough, their composition may be stored in memory. (O'Donnell (2015) proposes a similar idea in a different context.)

These facts are a part of the evidence that provides a sound basis for incorporating a principle of rule composition into morphological theory. (For additional supporting evidence, see Stump 2019a, b; 2020; to appear.) With this principle in mind, I now return to the problem of providing a formal conception of conditional exponence.

4. Conditional exponence as a kind of rule combination

Ordinarily, the content realized by a sequence of rule applications is derivable from the content realized by the individual rules in that sequence. Yet, some rule sequences seem to realize a more specific content than can be deduced from their members. Neither rule of affixation that applies in the realization of the Breton form *skriv-o-r* "one will write" unambiguously realizes the future tense (i.e., both [{} : -*o*] and [{IMPS} : -*r*] apply outside the future, as, e.g., in the present indicative forms *skriv-o-m* "we write", *skriv-e-r* "one writes"), yet *skrivor* is itself unambiguously future-tense. This suggests that more specific content may be associated with a combined rule than is derivable from its component rules – that the future tense of *skrivor* is realized neither by [{} : -*o*] nor by [{IMPS} : -*r*], but by the combination of [{IMPS} : -*r*] with [{} : -*o*]. There is, however, a conundrum here.

Definition (9) defines one way of combining rule [τ : -*y*] with rule [σ : -*x*]: as the composition of [τ : -*y*] with [σ : -*x*]. But if the combination of [{IMPS} : -*r*] with [{} : -*o*] differs from both rule [{IMPS} : -*r*] and rule [{} : -*o*] in expressing future tense, then this combination cannot simply be equated with the composition of those two rules. In general, the phenomenon of conditional exponence comprises instances in which part of the content realized by the combination of two rules derives from neither of those rules, being instead associated with the combination as a matter of stipulation. This suggests that composition is only one of the ways in which two rules might be combined.[4]

The consequence of this conclusion is that we must distinguish between two kinds of rule combination. In the default case, rule combinations constitute ordinary instances of rule composition, as defined in (9); rule combinations of this sort

---

4. Rule composition, as it is characterized here, might be seen as the notational equivalent of schema unification in Construction Morphology (Booij 2010: 42–50; Kempf & Hartmann 2018); if so, then the evidence presented here shows that schema unification cannot be the only means of combining schemas in Construction Morphology.

possess the property of realizational compositionality in (11a). Some rule combinations, however, do not amount to simple rule compositions, but are instead supplementational rule combinations (SupCs) of the sort defined in (12). I shall represent the α-supplementational rule combination of [τ : -*y*] with [σ : -*x*] as

$$([\tau : \text{-}y] \; Ⓢ_{\alpha} \; [\sigma : \text{-}x]),$$

and I shall refer to the property set α in this SupC as its ADDEND. Unlike ordinary rule compositions, SupCs possess the property of realizational discriminativity in (11b).

(11) Given two rules [σ : -*x*], [τ : -*y*],
  a. a combination of [τ : -*y*] with [σ : -*x*] is REALIZATIONALLY COMPOSITIONAL if it realizes the property set σ ∪ τ.
  b. a combination of [τ : -*y*] with [σ : -*x*] is REALIZATIONALLY DISCRIMINATIVE if it realizes a proper superset of σ ∪ τ.

(12) Given two rules [σ : -*x*], [τ : -*y*] and some property set α, [α ∪ σ ∪ τ: -*x*-*y*] is the α-SUPPLEMENTATIONAL COMBINATION of [τ : -*y*] with [σ : -*x*].

Thus, consider again the case of *skrivor* "one will write". In this case, the combination of [{IMPS} : -*r*] with [{} : -*o*] is not the simple composition ([{IMPS} : -*r*] ∘ [{} : -*o*]) (= the rule [{IMPS} : -*o*-*r*]), but is instead the SupC ([{IMPS} : -*r*] $Ⓢ_{\{\text{IND FUT}\}}$ [{} : -*o*]) (= the rule [{IND FUT IMPS} : -*o*-*r*]), whose addend is {IND FUT}. The definition of this SupC has the effect of making -*o* an exponent of future tense in the context of the impersonal termination -*r*.

The same reasoning affords an account of all of the instances of conditional exponence observed in Table 1, as in the analysis that I now briefly sketch.

5. A formal analysis of conditional exponence Breton verb inflection

In accordance with the discussion in § 1, I assume that the rules of exponence that define a Breton regular verb's inflection fall into four groups: the initial-suffix group (13a, b), the vocalic-suffix group (13c–e), the termination group (13f–n), and the portmanteau-suffix group (13o–s). I assume that, like the [{[1SG AGT]} : -*ŋ*] rule in Limbu (§ 3.2), the termination rules in (13f–n) are all dependent, and that the carrier rules for these dependent rules are the vocalic-suffix rules in (13c–e) and certain combinations of vocalic-suffix rules with the initial-suffix rules in (13a, b).

(13) Initial-suffix rules:
a. [{PST} : *-j*]
b. [{IRR PRS} : *-f*]

Vocalic-suffix rules:
c. [{} : *-i*]
d. [{} : *-o*]
e. [{} : *-e*]

Termination rules:
f. [{1SG} : *-n*]
g. [{2SG} : *-s*]
h. [{1PL} : *-m*]
i. [{} : *-t*]
j. [{2PL} : *-c'h*]
k. [{3PL} : *-nt*]
l. [{IMPS} : *-r*]
m. [{IMPS} : *-d*]
n. [{} : –] (default identity function)

Portmanteau-suffix rules:
o. [{2SG} : *-out*]
p. [{IND PRS 1SG} : *-an*]
q. [{IND PRS 2SG} : *-ez*]
r. [{IND PST 1SG} : *-is*]
s. [{IND PST 3SG} : *-as*]

The SupCs into which the rules in (13c–n) enter can be distinguished according to the addends that they involve. The SupCs in (14) have {IND PRS} as their addend; those in (15) have {IND FUT} as their addend; and those in (16) have {IMPF} as their addend. The SupCs in (14)–(16) are alike in that their addends only involve properties of tense and sometimes mood. Rules (13i) and (13n), however, are special in that the SupCs into which they enter have addends that include person/number properties, as in (17).

(14) {IND PRS}-supplemental combinations

Rule label	Definition	Extensional definition
SupC[14a]:	((13h) Ⓢ$_{\{\text{IND PRS}\}}$ (13d))	= [{IND PRS 1PL} : *-o-m*]
SupC[14b]:	((13k) Ⓢ$_{\{\text{IND PRS}\}}$ (13d))	= [{IND PRS 3PL} : *-o-nt*]
SupC[14c]:	((13l) Ⓢ$_{\{\text{IND PRS}\}}$ (13e))	= [{IND PRS IMPS} : *-e-r*]

(15) {IND FUT}-supplemental combinations

SupC[15a]:	((13f) Ⓢ$_{\{\text{IND FUT}\}}$ (13c))	= [{IND FUT 1SG} : *-i-n*]
SupC[15b]:	((13h) Ⓢ$_{\{\text{IND FUT}\}}$ (13c))	= [{IND FUT 1PL} : *-i-m*]
SupC[15c]:	((13k) Ⓢ$_{\{\text{IND FUT}\}}$ (13c))	= [{IND FUT 3PL} : *-i-nt*]
SupC[15d]:	((13l) Ⓢ$_{\{\text{IND FUT}\}}$ (13d))	= [{IND FUT IMPS} : *-o-r*]

(16) {IMPF}-supplemental combinations
SupC[16a]: ((13f) Ⓢ$_{\{\text{IMPF}\}}$ (13e)) = [{IMPF 1SG} : *-e-n*]
SupC[16b]: ((13g) Ⓢ$_{\{\text{IMPF}\}}$ (13e)) = [{IMPF 2SG} : *-e-s*]
SupC[16c]: ((13h) Ⓢ$_{\{\text{IMPF}\}}$ (13e)) = [{IMPF 1PL} : *-e-m*]
SupC[16d]: ((13j) Ⓢ$_{\{\text{IMPF}\}}$ (13e)) = [{IMPF 2PL} : *-e-c'h*]
SupC[16e]: ((13k) Ⓢ$_{\{\text{IMPF}\}}$ (13e)) = [{IMPF 3PL} : *-e-nt*]
SupC[16f]: ((13m) Ⓢ$_{\{\text{IMPF}\}}$ (13e)) = [{IMPF IMPS} : *-e-d*]

(17) SupCs whose addends include person/number properties
SupC[17a]: ((13i) Ⓢ$_{\{\text{IND PRS 2PL}\}}$ (13c)) = [{IND PRS 2PL} : *-i-t*]
SupC[17b]: ((13n) Ⓢ$_{\{\text{IND FUT 2SG}\}}$ (13c)) = [{IND FUT 2SG} : *-i*]
SupC[17c]: ((13i) Ⓢ$_{\{\text{IND FUT 2PL}\}}$ (13d)) = [{IND FUT 2PL} : *-o-t*]
SupC[17d]: ((13n) Ⓢ$_{\{\text{IND FUT 3SG}\}}$ (13d)) = [{IND FUT 3SG} : *-o*]
SupC[17e]: ((13i) Ⓢ$_{\{\text{IMP 3SG}\}}$ (13e)) = [{IMP 3SG} : *-e-t*]
SupC[17f]: ((13n) Ⓢ$_{\{\text{IMPF 3SG}\}}$ (13e)) = [{IMPF 3SG} : *-e*]

The vocalic-suffix rules (13d, e) combine with the initial-suffix rules (13a, b) as in (18). The combined rules in (18) in turn serve as carriers for various termination rules, combining with them to form the sixteen default rule compositions (DRCs) in (19)–(21). SupC[18a] serves as the carrier in the default composition rules in (19); the default rule composition DRC[18b] serves as the carrier in the default composition rules in (20); and SupC[18c] serves as the carrier in the default composition rules in (21).

(18) SupC[18a]: ((13d) Ⓢ$_{\{\text{IND}\}}$ (13a)) = [{IND PST} : *-j-o*]
DRC[18b]: ((13e) ∘ (13b)) = [{IRR PRS} : *-f-e*]
SupC[18c]: ((13e) Ⓢ$_{\{\text{IRR}\}}$ (13a)) = [{IRR PST} : *-j-e*]

(19) Default rule compositions based on carrier rule SupC[18a]

Rule label	Definition	Extensional definition
DRC[19a]:	((13h) ∘ SupC[18a])	= [{IND PST 1PL}: *-j-o-m*]
DRC[19b]:	((13j) ∘ SupC[18a])	= [{IND PST 2PL}: *-j-o-c'h*]
DRC[19c]:	((13k) ∘ SupC[18a])	= [{IND PST 3PL}: *-j-o-nt*]
DRC[19d]:	((13m) ∘ SupC[18a])	= [{IND PST IMPS}: *-j-o-d*]

(20) Default rule compositions based on carrier rule DRC[18b]
DRC[20a]: ((13f) ∘ DRC[18b]) = [{IRR PRS 1SG}: *-f-e-n*]
DRC[20b]: ((13g) ∘ DRC[18b]) = [{IRR PRS 2SG}: *-f-e-s*]
DRC[20c]: ((13h) ∘ DRC[18b]) = [{IRR PRS 1PL}: *-f-e-m*]
DRC[20d]: ((13j) ∘ DRC[18b]) = [{IRR PRS 2PL}: *-f-e-c'h*]
DRC[20e]: ((13k) ∘ DRC[18b]) = [{IRR PRS 3PL}: *-f-e-nt*]
DRC[20f]: ((13m) ∘ DRC[18b]) = [{IRR PRS IMPS}: *-f-e-d*]

(21) Default rule compositions based on carrier rule SupC[18c]
DRC[21a]: ((13f) ∘ SupC[18c]) = [{IRR PST 1SG}: *-j-e-n*]
DRC[21b]: ((13g) ∘ SupC[18c]) = [{IRR PST 2SG}: *-j-e-s*]
DRC[21c]: ((13h) ∘ SupC[18c]) = [{IRR PST 1PL}: *-j-e-m*]
DRC[21d]: ((13j) ∘ SupC[18c]) = [{IRR PST 2PL}: *-j-e-c'h*]
DRC[21e]: ((13k) ∘ SupC[18c]) = [{IRR PST 3PL}: *-j-e-nt*]
DRC[21f]: ((13m) ∘ SupC[18c]) = [{IRR PST IMPS}: *-j-e-d*]

Rule (13a) serves as a carrier for the portmanteau-suffix rule (13o), combining with it to form the SupC in (22). No rule combination is needed to define the realization of the property sets {IND PRS 1SG}, {IND PRS 2SG}, {IND PST 1SG}, and {IND PST 3SG}, since the realization of these property sets is specified by the portmanteau rules (13p–s).

(22) SupC[22]: ((13o) Ⓢ$_{\{\text{IND}\}}$ (13a)) = [{IND PST 2SG} : *-j-out*])

There are five forms in Table 1 that the rules in (13)–(22) do not directly account for: these are the three instances of syncretism in the plural imperative paradigm of SKRIVAÑ and the suffixless realizations of the property sets {IND PRS 3SG} and {IMP 2SG}. As noted above, I assume that the lack of a dedicated rule of exponence is a systematic property of the plural imperative; I therefore attribute the plural imperative syncretisms to the default stipulations in (23). I attribute the lack of suffixal morphology in *skriv* "s/he writes" and *skriv* "you (SG) write!" to the stipulation in (24); this precludes the default appearance of any of the three vocalic suffixes in these forms.

(23) For any verbal lexeme L,
a. if PF(⟨L, {IND PRS 1PL}⟩) = ⟨*w*, {IND PRS 1PL}⟩,
then PF(⟨L, {IMP 1PL}⟩) = ⟨*w*, {IMP 1PL}⟩
b. if PF(⟨L, {IND PRS 2PL}⟩) = ⟨*w*, {IND PRS 2PL}⟩,
then PF(⟨L, {IMP 2PL}⟩) = ⟨*w*, {IMP 2PL}⟩
c. if PF(⟨L, {IND IMPF 3PL}⟩) = ⟨*w*, {IND IMPF 3PL}⟩,
then PF(⟨L, {IMP 3PL}⟩) = ⟨*w*, {IMP 3PL}⟩

(24) Where Z is the bare stem of a verbal lexeme L and σ is {IND PRS 3SG} or {IMP 2SG}, PF(⟨L, σ⟩) = ⟨Z, σ⟩.

In (23) and (24), PF represents the Breton paradigm function. In Paradigm Function Morphology (Stump 2001, 2016), a language's paradigm function is a function that applies to a paradigm cell[5] to yield the realization of that cell; the Breton paradigm

5. Stump 2016 argues that there are two kinds of inflection paradigms: each cell in a content paradigm determines the syntax and semantics of a particular form, while the corresponding cell in the corresponding form paradigm determines the morphological realization of that form. I assume that a language's paradigm function may apply either to a content cell or to a form cell to yield its realization.

function PF, for example, applies to the paradigm cell ⟨SKRIVAÑ, {IND PRS 3PL}⟩ to yield its realization ⟨*skrivont*, {IND PRS 3PL}⟩. The rules in (23) cause the default realizations of a verbal paradigm's plural imperative cells to have the same form as those of certain other cells in the paradigm; the stipulation in (24) causes the default realization of a verbal paradigm's third-person singular present indicative and second-person singular imperative cells to take the form of the verb's bare stem.

The analysis sketched here provides an explicit account of the definition of each form in the paradigm of SKRIVAÑ "write" in Table 1. Each cell in Table 6 specifies the rule defining the inflectional morphology of the corresponding form in Table 1. In formal terms: for each cell ⟨SKRIVAÑ, σ⟩ in the paradigm of SKRIVAÑ, PF(⟨SKRIVAÑ, σ⟩) is the result of applying the corresponding rule in Table 6 to ⟨*skriv*-, σ⟩. In more general terms: the value of PF(⟨SKRIVAÑ, σ⟩) conforms to (23) and (24) but otherwise defaults to the result of applying the narrowest applicable rule of exponence to ⟨*skriv*-, σ⟩; in some instances, the narrowest applicable rule of exponence is simply one of the rules in (13); in most instances, however, it is either a supplementational rule combination or a default rule composition.

In the accompanying appendix, I discuss extensions of this analysis to a wider range of Breton data.

Table 6. Rules defining the finite forms of Breton SKRIVAÑ

a. Indicative

	Present	**Imperfect**	**Future**	**Past**
1SG	(13p)	SupC[16a]	SupC[15a]	(13r)
2SG	(13q)	SupC[16b]	SupC[17b]	SupC[22]
3SG	(24)	SupC[17f]	SupC[17d]	(13s)
1PL	SupC[14a]	SupC[16c]	SupC[15b]	DRC[19a]
2PL	SupC[17a]	SupC[16d]	SupC[17c]	DRC[19b]
3PL	SupC[14b]	SupC[16e]	SupC[15c]	DRC[19c]
IMPS	SupC[14c]	SupC[16f]	SupC[15d]	DRC[19d]

b. Irrealis			**c. Imperative**
	Present	**Past**	
1SG	DRC[20a]	DRC[21a]	
2SG	DRC[20b]	DRC[21b]	(24)
3SG	DRC[18b]	SupC[18c]	SupC[17e]
1PL	DRC[20c]	DRC[21c]	(23a)
2PL	DRC[20d]	DRC[21d]	(23b)
3PL	DRC[20e]	DRC[21e]	(23c)
IMPS	DRC[20f]	DRC[21f]	

6. Conclusions

The phenomenon of conditional exponence evidenced by Breton verb inflection provides additional motivation for the theoretical assumption that rules of morphology may combine to form more complex rules of morphology. This assumption has important implications in the domains of morphological theory, of morphological typology, of language processing, and of language change.

In inferential-realizational theories of morphology, it has been customary to assume that a language's rules of exponence fall into blocks such that (i) the rules in a given block are mutually exclusive (or disjunctive) in their application and (ii) the rules' sequence of application in the definition of a given word form reflects a fixed ordering of the blocks to which they belong. This way of formalizing a system of rules of exponence captures the traditional notion of affix position classes: affixes belonging to the same position class are introduced by rules belonging to the same block, and the ordering of affix positions reflects the ordering of the corresponding blocks.

Given these assumptions, the definition of a language's paradigm function has been assumed to be what determines the sequence of rule blocks in that language. For instance, the definition of the Swahili paradigm function PF might be assumed to include the clause in (25) for verb inflection: this clause accounts for the sequence of four rule blocks (I through IV) relevant for the inflection of a form such as (26). In the absence of rule combination, the narrowest applicable rules of exponence relevant for the definition of the form in (26) are those in (27).

(25) PF(⟨Z, σ⟩) = [IV – [III – [II – [I – ⟨Z, σ⟩]]]],
where [*n* – ⟨Z, σ⟩] is the result of applying to ⟨Z, σ⟩ the narrowest applicable rule in Block *n*.

(26)

ha-	*wa-*	*ta-*	*ni-*	*piga*
NEG-	SBJ.3PL-	FUT-	OBJ.1SG-	beat
IV	III	II	I	

"they will not beat me"

(27)
a. [{OBJ:{1SG}} : *ni*-] (∈ Block I)
b. [{TNS:FUT} : *ta*-] (∈ Block II)
c. [{SBJ:{3PL}} : *wa*-] (∈ Block III)
d. [{POL:neg} : *ha*-] (∈ Block IV)

The notion of rule combination advocated here raises the possibility of defining every inflected word form *w* as the result of applying a single, combined rule of exponence to the pairing ⟨Z, ρ⟩ of *w*'s stem Z with the property set ρ that *w* realizes; this will, in general, be the single most narrowly applicable rule of exponence available for the realization of ⟨Z, ρ⟩. Consider again the word form *hawatanipiga* in (26); by assumption, this is the result of applying the Swahili paradigm function

PF to the pairing in (28). Rather than define (the relevant clause of) this paradigm function as in (25), we can define it more simply as in (29). On this assumption, rules such as those in (27) combine recursively to form more specific rules, and as a result of such combinations, the narrowest rule of exponence applicable to (28) is the combined rule of exponence in (30a) (whose extensional equivalent is (30b)). In this new approach, the order of application of the members of some set of rules of exponence is not directly determined by defining a language's paradigm function in terms of ordered rule blocks; rather, it is determined by the definition of the combinations into which the members of that set enter.

(28) ⟨ *-piga*, {POL:neg, SBJ:{3PL}, TNS:FUT, OBJ:{1SG}} ⟩

(29) PF(⟨Z, σ⟩) is the result of applying to ⟨Z, σ⟩ the narrowest applicable rule of exponence.

(30) a. ((27d) ∘ ((27c) ∘ ((27b) ∘ (27a))))
b. [{POL:neg, SBJ:{3PL}, TNS:FUT, OBJ:{1SG}} : *ha-wa-ta-ni-*]

The notion of supplementational rule combinations is of typological interest because it suggests an explicit measure of the opacity of a paradigm's exponence relations, namely as the ratio of the number of paradigm cells whose exponence is defined by means of SupCs to the total number of cells. In a paradigm whose cells are fully realizable by compositional means, there is no motivation for SupCs; instead, all realizations are effected either by simple rules of exponence or by default composition, hence the opacity of the paradigm's exponence relations approximates zero. Clearly, the morphology of a Breton regular verb such as SKRIVAÑ doesn't come close to attaining this ideal of fully transparent exponence relations: because thirty-one of the forty-seven cells in Table 5 are defined by means of SupCs (including the nineteen SupCs in (14)–(17), the ten compositions based on SupC[18a] and SupC[18c] in (19) and (21), SupC[18c] itself, and SupC[22], the degree of opacity exhibited by its exponence relations (under the proposed analysis) is 31/47 (= 0.66).

The notion of rule combination also entails an interesting elaboration of the typology of morphomes. The vocalic suffixes *-i, -o* and *-e* are morphomic, but they differ from many of the morphomes discussed in the literature. In the most familiar cases, a morphome can be seen as a category to which a language's morphology is sensitive but to which its phonology, syntax and semantics are blind. Morphomic categories are in general of three kinds: lexeme categories, stem categories, and exponent categories.

- Morphomic lexeme categories distinguish classes of lexemes; in English, 'weak verb' is a category of this kind.
- Morphomic stem categories distinguish classes of stems that cross-cut classes of lexemes. In Sanskrit, 'strong stem', 'middle stem' and 'weakest stem' are

categories of this kind (Stump 2001: Chapter 6). Sanskrit nominals that differ in their declension-class membership are alike in that they use these three stems to realize parallel sets of cells in their paradigms; yet, those sets of cells are not associated with coherent sets of morphosyntactic properties, and the formation of the strong, middle and weakest stems varies from one subclass of nominals to another.

– Morphomic exponent categories distinguish sets of inflectional exponents that serve in the realization of word forms that cross-cut natural classes of morphosyntactic properties in the same way. In Hua [Trans-New-Guinea; Papua New Guinea], the 'C' suffixes are a category of this kind (Stump 2016: Chapter 8). Hua verbs inflect with suffixes that cumulatively express mood and person/number agreement. For each mood, there is a distinct 'C' suffix that marks both second-person singular forms and first-person plural forms; but these word forms are not syncretic, since they differ in their stem ablaut.

The vocalic suffixes *-i, -o* and *-e* might be seen as defining three morphomic exponent categories, where each category contains its defining suffix and all of the combined suffixes based on it as an effect of rule combination. The membership of the I, O and E categories is thus as in (31). These morphomic exponent categories are unlike the Hua 'C' category in an important way. In Hua, the 'C' category has a kind of internal consistency, in that each suffix in the 'C' category realizes one or another morphosyntactic property set instantiating the pattern in (32). The Breton categories in (31), by contrast, do not exhibit this degree of internal consistency; the members of the O category, for example, realize the miscellaneous morphosyntactic property sets in (33).

(31) I category: *-i-t, -i-n, -i, -i-m, -i-nt*
O category: *-o-m, -o-nt, -o, -o-t, -o-r, -j-o-m, -j-o-c'h, -j-o-nt, -j-o-d*
E category: *-e-r, -e-n, -e-s, -e, -e-m, -e-c'h, -e-nt, -e-d,*
-f-e-n, -f-e-s, -f-e, -f-e-m, -f-e-c'h, -f-e-nt, -f-e-d,
-j-e-n, -j-e-s, -j-e, -j-e-m, -j-e-c'h, -j-e-nt, -j-e-d, -e-t

(32) {MOOD: α, [AGR:{PER: 2, NUM: SG} or AGR:{PER: 1, NUM: PL}]},
where α ∈ {Indicative, Interrogative, Relative, Purposive, Concessive-expectant, Inconsequential, Medial/Coordinate, Medial/Subordinate, Exclamatory, Assertive, Counterfactual/Protasis, Counterfactual/Apodosis}

(33) Property sets realized by members of the O category

	{3SG FUT IND}		
{1PL PRS IND}		{1PL PST IND}	{1PL IMP}
	{2PL FUT IND}	{2PL PST IND}	
{3PL PRS IND}		{3PL PST IND}	
	{IMPS FUT IND}	{IMPS PST IND}	

The notion of rule combination also has probable relevance both to morphological processing and to language change. Psycholinguistic research (Teddiman 2012; Bilgin 2016) has shown that affixes that are equally frequent are processed more quickly in frequent combinations than in infrequent combinations, suggesting that frequent affix combinations may be stored as easily retrieved units. My proposals further suggest that these units are rule combinations and thus raise the possibility that the rich morphology of a word form such as *skriv-o-r* "one will write" isn't necessarily processed by reference to simple rules of exponence such as [{} : -*o*] and [{IMPS} : -*r*] (= (13d) and (13l) above), but may proceed purely by reference to rule combinations such as [{IND FUT IMPS} : -*o-r*] (= SupC[15d] above). To the extent that reference to a combination exceeds the frequency of reference to its component rules, the combination may, through time, supplant the component rules and come to be reanalyzed as a simple rule itself; this is the process of affix telescoping (Haspelmath 1995; Booij 2005: 273).

7. Appendix: Further implications of rule combination in Breton

The analysis of conditional exponence in Breton conjugation in § 5 focuses purely on regular verb inflection, as exemplified by the paradigm of the verb SKRIVAÑ "write". The Breton inflectional system does, however, present some interesting complications; here, I briefly sketch an approach to accommodating some of these within the proposed analysis.

Irregular verbs

Breton exhibits a small number of verbs exhibiting irregular inflectional patterns. I illustrate with the verb GOUZOUT "know", whose finite paradigm is given in Table 7. Two kinds of irregularity stand out in this paradigm. First, unlike SKRIVAÑ, which exhibits the same stem *skriv-* throughout its paradigm, the verb GOUZOUT exhibits four different stems: *goar-* (in the 3SG present indicative), *gouz-* (elsewhere in the present indicative), *goui-* (in imperfect and irrealis forms), and *gouez-* (in future and past indicative forms and in the imperative); in view of this stem alternation, the imperative forms of GOUZOUT (unlike those of SKRIVAÑ) cannot be seen as syncretic. Second, the present indicative forms of GOUZOUT deviate in three ways from the regular pattern. These deviations are formalized in (34): the 1SG form is defined by SupC[34a] rather than by the portmanteau (13p); the 2PL form, by SupC[34b] rather than by SupC[17a]; and the 3PL form, by SupC[34c] rather than by SupC[14b]. Although their component rules apply widely in Breton conjugation, the SupCs in (34) are restricted in application to a small set of forms, among them the present-tense stem *gouz-* of GOUZOUT.

Table 7. Finite forms of the Breton irregular verb GOUZOUT "know"

a. Indicative

	Present	Imperfect	Future	Past
1SG	*gouz -o -n*	*goui -e -n*	*gouez -i -n*	*gouez -is*
2SG	*gouz -ez*	*goui -e -s*	*gouez -i*	*gouez -j -out*
3SG	*goar*	*goui -e*	*gouez -o*	*gouez -as*
1PL	*gouz -o -m*	*goui -e -m*	*gouez -i -m*	*gouez -j -o -m*
2PL	*gouz -o -c'h*	*goui -e -c'h*	*gouez -o -t*	*gouez -j -o -c'h*
3PL	*gouz -i -nt*	*goui -e -nt*	*gouez -i -nt*	*gouez -j -o -nt*
IMPS	*gouz -e -r*	*goui -e -d*	*gouez -o -r*	*gouez -j -o -d*

	b. Irrealis		c. Imperative
	Present	Past	
1SG	*goui -f -e -n*	*goui -j -e -n*	
2SG	*goui -f -e -s*	*goui -j -e -s*	*gouez*
3SG	*goui -f -e*	*goui -j -e*	*gouez -e -t*
1PL	*goui -f -e -m*	*goui -j -e -m*	*gouez -o -m*
2PL	*goui -f -e -c'h*	*goui -j -e -c'h*	*gouez -i -t*
3PL	*goui -f -e -nt*	*goui -j -e -nt*	*gouez -e -nt*
IMPS	*goui -f -e -d*	*goui -j -e -d*	

Source:
Trépos (n.d.: §§ 493, 507, 518, 530, 541, 549, 477); but cf. Kervella (1976: § 206)

(34) Some SupCs for the inflection of GOUZOUT "know"
SupC[34a]: ((13f) Ⓢ$_{\{\text{IND PRS}\}}$ (13d)) = [{IND PRS 1SG} : *-o-n*])
SupC[34b]: ((13j) Ⓢ$_{\{\text{IND PRS}\}}$ (13d)) = [{IND PRS 2PL} : *-o-c'h*])
SupC[34c]: ((13k) Ⓢ$_{\{\text{IND PRS}\}}$ (13c)) = [{IND PRS 3PL} : *-i-nt*])

Conjugating prepositions

Breton prepositions inflect for the person, number, and (in the 3SG) gender of a pronominal object. Prepositional inflection follows two main patterns, the *o* type and *i* type; these are exemplified by the paradigms of the prepositions EVIT "for" and OUZ "against" in Table 8. As these examples show, some of the patterns of conditional exponence observed in the inflection of verbs have analogues in the inflection of prepositions. Prepositions of the *o* type inflect for 1SG, 2SG, 1PL and

2PL objects by means of *-o-n, -out, -o-m*, and *-o-c'h*, recalling the subject agreement of verb forms such as *gouz-o-n* "I know", *skriv-j-out* "you (SG) wrote", *skriv-o-m* "we write", and *gouz-o-c'h* "you (PL) know". Prepositions of the *i* type inflect for 1SG and 2SG objects by means of *-i-n* and *-i-t*, recalling the subject agreement of verb forms such as *skriv-i-n* "I will write" and *skriv-i-t* "you (PL) write". Plural forms of *i* type prepositions are like those of the *o* type, though *i* type prepositions may alternatively inflect for a 1PL object by means of *-i-m*, recalling the subject agreement of verb forms such as *skriv-i-m* "we will write".

Table 8. Inflection of Breton EVIT "for"and OUZ "against" for a pronominal object

			***o* type**	***i* type**
Singular	1st		*evid-o-n*	*ouz-i-n*
	2nd		*evid-out*	*ouz-i-t*
	3rd	masc.	*evit-añ*	*out-añ*
		fem.	*evit-i*	*out-i*
Plural	1st		*evid-o-m*	*ouz-o-m, ouz-i-m*
	2nd		*evid-o-c'h*	*ouz-o-c'h*
	3rd		*evit-o, evit-e*	*out-o, out-e*

Sources:
Trépos (n.d.: §§ 338–339); Kervella (1976: § 583)

Here we see another manifestation of conditional exponence: the same morphological marking may express different content depending on the category of the expression that it serves to inflect. In the inflection of verbs, the SupCs for *-o-n, -i-n, -o-m, -i-m, -o-c'h* and *-i-t* (= SupC[34a], SupC[15a], SupC[14a], SupC[15b], SupC[34b] and SupC[17a]) all express tense and mood; but because prepositions don't inflect for tense or mood, these same markings only express person and number in the inflection of a preposition, in accordance with the rule combinations in (35).

(35) Some rule combinations for the inflection of Breton prepositions
SupC[35a]: ((13f) $\circledS_{\{o \text{ type}\}}$ (13d)) = [{1SG, *o* type} : *-o-n*]
SupC[35b]: ((13f) $\circledS_{\{i \text{ type}\}}$ (13c)) = [{1SG, *i* type} : *-i-n*]
DRC[35c]: ((13h) ∘ (13d)) = [{1PL} : *-o-m*]
SupC[35d]: ((13h) $\circledS_{\{i \text{ type}\}}$ (13c)) = [{1PL, *i* type} : *-i-m*]
DRC[35e]: ((13j) ∘ (13d)) = [{2PL} : *-o-c'h*]
SupC[35f]: ((13i) $\circledS_{\{2\text{SG}, i \text{ type}\}}$ (13c)) = [{2SG, *i* type} : *-i-t*]

Dialectal variation

Breton exhibits considerable dialect variation, and one dimension of this variation is its system of conditional exponence. In the Trégor dialect of Breton, for example, the suffix *-f* observed in the irrealis present is also employed in plural forms of the future (e.g., *kar-f-o-mp, kar-f-e-t, kar-f-o-nt* "we, you (PL), they will love"; Le Clerc 1908: 76). Thus, for this dialect area, the *-f* rule (13b) might be better represented as in (36), so that it, too, is a locus of conditional exponence: in combination with the *-o* rule (13d) or the *-e-t* SupC ((13i) Ⓢ$_{\{2PL\}}$(13e)), it serves as an exponent of the future tense; in other combinations, as an exponent of the irrealis present. Numerous other instances of dialectal variation in the Breton system of conditional exponence can be cited; a full and precise account of these will be quite extensive and must await further research.

(36) [{} : *-f*]

Abbreviations

1, 2, 3	1st, 2nd, 3rd person	IRR	irrealis
AGT	agent	PASS	passive
DRC	default rule composition	PAT	patient
FUT	future	PL	plural
IMP	imperative	PRS	present
IMPF	imperfect	PST	past
IMPS	impersonal	SG	singular
INCL	inclusive	SupC	supplementational rule combination
IND	indicative		

References

Anderson, Stephen R. 1992. *A-morphous morphology*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511586262
Aronoff, Mark. 1994. *Morphology by itself*. Cambridge, MA: MIT Press.
Bilgin, Orhan. 2016. Frequency effects in the processing of morphologically complex Turkish words. Istanbul, Turkey: Boğaziçi University MA thesis.
Booij, Geert. 2005. *The grammar of words*. Oxford: Oxford University of Press.
Booij, Geert. 2010. *Construction morphology*. Oxford: Oxford University Press.
Driem, George van. 1987. *A grammar of Limbu*. Berlin and New York: Mouton de Gruyter. https://doi.org/10.1515/9783110846812
Favereau, Francis. 1997. *Grammaire du breton contemporain*. Morlaix: Skol Vreizh.
Harris, Alice C. 2017. *Multiple exponence*. Oxford: Oxford University Press.

Haspelmath, Martin. 1995. The growth of affixes in morphological reanalysis. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 1994*, 1–29. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-017-3714-2_1

Kempf, Luise & Stefan Hartmann. 2018. Schema unification and morphological productivity: A diachronic perspective. In Geert E. Booij (ed.), *The construction of words: Advances in Construction Morphology*, 441–474. Berlin: Springer. https://doi.org/10.1007/978-3-319-74394-3_16

Kervella, F. 1976. *Yezhadur bras ar brezhoneg*. Brest: Al Liamm.

Le Clerc, L. 1908. *Grammaire bretonne du dialecte de Tréguier*. Saint-Brieuc: René Prud'homme.

Matthews, P. H. 1972. *Inflectional morphology: A theoretical study based on aspects of Latin verb conjugation*. Cambridge: Cambridge University Press.

O'Donnell, T. J. 2015. *Productivity and reuse in language*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9780262028844.001.0001

Stump, Gregory. 2001. *Inflectional morphology*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486333

Stump, Gregory. 2016. *Inflectional paradigms*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781316105290

Stump, Gregory. 2017a. Polyfunctionality and the variety of inflectional exponence relations. In Ferenc Kiefer, James P. Blevins & Huba Bartos (eds.), *Perspectives on morphological organization: Data and analyses*, 11–30. Leiden: Brill.

Stump, Gregory. 2017b. Rule conflation in an inferential-realizational theory of morphotactics. *Acta Linguistica Academica* 64(1). 79–124. [http://akademiai.com/loi/2062] https://doi.org/10.1556/2062.2017.64.1.3

Stump, Gregory. 2017c. Rules and blocks. In Claire Bowern, Laurence Horn & Raffaella Zanuttini (eds.), *On looking into words (and beyond)*, 421–440. Berlin: Language Science Press.

Stump, Gregory. 2019a. An apparently noncanonical pattern of morphotactic competition. In Franz Rainer, Francesco Gardani, Wolfgang Dressler & Hans Christian Luschützky (eds.), *Competition in inflection and word-formation*, 259–278. Berlin: Springer. https://doi.org/10.1007/978-3-030-02550-2_10

Stump, Gregory. 2019b. Some sources of apparent gaps in derivational paradigms. *Morphology* 29. 271–292. https://doi.org/10.1007/s11525-018-9329-z

Stump, Gregory. 2020. Complex exponents. In Lívia Körtvélyessy & Pavol Štekauer (eds.), *Complex words: Advances in morphology*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108780643.010

Stump, Gregory. To appear. Rule conflation, potentiation, affix telescoping. In Andrea Sims et al. (eds), *Morphological typology and linguistic cognition*. Cambridge: Cambridge University Press.

Teddiman, Laura Gene Helena. 2012. An experimental investigation of complexity-based ordering. Edmonton, Canada: University of Alberta dissertation.

Trépos, Pierre. N.d. *Grammaire bretonne*. Rennes: Ouest France.

Zwicky, Arnold M. 1985. How to describe inflection. In Mary Niepokuj, Mary Van Clay, Vassiliki Nikiforidou & Deborah Feder (eds.), *Proceedings of the Eleventh Annual Meeting of the Berkeley Linguistics Society*, 372–386. Berkeley: Berkeley Linguistics Society.

CHAPTER 15

My favorite morphome

The Arabic suffix AT

Robert Hoberman
Stony Brook University

Arabic has a suffix, glossed here as AT, that is a clear and simple example of a morphome. It most frequently and productively marks feminine gender in singular nouns and adjectives, but in fact it has diverse morphological, syntactic, and semantic functions that cannot be unified. That all these functions are expressed by a single element AT, rather than a clutch of accidentally homophonous suffixes, is proven by the fact that AT, in all its functions, has two allomorphs, /at/ and /ah/, with identical distributional patterns no matter which function it is an exponent of. Because AT is not unifiable on the function side and not simplex on the form side, it is a purely morphological entity, a morphome.

Keywords: Arabic, *ta' marbuta*, morphome, allomorph, gender, number, collective, singulative, feminine

1. Introduction

Naming is not trivial. Naming of scientific entities, at least in our field, is often treated casually, sometimes obscurely, sometimes jocularly. The term *morphome*, which Mark Aronoff coined in 1994, is an excellent term, and an illustration of the non-trivial impact of naming. A sizeable literature has grown up around the term morphome in the years since its creation, with a wide range of understandings of that concept. If Aronoff had described the same patterns that he wrote about in that book without creating the name, it is likely that the commonalities among the topics discussed in the subsequent literature would not have been noticed and the topics not seen as bearing on each other.

My understanding of the concept of morphome is simple: any entity, pattern, or process that cannot be attributed entirely to phonology, syntax, or semantics is morphomic, and if it is an entity (word, root, affix), it is a morphome. In this chapter I will describe a simple morphome, one that to me is the simplest and clearest example, though for that reason not the most complex or interesting.

https://doi.org/10.1075/cilt.353.15hob

The unit that I will discuss is a suffix in Arabic[1] which I will label *AT*. Most often it marks feminine gender. In order to show that it is truly a morphome, I will need to show two things. On one hand, I will show that it has no clearly unified or unifiable function in syntax and semantics, that it has such diverse syntactic and semantic functions that it cannot be said to be an exponent of a single cluster of features. On the other hand, at the level of phonological form, it is sufficiently complex that it is unreasonable to think of it as simply a set of accidentally similar suffixes. In other words, by showing that it has more than one allomorph and it has more than one function, we demonstrate that it is indeed an *it*.

2. The phonological level

First at the phonological level, the Arabic suffix AT has two allomorphs: a long form *at* and a short form *a* or *ah*. The selection of the long or short form is different in Classical Arabic and in the modern Vernaculars, while Modern Standard Arabic as spoken in practice varies between the two patterns. In Classical Arabic and the most formal registers of Modern Standard Arabic, AT is always pronounced *at* except at the end of a phrase (known as pausal position), where it is *ah*. For example, *mudarris-at-u-n* "teacher-AT-NOM-ABSL" = "(female) teacher", *kabiir-at-u-n* "big-AT-NOM-ABSL",[2] are pronounced in phrase-final (pausal) position *mudarrisa(h), kabiira(h)*. In most modern Vernacular Arabic, AT has its long *t*-form *at, it, et, ət* when it is followed by another suffix in the word or by another word closely bound to it in a construct phrase. Otherwise, it is *a, e*, or *i*, according to the dialect. The alternation of the two allomorphs is analyzed in Hoberman 1996.[3]

1. I use the term Standard Arabic to refer to the common features of Classical Arabic and Modern Standard Arabic, which are essentially the same language, differing mainly in lexical changes and innovations and in stylistic features. Quite different from Standard Arabic are the many modern Vernacular Arabic dialects. Examples given without an indication of which variety of Arabic they are from represent Modern Standard Arabic.

2. The suffix *-n*, glossed here as ABSL for 'absolute', appears by default in most classes of nouns and adjectives when no other affix of certain sorts is present. Although it is often referred to as an indefinite marker, that cannot be right, as it appears on many proper names, such as *muħammad-u-n* "Muhammad". As it is a morphological default with no expressive function, it too could be considered a morphome.

3. Transcription: A dot under a letter indicates pharyngealization. I use the symbol *ɟ* in all dialects for a phoneme that is variously pronounced as *dʒ, ʒ, g*, and various other ways. All other symbols are used as in IPA.

For simplicity and legibility in the rest of this paper, words containing AT will be cited in their pausal forms, ending in *ah*, with the case and absolute markers deleted as they are in pause. Thus the words in the preceding paragraph would be cited as *mudarris, mudarrisah, kabiir, kabiirah*. The sources of data, if not specified, are the following: for Classical Arabic, Wright 1896; for Modern Standard Arabic, Wehr 1976 and Ryding 2005; for Syrian Arabic, Cowell 1964; for Palestinian Arabic, Seeger 2018; for Egyptian Arabic, Hinds & Badawi 1986.

3. The morphological level

Is the alternation of the long form *at* and the short form *a(h)* simply phonological? No: word-final *t* does not otherwise delete in pause. In particular, there is a different suffix, likewise *at*, that must be distinguished from AT. That *at*, the marker for the third person feminine singular in the perfective aspect of verbs, never loses the *t*.

The most frequent and most productive function of AT is to mark feminine gender in nouns and adjectives: *mudarris* "(male) teacher", *mudarrisah* "(female) teacher", *kabiir* "big (MASC)", *kabiirah* "big (FEM)". Because this function is so prevalent, AT is often called the feminine marker. However, as we will see, it has many other functions.

3.1 Morphological functions: Feminine forms

AT is extremely productive in forming nouns representing females (even in borrowed nouns like *duktuur, duktuurah*) and feminine adjectives. There are also a great many lexically feminine nouns with AT that are not paired with AT-less masculines. Some examples are *madrasah* "school", *luɣah* "language", *ɣurfah* "room", *muħaadaθah* "conversation" *wilaadah* "birth", *wizaarah* "ministry". Some bases occur both with and without AT, with different meanings: *maktab* "office", *maktabah* "library, bookstore". A very numerous group consists of nouns formed with the suffix *ijj* (which otherwise forms adjectives) followed by AT, denoting abstractions like *iʃtiraakijjah* "socialism", *masiiħijjah* "Christianity", *ʃarʕijjah* "legitimacy", *ʕamalijjah* "operation", *waṭanijjah* "patriotism", and some are concrete: *ɟumhuurijjah* "republic", *baʃarijjah* "humankind". All of these are feminine in gender. It is perhaps worth noting that there are two other suffixes that mark feminine gender in certain environments, *aa* and *aaʔ*, and a few feminine nouns have no overt mark. Thus we cannot say that feminine gender is uniquely expressed by AT.

3.2 Morphological functions: Singulative for mass nouns

It marks singulative, count meaning when attached to mass or collective nouns: *tuffaaħ* "apple (mass)", *tuffaaħah* "an apple"; *ʃaɟar* "trees (collective)", *ʃaɟarah* "a tree", *daɟaaɟ* "chicken(s)", *daɟaaɟah* "a chicken", *riiʃ* "feathers", *riiʃah* "a feather". At the same time, the gender changes from masculine to feminine.

3.3 Morphological functions: Gerund for action

Similarly, when attached to a gerund or a generic noun referring to action, it marks a single act: *fikr* "thinking, cogitation", *fikrah* "a thought"; *ḍaħk* "laughing", *ḍaħkah* "a laugh", *ʃaħn* "loading", *ʃaħnah* "a shipment", *intifaaḍ* "shaking off", *intifaaḍah* "an act of shaking off", *raqṣ* "dance", *raqṣah* "a dance". Here too the gender changes, and the forms with AT are grammatically feminine.

3.4 Morphological functions: Plural for human nouns

The suffix AT forms the plural of a good number of nouns referring to humans, and, perhaps surprisingly, these are all masculine nouns (in both the singular and plural): *ħaraamijj* "a thief", pl. *ħaraamijjah; raɟɟaal* "a pedestrian", pl. *raɟɟaalah*; and Syrian *kəndarɟi* "cobbler", pl. *kəndarɟijje; sammaan* "grocer", pl. *sammaane*. Grammars of Standard Arabic, Classical or Modern, barely mention these if at all (they are described in Wright 1896: 232–233 in a fine-print subordinated 'remark'), but they are not uncommon, probably increasingly frequent. Though Wright and others call them 'collectives', a Google search turns up many examples of them as subjects of verbs showing plural agreement, rather than feminine singular which would be expected if they were collectives. The best description of them that I know is by Cowell (1964: 213–214). Because these are generally neglected in grammars, I will list examples of them from different Arabic varieties in an Appendix.

3.5 Morphological functions: Component of broken plural

To be distinguished from the function of AT by itself as an exponent of plural number, the same suffix AT forms an integral component of several so-called 'broken plurals', that is, plurals formed by ablaut and template-manipulation. These are restricted in form, but there are quite a lot of them and some of the patterns are productive. All of these are masculine nouns, nearly all referring to humans. Table 1 shows a selection of Standard Arabic words with plurals containing AT.

Table 1. Broken plural patterns containing AT

Singular	Plural	Gloss
ʔustaað	ʔasaatiðah	"professor"
duktoor	dakaatirah	"doctor"
maṭraan	maṭaarinah	"metropolitan (bishop)"
ʔiswaar	ʔasaawirah	"bracelet, bangle"
ṭaalib	ṭalabah	"student"
ɟanaaħ	ʔaɟniħah	"wing"
zamaan	ʔazminah	"time"
ʕaziiz	ʔaʕizzah	"honorable, dear"
ħabiib	ʔaħibbah	"dear, friend"
sajjid	saadah	"sir, gentleman"
qaaḍii	quḍaah	"Islamic judge"
dubb	dibabah	"bear"
ʔax	ʔixwah	"brother"
fataa	fitjah	"young man"
xaal	xuʔuulah	"maternal uncle"
faħl	fuħuulah, fiħaalah	"stallion"

3.6 Morphological functions: Gerunds of certain verb classes

AT occurs productively in gerunds of certain verb classes. Every Arabic verb has at least one, and often several, gerunds or nouns of action, which are formed by templatic change from the base verb, and some of the patterns include AT: *kitaabah* "writing", *taɟribah* "experimenting", *tasmijah* "naming", *taɣṭijah* "covering", *muħaawalah* "attempt", *munaaqaʃah* "debate".

In one class of verbs the presence of AT is conditioned by a phonological property, with complete regularity. Verbs of class IV, with the shape *ʔaCCaCa* in the perfective aspect, such as *ʔarsala* "he sent", *ʔaslama* "he submitted", have gerunds of the form *ʔiCCaaC: ʔirsaal* "sending", *ʔislaam* "submission, Islam". If the verb has the form *ʔaCaaCa* (based on roots with a semivowel as the middle consonant of the three-consonant root), AT is added in the gerund: *ʔaðaaʕa* "he broadcast", *ʔiðaaʕah* "broadcasting", *ʔadaara* "he administered", *ʔidaarah* "administration". All the gerunds with AT, and only those, are feminine in gender.

3.7 Morphological functions (unproductive): In numerals

The remaining types of occurrences of AT are not productive. In the numerals 3–10 and 13–19, the suffix AT appears on numbers that agree with masculine nouns, not feminine: for example, *xams* "five (FEM)", *xamsah* "five (MASC)". This polarity was inherited from proto-Semitic intact in Standard Arabic. In many of the modern Vernacular dialects, the long and short forms of these numerals still exist but with different distribution. The long form appears on numbers in isolation, and the short form when the numeral appears before the counted noun, regardless of gender, as in Syrian Arabic *xamse* "five" but *xams banaat* "five girls", *xams wlaad* "five boys".

3.8 Morphological functions (unproductive): Masculine nouns of esteem

Strikingly, there are a few *masculine* nouns in AT, all of them denoting especially esteemed male humans: *xaliifah* "caliph", *ʕallaamah* "erudite, very learned".

3.9 Summary of morphosyntactic properties

I will now summarize the morphosyntactic properties of AT, illustrating each type with one or two examples in Table 2.

Table 2. Morphosyntactic properties of words containing AT

	Gloss	Base	Base+AT	Gender of base	Gender of word with -AT	Number of word with AT	Increment of meaning or function
1.	"teacher"	mudarris	mudarrisah	M	F	SG	female
	"big"	kabiir	kabiirah	M	F	SG	feminine
	"school"	–	madrasah	–	F	SG	–
2.	"apple"	tuffaaħ	tuffaaħah	M	F	SG	singulative (object)
3.	"laugh"	ḍaħk	ḍaħkah	M	F	SG	singulative (action)
4.	"thief"	ħaraamijj	ħaraamijjah	M	M	PL	plural
5.	"doctor"	duktoor	dakaatirah	M	M	PL	plural
6.	"administration"	–	ʔidaarah	–	F	SG	–
7.	"five"	xams	xamsah	F	M	PL	masculine
8.	"caliph"	–	xaliifah	–	M	SG	–

The table shows that words with AT can be feminine or masculine, singular or plural. This is so even if we ignore the unproductive types 7 and 8. Thus AT can be an exponent of feminine or masculine gender, of plural number or singulativity. In its productive functions, it can be defined as in (1):

(1) AT is an exponent of feminine singular or of masculine plural.

4. Discussion

Can we say more about the meaning or function of AT? We can look at this two ways, using the Jakobsonian concepts of *Gesamtbedeutung* and *Grundbedeutung*. The clearest exposition I know of the difference between these is by Christian Lehmann (n.d.). Here are Lehmann's definitions, in my translation:

(2) a. **Gesamtbedeutung**: "The OVERALL MEANING of an ambiguous expression is the intersection of the various meanings, i.e., the portion that is common to all of them. The overall meaning can be very abstract."

b. **Grundbedeutung**: "The CORE MEANING of an ambiguous expression is that meaning from which the other meanings can be (diachronically) derived, where the latter are variants of the basic meaning that have come about in various ways. The basic meaning is typically more concrete than the other variants."

The reason Lehmann defines the "core meaning" in terms of diachrony is that he uses the terms in the context of grammaticalization, where the *Grundbedeutung* is original and the *Gesamtbedeutung* includes all later developments. To determine the diachronically original function of AT, we can look at the two most thoroughly documented and studied Central Semitic languages alongside Arabic: Hebrew and Aramaic. In these two languages the cognate of AT always marks feminine gender in singular nouns and adjectives (except for the polarity in numerals like those in § 3.7). As summarized in (3), this comparison supports the result we get from looking only at the most productive functions in Arabic: the Core Meaning of AT is to mark feminine gender. What is its Overall Meaning, the intersection of its various functions? It has none. It can mark masculine gender as well as feminine, plurality as well as singulativity. If our arm is twisted and we must attribute a *Gesamtbedeutung* to it, the Overall Meaning is simply 'marked', which isn't saying much at all.

(3) a. The Core Meaning of AT is as the exponent of feminine gender (in singular nouns and adjectives).

b. There is no Overall Meaning.

5. Conclusion

The Arabic writing system recognizes the status of AT as an entity in an astonishingly effective and revealing way. Recall that AT has two allomorphs: /at/ and /ah/. The suffix AT has its own unique form in the Arabic script: ة, which is a hybrid of the shape of ه <h> and the dots of ت <t>. This ة is called *taaʔ marbuuṭah* "bound <t>", as opposed to ت *taaʔ maftuuħah* "open <t>". It represents AT and only AT, and is pronounced as either allomorph, as appropriate in any context. Though it is not counted as a separate letter of the alphabet, it is obligatory in normative orthography. It probably originated as a historical accident, as a product of the recording of two dialects at two periods of history, but the result is a beautiful bit of morphomic orthography.

Acknowledgement and Dedication

It is a pleasure to dedicate this paper to Mark Aronoff. Mark and I have been friends for roughly forty years, and I have much to thank him for. He has supported and encouraged me in many ways throughout my career. I can't think of any linguist other than Mark whose ideas nearly always strike me as perfectly plausible, full of potential, and probably correct. After reading or hearing his work, I usually feel that that is exactly what I would have said if I were as creative and insightful as he is. I would think that that is because of shared elements of our family backgrounds and earliest education and the languages we studied, but surely other people have the same backgrounds but have different ideas and interests. Who knows? I'm just grateful to have been Mark's colleague and to have learned from him all these years.

References

Aronoff, Mark. 1994. *Morphology by itself: Stems and inflectional classes*. Cambridge, MA: MIT Press.

Cowell, Mark W. 1964. *A reference grammar of Syrian Arabic*. Washington, DC: Georgetown University Press.

Hinds, Martin & El-Said Badawi. 1986. *A dictionary of Egyptian Arabic, Arabic-English*. Beirut: Librairie du Liban.

Hoberman, Robert D. 1996. Subtractive morphology and morpheme identity in Arabic pausal forms. In Geert Booij & Jaap van Marle, (eds.), *Yearbook of morphology 1995*, 161–174. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-017-3716-6_9

Lehmann, Christian. n.d. Grundbedeutung und Gesamtbedeutung. www.christianlehmann.eu/ling/lg_system/sem/grundbedeutung_gesamtbedeutung.php. (8 July 2020.)

Ryding, Karin C. 2005. *A reference grammar of Modern Standard Arabic*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486975

Seeger, Ulrich. 2018. *Wörterbuch Palästinensisch–Deutsch*. www.uni-heidelberg.de/fakultaeten/philosophie/ori/semitistik/seeger_woerterbuch.html. (16 December 2018.)
Wehr, Hans. 1976. *A dictionary of modern written Arabic*, ed. by J. Milton Cowan, 3rd edn. Ithaca, NY: Spoken Language Services.
Wright, William. 1896. *A grammar of the Arabic language*. 3rd edn. Volume 1. Cambridge: Cambridge University Press. Reprinted: Beirut: Librairie du Liban, 1974.

Appendix. Words with AT as the sole marker of plural number

Singular	Plural	Gloss
Modern Standard Arabic (Wehr 1976)		
ħaraamijj	ħaraamijjah	"thief"
kutubijj	kutubijjah	"bookseller"
saaʕaatijj	saaʕaatijjah	"watchmaker"
ʔaalaatijj	ʔaalaatijjah	"musician"
kawaaliinijj	kawaaliinijjah	"locksmith"
jaziidijj	jaziidijjah	"Yazidi"
ʔafandijj	ʔafandijjah	"gentleman"
inkiʃaarijj	inkiʃaarijjah	"Janissary"
baraanisijj	baraanisijjah	"maker of burnooses"
buruuɟijj	buruuɟijjah	"trumpeter"
balṭaɟijj	balṭaɟijjah	"gangster"
bahlawaan	bahlawaanah	"acrobat"
ɟazmaɟijj	ɟazmaɟijjah	"shoemaker"
ħamalijj	ħamalijjah	"ambulant water vendor"
ħaanuutijj	ħaanuutijjah	"mortician"
rubbaanijj	rubbaanijjah	"naval captain"
qandalaft	qandalaftah	"sexton, sacristan"
Classical Arabic (Wright 1896: 232–233)		
ṣuufijj	ṣuufijjah	"Sufi"
raɟɟaal	raɟɟaalah	"pedestrian"
nað̣ð̣aar	nað̣ð̣aarah	"spectator"
Palestinian Vernacular (Seeger 2018)		
ʔafɣaani	ʔafɣaanijje	"Afghan"
ʔooroobbi	ʔooroobbijje	"European"
ʔaalaati	ʔaalaatijje	"musician"
ʔiiṭaali	ʔiiṭaalijje	"Italian"
baħri	baħrijje	"maritime"
badawi	badawijje	"Beduin"
bustaani	bustaanijje	"gardener"

Singular	Plural	Gloss
beeruuti	beeruutijje	"Beiruti"
taħtaani	taħtaanijje	"lower, below"
ħaḏ̣ari	ħaḏ̣arijje	"sedentary" (non-nomadic)
ʔakkiil	ʔakkiile	"voracious"
baqqaal	baqqaale	"grocer"
ħaɟɟaar	ħaɟɟaara	"stone quarryman"
xarriiṭ	xarriiṭa	"liar"
Egyptian Vernacular (Hinds & Badawi 1986)		
ʔismaʕallaawi	ʔismaʕalawijja	"one from Ismailia"
zalanṭaħi	zalanṭaħijja	"destitute"
ʔiskandaraani	ʔiskandaraanijja	"one from Alexandria"

CHAPTER 16

In further pursuit of the adjective

Evidence from the Siouan language Osage

Marcia Haag
University of Oklahoma

The Osage language is shown to have a lexical category Adjective based primarily on morphological and secondarily on syntactic evidence. The hypothesis with the most currency is that Siouan predicates are either active or stative verbs, and that given the lack of distinguishing morphology, adjectives cannot be classified as separate from stative verbs. I show that Osage treats adjectives differently from stative verbs in compound formation, in noun phrase modification, as clausal predicates, and in the exponence of Scale. The number of predicates that could be thus classified as adjectives seems to be larger than those in some other Siouan languages such as Hocąk and Lakota.

Keywords: Osage, Siouan, adjectives, lexical properties

1. Outline of the problem

Attempts to define what constitutes the category Adjective have led researchers down a number of paths, as semantic, syntactic, lexical, and morphological criteria have all been deployed to shed light on this elusive lexical category. I will refer only sparingly to this very large body of research and instead limit this investigation to lexical and morphological considerations that pertain particularly to the Siouan language Osage.

Linguists have noted for some time (Dixon 1982 is a good summary) that while (virtually) all languages have the categories Noun and Verb, the category Adjective may not be present in all, or it may consist of a small, closed set of members. Dixon (1982: 3–5) further notes that in those languages with a minor adjective class, the semantic content of members tends to be of 'descriptive adjectives.' Color words, *big, small, good*, and *bad* are representative of the members of small closed classes of adjectives.

https://doi.org/10.1075/cilt.353.16haa

So if languages do not encode the kinds of semantic information that adjectives typically provide in a large, open lexical category, they must do so in their nominal or verbal systems. Immediately we see that status as an adjective cannot be purely a matter of notional content, but this does not preclude a tendency for adjectives to have an affinity for certain kinds of notional content; surely this content has shown itself to be descriptive and attributive.

In what we might in jocular fashion call ‘well behaved’ languages with respect to an adjective class, members of that class could be diagnosed using a number of criteria. Among these are predictable syntactic distribution, derivational morphology that creates adjectives from other classes and turns adjectives into other classes as well, and straightforward inflectional morphology that marks the unique lexical properties of adjectives. These are Gradation or Scale (Beard 1995; Haag 1996), and their subcategories intensification, comparative, and superlative, as in English.

In Native American linguistics, the debate over the differentiation of verbs from adjectives remains a lively one. I will limit this discussion of the properties of verbs and adjectives, which constitutes a sizeable bibliography, to those most relevant to Siouan languages. The principal characteristics of verbs are argument structure and aspect (Beard 1995; Bybee 1985; Haag 1996). In contrast, syntactic evaluations of adjectives such as Baker’s (2003: 190) hold that they “have no essence” or in Croft’s (1991: 91) view are “the least marked,” of the lexical categories. Beard’s lexical evaluation (1995: 224) proposes a feature +Gradable as the essential property of the qualitative adjective. Haag (1996) goes further in positing the categorial property Gradation (also termed Scale) as inherent to the adjective and its differentiating property.

A question that has long occupied linguists who research American languages is whether a class Adjective exists in languages that do not show any of these linguistic hallmarks in a straightforward way, in particular the lack of dedicated morphology.

2. Three pertinent hallmarks of Siouan and other American languages

The most important feature with respect to the lexicosemantic class Verb that we commonly see in many American languages is an argument-marking system based on semantic roles rather than grammatical ones. Hence, the direct object of a transitive verb, which would very often carry a patient role, has the same pronominal marking as the subject of an intransitive stative verb, which would also carry the semantic role of patient or experiencer; in any case, an affected argument rather than an active one.

Siouan languages routinely, but not always, infix person agreement morphemes into the verb root in phonologically and morphologically complicated ways, which will not be discussed here. In glossing, I will repeat the second part of the verb after the infix. Other researchers (particularly, Quintero 2004, 2009) term the part before the infix a 'preverb.'

(1) a. *sitǫi žįkažį wį o-**ða**-hką* (Siouan, Osage)[1]
yesterday child INDF help-2A-help
"yesterday you helped a child"[1]
b. *sitǫi žįkažį apa o-**ði**-hką*
yesterday child SBJ help-2P-help
"yesterday the child helped you"
c. ***ði**-ðuheka nįkšé*
2P-sick 2.CONT
"you are sick" CONT = continuous aspect

(2) a. *pilashash, alla **ish**-apiila-h* (Muskogean, Choctaw)
yesterday child 2A-help-TNS
"yesterday you helped a child"
b. *pilashash, alla yat **chi**-apiila-h*
yesterday child SUBJ 2P-help-TNS
"yesterday the child helped you" TNS = present or past tense
c. ***chi**-abiika-h*
2P-sick-TNS
"you are sick"

A second important feature of Siouan and other American languages is the lack of marked third person pronominals, in other words, they have zero-marking of the third person. (These languages have other ways to mark semantic roles other than agent and patient, such as markers for dative and instrumental roles that can be associated with third person arguments.)

(3) a. *ðe ø-óhką pe* (Osage)
DEM 3.help COMPL
"he/she helped that one (him/her)"
b. *ø-apiila-h.* (Choctaw)
3A.3P.help-TNS
"someone helped/helps someone"

1. All Osage data is provided by my collaborators Cameron Pratt and Stephanie Rapp, unless attributed elsewhere. Siouan languages are traditionally transcribed in the American Phonetic Alphabet.

A third important feature that makes the identification of adjectives difficult in some American languages is the dearth of derivational morphology and the ubiquity of conversion.

(4)	a.	*íe*	"speak", "speech"	Osage
	b.	*wanǫbre*	"eat", "dinner"	
	c.	*pala*	"shine", "lamp"	Choctaw
	d.	*taloa*	"sing", "song"	

These features make singling out a definitive adjective difficult because, first, stative verbs are not differentiable from would-be adjectives by pronominal marking. Second, the lack of third person pronominals means that only those predicates that can take first or second person marking are eligible for analysis. Pragmatically, a large number of attributive lexemes are almost always used to describe things. Third, the ease of and ubiquity of conversion from one lexical category to another makes it harder to determine if a lexeme is in fact a noun, or a verb, or indeed an adjective.

The course of this analysis will be to examine the modification of nouns first in compounding contexts and then in phrasal contexts. I will show that in both compounds and phrases, the adjective and verb components behave quite differently. Last, I will show that the categorial property Scale or Gradation can be demonstrated using markers of intensification. Again, in Osage intensification marking is different if used with adjectives or with stative verbs.

I assume that a difference in distribution and use of morphology between verbs and adjectives is evidence that Adjective is indeed a lexical category. I will use the term adjective to classify these lexemes under discussion in making my analysis here to avoid clumsy locutions such as 'stative predicates that behave differently.'

3. Related work on Siouan adjectives

Johannes Helmbrecht and Bryan Rosen have done extensive research into the status of adjectives in the Siouan language Hocąk, with different outcomes. In a good summary of this work, Helmbrecht (2006) finds that adjectives cannot be found to be distinct from verbs based on criteria similar to the ones I will use to analyze Osage. He finds that Hocąk has a subclass of intransitive inactive verbs that are not syntactically distinguishable from other verbs based on lack of specialized derivation and the lack of a need for a copula when in predicate position. He finds that this type verb can appear in any noun phrase construction or compound without distinctions. Most important, he finds that that gradation markers – intensifiers – can occur with any class of word. He says (2006: 310) "if *-xji* (the intensive marker) would occur only with property words (those that describe properties), one could make an argument out of this fact for a special adjective construction."

Rosen (2015) uses syntactic analyses of the ordering of adjectives in the noun phrase to find such a distinction. He finds that adjectives in noun phrases do not have to agree with the head noun as do verbs in relative clauses, and that adjectives do indeed have degree specifiers.

In the more closely related Omaha and Ponca, both of the Dhegihan branch of the Siouan family along with Osage, Catherine Rudin (2020) summarizes arguments for and against a category Adjective. Her evidence for them is the need for copular support in predicative positions, and the lack of verbal inflection when in noun modifying position (nearly impossible when only first and second person are marked). She concludes that while there may be an adjective class, it is very small.

Caroline Quintero's 2004 grammar of Osage makes some mention of various behaviors of predicative words that are adjective-like without explicitly arguing for a separate lexical category. Much of the evidence in this chapter comes from a reanalysis of examples that she has given and will be examined in § 5.

Quintero generally acquiesces to the 'stative verb' category for adjectives, saying, "All adjectives can function as verbs in Osage (or may be thought of as verbs) inasmuch as they appear as predicates in clause-final position." But she goes on to say,

> [W]hat is needed is a characterization of each adjective based on its ability to carry either agent or patient inflection (that is, whether inflection marking is obligatory, optional, or impossible, and which kind is used), noting any change in meaning encountered if it is inflected. (Quintero 2004: 397)

Revealingly, Quintero's 2009 dictionary has a lexical category Adjective, suggesting that while she may not have wanted to enter the linguistic fray about their lexicosemantic status, she made a distinction between verb and adjective as a pedagogical and practical matter.[2]

4. Varieties of compounded nouns in Osage

Osage is a moderately compounding language in which endocentric noun-noun compounds are right-headed and compounds made with the predicative categories verb and adjective are left-headed. Many compounds are exocentric, hence metaphorical, however.

2. Carolyn Quintero died in 2008, before the publication of her dictionary.

Some examples of noun-noun compounds are:

(5) a. *hkąące-hpaalą*
apple-nose
"pear"
b. *hápa-wacue*
corn-bread
"cornbread"

Some examples of noun-adjective compounds are:

(6) a. *hkąące-zihcaaðe*
apple-sour
"lemon"
b. *wacue-brehka*
bread-thin
"pancake"

Some examples of noun-verb compounds are:

(7) a. *ho-i-ðuze*
fish-INS-catch
"net"
b. *maze-o-ðįįke*
metal-LOC-grab
"trap"
alternatively:
*maze o-**ða**-ðįįke*
metal-LOC-2A-grab
"trap"

Notice the alternative forms of "trap." In both versions, a locative marker is affixed to the verb root *ðįĩke* "grab," and the second version also includes the second person active morpheme. (Some compounds include the first-person morpheme.) The literal gloss is "metal thing to grab there" or "metal thing you grab there."

The examples of noun-verb compounds are illustrative of a crucial distinction between the categories Adjective and Verb. One of the categorial properties of verbs is argument structure (Beard 1995; Haag 1996; Bybee 1985; Croft 2000). The two examples here show overt marking of the arguments 'instrument,' 'locative,' and 'person'. The category Adjective lacks inherent argument structure. But since it is not necessary for verbal members of compounds to show argument structure, this is not a conclusive argument for the existence of adjectives, though suggestive.

5. Phrasal modification

Osage noun phrases that contain a predicative modifier category, in this case verbs and adjectives, show decided differences. The modifiers that I argue to be adjectives straightforwardly appear after the noun they modify without derivational or other kinds of morphology.

(8) a. *wasápe ska*
bear white
"white bear"
b. *nihka lą̨ą̨ðe*
man big
"big man"
c. *wazį̨ka žuce*
blanket yellow
"yellow blanket"

These phrases appear in larger phrases without further modification.

(9) a. *nihka lą̨ą̨ðe wį̨ i-ða-ðe nį̨kšé*
man big INDF see-2A-see 2CONT
"you see a big man"

Modifying verbs that appear in noun phrases, on the other hand, require particular morphology, generally with the iterative aspect marker *štą̨* suffixed to the verb stem.

(10) a. *wakʔo waðihtą̨-štą̨ wį̨*
woman work-ITER INDF
"a working woman" ITER=iterative aspect
(example from Quintero 2004: 404)

In a more complicated example Quintero states (2004: 405)

> The Osage counterparts of English deverbal adjectives will often retain their verbal status … The adjectival form "sticky" is derived from the valence-reduced form *wastasta* "stick to stuff", with detransitivizing prefix *wa*, plus reduplication to communicate either the plasticity or relentless nature of the event or state.

(11) a. *onǫbre wa-asta-sta-ðe akxa ðaalį̨ akxai*
food VAL-stick-REDUP-CAUS SBJ good CONT.3.PRES
"the sticky food is good" VAL = valence reducer, REDUP=reduplicated

Examine these same verbs when they are in clausal predicative form:

(12) a. *wakʔo apa waðihtą apai*
woman SBJ work CONT.3.MOV
"the woman is working" MOV=moving, out of sight

b. *wanǫbre akxa asta akxai* (Quintero 2009: 18)
food SUBJ stick CONT.3.PRES
"the food sticks or the food is sticking (to something)"
PRES=present, standing

Although Quintero does not delve into the appearance of the iterative aspect marker *štą* on modifying verbs, it is important to note that this aspect marker does not in fact mark a true iterative as much as it distinguishes the verb from a clausal predicate. This state of affairs is entirely comparable to the *-ing* marking of English present participles that are used to modify nouns in phrases, e.g., *working woman*. The suffix *-ing* does not mark a true progressive aspect so much as serve as a morphological cue that the verb is not a clausal predicate. This analysis is in keeping with Aronoff's (1994) inquiry into morphological forms as a separate component of the grammar that may be mapped to different and seemingly unrelated inflections. These repurposed forms are termed *morphomes* in the morphology literature.

I discuss this phenomenon in my discussion of Choctaw intensive marking in adjectives, which is done with the same inflectional tools as aspect marking in verbs, as in the following example from Haag (1996: 62–63):

(13) a. *shokata hasimbish at tohbi-t wanoksho-t biyyika-tok*
opossum tail SBJ white-SS fluffy-SS alone.INTNS-PST
"the opossum's tail used to be white and fluffy"

b. *hasimbish at toyyohbi-t wannoksho-t toba-tok*
tail SBJ white.INTNS-SS fluffy.INTNS-SS become PST
"the tail became extremely white and frazzled"
SS=same subject, INTNS = intensive

The important result shown in Example (13) is that a morphological marker of aspect (a fairly complicated reduplication of an internal consonant) that marks a resolutional aspect 'finally V' in verbs produces intensification when applied to predicates such as *tohbi* "white" and *wanoksho* "fluffy." A single morphological operation has been deployed to give different linguistic results depending on the properties of the words with which they are construed (Haag 1998). Following Aronoff (1994), who first analyzed the morphome as a persistent phenomenon in disparate languages, we need not try to find and force semantic correspondence based on morphological form. Hence, the form *štą* does not need to have one-to-one correspondence with iterative aspect. It is a marker of a verbal modifier of nouns when used in noun phrases.

6. Clausal predicates: Adjectives or stative verbs?

As briefly discussed in § 1, the categorial properties pertaining to Verbs are aspect and argument structure. The main argument for declaring all predicative categories to be verbs (generally excluding predicate nouns) in Siouan and other American languages has been that, as illustrated in Examples (1)–(3), predicates with non-agent subjects are not morphologically distinguishable. Of particular interest are those classified as 'stative verbs.' I have pointed out that since these languages have zero-marked third person agreement, the only predicates that can be inspected are those with first or second person arguments.

Aspect has been much less investigated as a distinction between stative verbs and adjectives. Aspect is an obligatorily marked category in Osage, although there are a number of instances when such marking defaults to zero, particularly in the completive aspect (which is a subcategory of perfective, if even distinguishable from it). Like other Siouan languages, the continuous aspect (termed 'continuative' in Quintero's grammar), bears agreement for person. These aspect markers also suppletively mark the position of the subject, a complex and incompletely studied topic. I have glossed the most neutral position as simply PRES 'present.'[3] Both active and stative verbs have person agreement in continuous aspect markers.

The following Examples (14)–(15) show person agreement in both the verb itself and the continuous aspect marker. Examples (14a)–(d) depict agreement in the stative verb *nǫhpehi* "hungry." The aspect is the 'present, standing, not moving' continuous, which is glossed as PRES.

(14) a. *nǫhpehi* "hungry" (stative verb)
šį̨žižį *akxa* *nǫhpehi* *akxai*
girl SBJ.PRES hungry CONT.3.PRES
"the girl (who is present) is hungry"

b. *nǫhpe-**ą**-hi* *mįkšé*
hungry-**1P**-hungry CONT.1.PRES
"I am hungry (present)"

c. *nǫhpe-**ði**-hi* *nįkšé*
hungry-**2P**-hungry CONT.2.PRES
"you (sg) are hungry (present)"

d. *nǫhpe-**wa**-hi* *ąkaðe*
hungry-**1P.PL.**-hungry CONT.1PL.PRES
"we are hungry (present)"

---

3. This position marker *akxa* and its agreement aspect marker for third person continuous *akxai* (or its variation *akxa)* is used when a subject is present, standing, and not moving, or the speaker's choice of which of these to emphasize if there is contradiction among these conditions.

Examples (15a)–(b) depict person agreement in the active verb *žįhe* "sleep." In (15a) the aspect is *akxai* 'present, not moving' continuous, while in (15b) it is *žąkšé* 'lying continuous.'

(15) a. *žįhe* "sleep" (active verb)
nihka akxa žįhe akxai
man SBJ.PRES sleep CONT.3.PRES
"the man is sleeping"
b. ***ða***-*žįhe žąkšé*
2A-sleep CONT.2.LIE
"you are sleeping (while lying)" LIE = lying position

In both sets of examples, verbs of either class have obligatory person agreement on both the verb and the aspect marker. This will not be the case when the predicates are not verbs.

6.1 Stative-like predicates without subject agreement

The group of clausal predicates of most interest here are those that should be classified as stative verbs but do not have person agreement on the lexeme while retaining it on aspect agreement marker.

The following examples show the use of *tąhe* "well" and *šįta* "fat."

(16) a. *tąhé mįkšé*
well CONT.1.PRES
"I am well"
b. *tąhé ątxąhé*
well CONT.1.STAND
"I am well (while standing)" STAND = standing position
c. *tąhé ąhé*
well CONT.1.MOV
"I am well (while moving)"

(17) a. *šįta mįkšé*
fat CONT.1.PRES
"I am fat"

In Examples (16)–(17), the person-marked aspect is the only necessary reference to person, in contrast with the resumptive person marking in Examples (14)–(15).

6.2 Stative-like predicates with use of copula

A second way that clausal predicates such as *šįta* can be produced is with a person inflected copula.

(18) a. *lą̨ðe brįe*
big 1s.be
"I am big"
b. *šįta ðįe*
fat 2s.be
"you are fat"

The same strategy is used for noun predicates.

(19) a. *wažaže brįe*
Osage 1s.be
"I am Osage"

Again, in Examples (18)–(19), person is marked on the copula only. There is no expectation that person marking should appear on noun phrases, but this is yet another way that adjectives and stative verbs behave differently, suggesting adjectives form a separate category.

6.3 Subjects of adjectival predicates without subject markers

Adjectives may appear as clausal predicates with NP subjects where the subject marker is omitted. An aspect marker may optionally appear.

(20) a. *hcee tą*
buffalo big
"the buffalo is big"
b. *hcee tą pe*
buffalo big COMPL
"the buffalo is big"

In Osage, the completive aspect marker *pe* is the default setting. Its absence is not unusual. However, the lack of a subject marker is another way that this adjectival predicate differs from verbs. The subject marker is necessary with NP subjects of verbs.

7. The property Scale

My final point will be to establish that Osage adjectives may be marked for intensification in a different way than are stative or other verbs. I will compare the placement of two intensifiers *walį* and *xci*, which may both be glossed "very."

In Examples (22) and (23) *xci* appears right-adjacent to the modifiers "mean" and "nice" inside the subject noun phrase, whose boundary is marked with the present/stationary subject marker *akxa*.

(21) a. *ðe šįtožį íisiwaðe* ***xci*** *akxa ą-wą-txą pe*
DEM boy mean **INTNS** SBJ kick-1P-kick COMPL
"that very mean boy kicked me"

(22) a. *ðe šimįžį ðalį* ***xci*** *akxa i-ði-štįhka pe*
DEM girl nice **INTNS** SBJ kiss-2P-kiss COMPL
"that very nice girl kissed you"

Note that the intensive *xci* appears inside the noun phrase.

An adjectival predicate may also be modified with *xci*.

(23) a. *ðe šįtožį akxa íisiwaðe xci akxai*
DEM boy SBJ mean INTNS CONT.3.PRES
"that boy is very mean"

In contrast, the intensive *walį* is used to modify both adjectives and verbs, but it may only appear in the predicate phrase. Additionally, it appears before the predicate, be it verb or adjective, in the syntactic position of an adverb.

(24) a. *ðe šįtožį akxa* ***walį*** *íisiwaðe akxai*
DEM boy SBJ **INTNS** mean CONT.3.PRES
"that boy is very mean"
b. *haace* ***walį*** *nuužužu che*
today **INTNS** rain EVID
"it rained a lot/ it really rained (I surmise)" EVID=evidential

In contrast with *xci, walį* may not appear in the noun phrase.

8. Adjective derivation from verbs

Osage adjectives may be derived from verbs with overt morphology. One common way for this to happen is with the valence reducing, also termed detransitivizing, morpheme *wa*, which is seen in a number of constructions. Recalling *wastasta* "sticky" and *íisiwaðe* "mean" from Examples (11) and (23, 24), these adjectives are derived from *astaðe* "to stick" and *iisiðe* "to hate." The detransitivizing morpheme *wa* yields "(be) sticky" and "(be) hateful." Because these derivations may appear inside the noun phrase, they can be considered adjectives and not verbs.

9. Summary of distinctions between Osage verbs and adjectives

1. Compounds containing a verbal member may include inflections on that verb, including instrumental, locative, and person. Adjectival modifiers in compounds do not bear morphology.
2. Adjectives appear straightforwardly as modifiers in noun phrases, but verbal modifiers in noun phrases need further morphological marking, generally by use of the morphome *štą*.
3. Stative verbs take both person marking on the verb lexeme itself and person agreement on the clausal aspect marker. Predicative adjectives are inflected by use of the aspect marker alone, and its omission is common when context makes interpretation very clear.
4. Adjectives may appear as clausal predicates with NP subjects where the subject marker is omitted. The NP subjects of clauses with verbal predicates require subject markers.
5. Adjectives may be marked for intensification with either *walį* or *xci*. Stative verbs are marked for intensification with *walį*.
6. There is some verb-to-adjective derivation.

10. Discussion

Evidence points to the existence of a class of predicates in Osage that differ from verbs, including stative verbs, in their syntactic but especially their morphological behavior. This class is also strongly associated semantically with those words that are usually classified as adjectives in numerous other languages: they are attributive words such as colors and size.

In §§ 1–8 I laid out tests for making distinctions in verbs and adjectives, the first being the presence of argument structure. Given this rubric for differentiating verbs and adjectives, even stative verbs should have argument structure, which would generally consist of an experiencer or patient subject and no object. Indeed, when we examine stative verbs, they very frequently are those that describe human conditions and internal states such as "be sick," "be weak," and the like. Such verbs readily take first and second person pronominals, as seen in Examples (14a)–(c) with *nǫhpehi* "hungry."

Adjectives, by cited theories, lack argument structure. They would not, then, be expected to morphologically mark experiencer or patient subjects. Recalling Dixon's canonical 'semantic types' for adjectives, we have a large number of attributives that are more easily applied to things: color, size, evaluations of good and bad, dimensions, temperature, and so forth. Such attributives may be less often applied to humans than to things. It is far likelier for one to say "a long hot day" than to say "I am long and hot." While a semantic sense may become incorporated into a language as a verb, those verbs should behave as do others: they should have argument structure. My examination of the Osage lexicon shows that demonstrable stative verbs have animate experiencer and patient subjects. When speakers want to say such utterances as "I am big," they use an aspect marker inflected for person, or a person-inflected copula. In Osage, "big" *łąąðe* or *tą* may have animate or human subjects but these lexemes do not have inherent arguments.

The second verbal property, aspect, is obligatorily marked at the clause level. Both predicate verbs and predicate adjectives have aspect marking, both with person agreement. Here we need to make a negative argument: adjectives that appear as modifiers in noun phrases do not have aspect marking, nor is it expected.

The adjectival property Scale/Gradation is robustly demonstrated in the way intensification is marked differently in adjectives than in verbs, as shown in Examples (21)–(24). Recalling Helmbrecht's comment from § 3, that demonstration of such intensification would provide a robust argument for a class Adjective, this is indeed the case in Osage.

It appears that Osage has evolved away from other Siouan languages in that it has a rather large class of words that behave in these described ways. It is possible that exposure to an adjective-heavy language such as English has contributed to this tendency, but it is also quite clear that Osage has not simply replaced its own lexicon with glossed versions of English adjectives, even under this pressure. The derivational processes that create both complex verbs and deverbal categories are not parallel to anything in English. Also, there is not across-the-board conversion from verbs to adjectives: conversion appears to be a relation between nouns and verbs.

The evidence supporting a separate category Adjective is sufficient to continue to categorize Osage lexemes this way: Quintero's 2009 dictionary, by far the most-referenced source, already does so. Having a category Adjective will greatly help with language learning as it will allow students, virtually all of whom are English-first speakers, to mentally realize this important part of the grammar.

References

Aronoff, Mark. 1994. *Morphology by itself: Stems and inflectional classes*. Cambridge: MIT Press.

Baker, Mark. 2003. *Lexical categories*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511615047

Beard, Robert. 1995. *Lexeme-morpheme base morphology: A general theory of inflection and word formation*. Albany, NY: State University of New York Press.

Bybee, Joan. 1985. *Morphology: A study of the relation between meaning and form*. [Typological Studies in Language 9]. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.9

Croft, William. 1991. *Syntactic categories and grammatical relations*. Chicago IL: University of Chicago Press.

Croft, William. 2000. Parts of speech as language universals and as language-particular categories. In Petra Vogel & Bernard Comrie (eds), *Approaches to the typology of word classes*, 65–102. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110806120.65

Dixon, R. M. W. 1982. *Where have all the adjectives gone? And other essays in semantics and syntax*. Berlin: Mouton. https://doi.org/10.1515/9783110822939

Haag, Marcia. 1996. Lexical categories in Choctaw and Universal Grammar. Stony Brook, NY: Stony Brook University dissertation.

Haag, Marcia. 1998. Word-level evidence for lexical categories in Salishan languages. *International Journal of American Linguistics* 64. 379–393. https://doi.org/10.1086/466367

Helmbrecht, Johannes. 2006. Are there adjectives in Hocąk (Winnebago)? In Ximena Lois & Valentina Vapnarsky (eds), *Lexical categories and root classes in Amerindian languages*, 289–316. Bern: Peter Lang.

Quintero, Carolyn. 2004. *Osage grammar*. Lincoln, NE: University of Nebraska Press.

Quintero, Carolyn. 2009. *Osage dictionary*. Norman, OK: University of Oklahoma Press.

Rosen, Bryan. 2015. The syntax of adjectives in Hocąk. Madison, WI: University of Wisconsin Madison dissertation.

Rudin, Catherine. 2020. On lexical and syntactic categories in Omaha-Ponca (Siouan). Paper presented at The Society for the Study of the Indigenous Languages of the Americas (SSILA), New Orleans, LA, January 2–5.

CHAPTER 17

Two-suffix combinations in native and non-native English

Novel evidence for morphomic structures

Stela Manova[1] and Georgia Knell[1,2]
[1]University of Vienna / [2]Vrije Universiteit Brussel

We demonstrate the existence of a novel type of morphomic structure: the suffix combination in word formation. We ran two psycholinguistic experiments with 45 native and 30 non-native English speakers, asking them to judge whether or not existing and non-existing two-suffix combinations presented without bases occur in English. Both groups identified existing and non-existing suffix combinations with very high accuracy; productive combinations were recognized more accurately than unproductive ones. Our research shows that suffix combinations are listed in the mental lexicon as morphomes. This finding accords with recent research in natural language processing that induces a language's patterns, rules, and semantics based entirely on form relations. We also discuss other possible applications of the morphome outside of theoretical linguistics.

Keywords: morphology, morphome, word formation, suffix ordering, productivity, language processing, foreign language learning, psycholinguistics, computational linguistics, English

> Так оставьте ненужные споры!
> Я себе уже все доказал –
> Лучше гор могут быть только горы,
> На которых никто не бывал.
> На которых еще не бывал.
> *Прощание с горами*, Владимир Высоцкий[1]

1. So stop the unnecessary debates.
I got it all figured out:
Only mountains can be better than mountains,
Those that nobody visited,
Those not yet visited.

Farewell to the mountains by Vladimir Vysotsky; the translation is based on the subtitles of https://www.youtube.com/watch?v=Da9UpNSFGIo

https://doi.org/10.1075/cilt.353.17man

1. Introduction

Since Aronoff's (1994) introduction of the concept of morphome to linguistic theory, there has been much debate on the exact definition and the diachronic development of mophomic patterns (recent overview in Luís & Bermúdez-Otero 2016; for the different definitions of 'morphome' see O'Neill 2014;[2] see also Aronoff 2012, 2016). Morphological forms and patterns of various kinds have been identified as morphomic: single morphemes (inflectional material), stems and whole words, sets of paradigm cells as well as phenomena such as stem formation, stem indexing and inflectional classes. As one can guess from this list, in the literature so far morphomes have been related to inflectional morphology. Additionally, the usefulness of the concept of morphome outside of theoretical linguistics has been neglected, somehow. Thus, the goal of this contribution is to fill these gaps. We discuss morphomes in English word-formation as well as two applications of the morphome: for language processing (in both humans and machines) and for foreign language learning.

Our approach is grounded in mathematics. A mathematician believes that all problems are solved in the real world, i.e., that the solution to a problem consists of making the right analogy with a phenomenon in the real world. To illustrate: a one-floor house consists of rooms, the rooms consist of walls, the walls consist of bricks, the bricks are made of clay (-bearing soil, sand, and lime). Analogizing to grammar: the house is the parallel to a sentence that consists of words (rooms) which are built up of morphemes (bricks) that are made up of phonemes (clay …). For the construction of a house, one can use not only bricks but also readymade walls, which will facilitate the building of the house. Such readymade walls are the parallel of morphomes.

These analogies can also help us understand why the relation of meaning and form in morphology is an issue:

1. bricks (morphemes) can be linked to room types (words), e.g., bathroom bricks (cf. verbalizing morphemes) or other specific features, e.g., glass bricks (cf. number morphemes); this is how lexical approaches to morphology (e.g., Distributed Morphology (DM), Halle & Marantz 1993) treat morphemes.

---

2. O'Neill (2014: 31) identifies three basic definitions for the term 'morphome' and the adjective 'morphomic' in the literature:

1. 'meaningless formatives' which show the same pattern of allomorphy; they can be stems, other inflectional material or even whole-word forms
2. a semantically and syntactically incoherent set of paradigm cells characterized by a particular type of allomorphy
3. phenomena that are not derived by semantico-syntactic features: stem formation, stem-indexing, inflectional classes.

2. bricks (morphemes) can be seen as unspecified units used for building up any type of room (word); this is how inferential approaches (e.g., Paradigm Function Morphology (PFM), Stump 2001) define morphological units.

Likewise for morphomes:

1. one can relate them to a word class (e.g., the L-shape morphome in Romance verb inflection, Maiden 2016 and earlier work) or to specific features (e.g., deponent verb stems and passive, Baerman et al. 2007, among many others)
2. one can also think of morphomes as unspecified pieces of morphological structure processed without any reference to meaning, including no word-class specification (our approach).

Based on the above analogies, we also differentiate between morphemes and morphomes. We see morphomes as readymade pieces of morphological structure consisting of two (or more) morpheme positions (cf. Manova et al. 2020) that can be identified and/or processed without any reference to meaning.

The mathematical nature of our approach to suffix ordering will be addressed in the next section.

This chapter has the following structure. In § 2, we discuss research on suffix ordering in English, introduce our own approach and the selection of the suffix combinations used in this study, including why we define them as morphomic. Section 3 describes two psycholinguistic experiments carried out with native and non-native speakers of English to confirm the morphomic nature of suffix combinations in English word formation. Section 4 contains the discussion and also addresses possible applications of morphomic structures. In § 5 conclusions are drawn.

2. Suffix ordering in English: History of research, our approach and why suffix combinations are morphomes

There has been much research on the ordering of English derivational suffixes[3] and a number of specific proposals have been put forward (in chronological order): level ordering or stratal approach (Siegel 1974; Allen 1978; Selkirk 1982; Kiparsky 1982; Mohanan 1986; Giegerich 1999); selectional restrictions (Fabb 1988; Plag 1996, 1999); the monosuffix constraint (Aronoff & Fuhrhop 2002), and the parsability hypothesis (Hay 2001, 2002, 2003) or complexity-based ordering (CBO) (Plag 2002; Hay & Plag 2004; Plag & Baayen 2009). In this list of approaches, every approach was formulated in response to the incorrect predictions of the preceding

3. On affix ordering in general, see Manova & Aronoff (2010).

approach. Likewise, our approach (Manova 2011) has been formulated as a reaction to CBO that has been illustrated primarily with data from English. In a series of articles, Manova (2010, 2011, 2015) shows that CBO does not hold for Slavic, nor for Romance languages (Manova & Talamo 2015; for Italian, see also Talamo 2015) and suggests an alternative approach based on logic from mathematics: Gauss-Jordan elimination. This method serves to solve large linear systems (i.e., systems with a large number of variables) numerically with only the help of elementary operations (addition, subtraction, multiplication). The idea is to reduce the options for each variable to one, because if there is only one option for a variable, this option is the variable's value, i.e., the solution to the problem. Gauss-Jordan elimination is a masterpiece of logical reasoning because one comes to the solution without 'solving' the problem. What happens is elimination; the problem is eliminated in an easy and elegant way. Due to the limited space of this chapter, we cannot go into detail about Gauss-Jordan elimination, but we refer curious readers to Manova (2011) where the method is illustrated and its application to morphological data is discussed in detail.

Manova (2011) sees derivational suffix combinations as binary structures of the type SUFF1-SUFF2, where SUFF1 has three valency positions for further suffixation: $\text{SUFF2}_{\text{Noun}}$, $\text{SUFF2}_{\text{Adjective}}$ and $\text{SUFF2}_{\text{Verb}}$ (see Table 1). The idea of this distribution of outputs according to the syntactic-class specification of SUFF2 is to have one combination of a kind. And indeed, Manova observes that in most cases there is a single SUFF2 of each syntactic category, i.e., N, A, V, that follows a given SUFF1 (N: *-dom* and V: *-ize* in Table 1).

Table 1. Combinability of the suffix *-ist* (data from Aronoff & Fuhrhop 2002, based on OED, CD 1994)

SUFF1	Syntactic category of SUFF1	SUFF2 according to syntactic category
-ist	N	N: *-dom* (2) A: *-ic* (631), *-y* (5) V: *-ize* (3)

If more than one SUFF2 of the same syntactic category is available (A: *-ic* and *-y* in Table 1), there is one SUFF2 that attaches by default, that is, one of the competing SUFF2 suffixes (*-ic* and *-y*) derives almost all outputs (types/lemmas), *-ic* in our case, because *-ist-ic* derives 631 types in comparison to *-ist-y*, with only 5 types. In other instances of more than one SUFF2 suffix, the available SUFF2 options express completely different semantics (e.g., an abstract noun and an object) and thus do not really compete for the SUFF1. Manova (2011) interprets all these facts as evidence for the uniqueness of suffix combinations. If something is unique, it should be memorized as one of a kind, which for suffix combinations means listedness in the mental lexicon – that is, native speakers should know them by heart. This led to the idea to test native speakers' intuitions about suffix combinability with morphomic pieces of word structure, i.e., with sequences of two suffixes without any relation to lexical bases, be they roots, stems or words.

Having counted suffix combinations in large dictionaries and corpora for different languages (Slavic and non-Slavic alike), Manova (2011) also observes that in cases in which more than one SUFF2 of the same syntactic category is available, one of the SUFF2 suffixes derives more than 10 types (the suffix *-ic* in Table 1). Such suffixes are referred to as default. The SUFF2 suffixes that compete with the default suffix always derive ten types or fewer each (e.g., the suffix *-y* in Table 1 derives 5 types). Thus, 10 types are seen as a threshold for productivity, i.e., suffixes that derive more than 10 types are productive, suffixes that derive 10 types or fewer are unproductive. Significantly, the number of types does not depend on the corpus size, and a dictionary of about 70–100 thousand words and a large corpus of millions of tokens give the same results (Manova & Talamo 2015), which is the case because at some point the suffix combinations start repeating. Since this study is part of a series of studies on the processing of morphological structure, for the sake of uniformity we follow Manova's (2011) understanding of productivity.[4] For alternative approaches to productivity in English word-formation, see, e.g., Baayen (1991, 1993, 2001), Baayen & Lieber (1991), Plag (1999), and Bauer (2001).

The suffix combinations used in the psycholinguistic experiments reported in this study are listed in Table 2 (existing combinations) and in Table 3 (non-existing combinations). In Table 2, the suffixes are given with their syntactic category specifications and the suffix combinations are illustrated with examples. The list of suffix combinations used in this study is mainly based on Manova (2011), i.e., these are the suffix combinations from Aronoff & Fuhrhop (2002) and Plag & Baayen (2009), plus some additional suffix combinations pointed out by native speakers as missing in the two studies. Aronoff & Fuhrhop describe the combinations of 44 English suffixes, based on the Oxford English Dictionary (1994). Plag & Baayen tackle the combinations of 31 English suffixes and their data come from the CELEX lexical database (Baayen, Piepenbrock & Gulikers 1995) and the subcorpus of written English of the British National Corpus. Both studies provide the number of the derivatives formed with a particular suffix combination. Suffix combinations in English appear first heavily restricted (of all possible combinations only a few exist) and second the majority of the existing combinations produce a very limited

4. Manova's approach is defined to make possible research on affix combinability in languages of the inflecting-fusional type where derivational suffixes are followed by inflection and for which the resources available for English do not exist, including corpora with annotated derivational suffixes such as for, e.g., CELEX for English (Baayen et al. 1995).

number of words (types), fewer than 10 types each. These facts also seem to favor listedness in the mental lexicon. Again, if morphology (morphological structure) plays a role in the organization of the mental lexicon, native speakers should know suffix combinations by heart, which means access without reference to meaning, i.e., morphomic processing.

Table 2. Existing suffix combinations used in the experiments reported in § 3

Existing combinations used in task							
Suffix combination		**SUFF1**	**SUFF1 syntactic category**	**SUFF2**	**SUFF2 syntactic category**	**Example word**	**Type frequency**
Productive	-**ative**	-ate	V	-ive	ADJ	neg-at-ive	> 10
	-**ership**	-er	N	-ship	N	lead-er-ship	> 10
	-**icism**	-ic	ADJ	-ism	N	sto-ic-ism	> 10
	-**ifiable**	-ify	V	-able	ADJ	just-ifi-able	> 10
	-**ifier**	-ify	V	-er	N	mod-ifi-er	> 10
	-**ional**	-ion	N	-al	ADJ	profess-ion-al	> 10
	-**ishness**	-ish	ADJ	-ness	N	fool-ish-ness	> 10
	-**ivate**	-ive	ADJ	-ate	V	capt-iv-ate	> 10
	-**ivist**	-ive	ADJ	-ist	N	relat-iv-ist	> 10
	-**ization**	-ize	V	-ation	N	immun-iz-ation	> 10
	-**lessness**	-less	ADJ	-ness	N	weight-less-ness	> 10
	-**liness**	-ly	ADJ	-ness	N	friend-li-ness	> 10
	-**mentary**	-ment	N	-ary	ADJ	comple-ment-ary	> 10
	-**ority**	-or	N	-ity	N	seni-or-ity	> 10
	-**osity**	-ous	ADJ	-ity	N	visc-os-ity	> 10
Unproductive	-**ageous**	-age	N	-ous	ADJ	advant-age-eous	≤ 10
	-**atee**	-ate	V	-ee	N	repudi-at-ee	≤ 10
	-**domless**	-dom	N	-less	ADJ	king-dom-less	≤ 10
	-**eeism**	-ee	N	-ism	N	absent-ee-ism	≤ 10
	-**eeship**	-ee	N	-ship	N	trust-ee-ship	≤ 10
	-**enment**	-en	V	-ment	N	Enlight-en-ment	≤ 10
	-**erish**	-er	N	-ish	ADJ	quak-er-ish	≤ 10
	-**ianness**	-(i)an	ADJ	-ness	N	Christ-ian-ness	≤ 10
	-**iless**	-y	N	-less	ADJ	merc-i-less	≤ 10
	-**ingless**	-ing	N	-less	ADJ	mean-ing-less	≤ 10
	-**isty**	-ist	N	-y	N	tour-ist-y	≤ 10
	-**lihood**	-ly	ADJ	-hood	N	live-li-hood	≤ 10
	-**mentous**	-ment	N	-ous	ADJ	fila-ment-ous	≤ 10
	-**orish**	-or	N	-ly	ADJ	act-or-ish	≤ 10
	-**thful**	-th	N	-ful	ADJ	heal-th-ful	≤ 10

Table 3. Non-existing suffix combinations used in the experiments reported in § 3

	Non-existing combinations used in task			
	Permutation	**Existing combination**	**Manipulation**	**Existing combination**
From used existing combinations	-izic	-icize	-adive	-ative
	-ationize	-ization	-liress	-liness
	-ablify	-ifiable	-isiable	-ifiable
	-ariment	-mentary	-tsful	-thful
	-itior	-ority	-osify	-osity
	-lessdom	-domless	-aguous	-ageous
	-ousment	-mentous	-ieship	-eeship
	-ishor	-orish	-emment	-enment
From unused existing combinations	-istion	-ionist	-ausness	-ousness
	-ionalate	-ational	-iunism	-ionism
	-istal	-alist	-ionier	-ioneer
	-mentish	-ishment	-ifilation	-ification
	-istage	-agist	-oryst	-orist
	-izive	-ivize	-iarship	-(i)anship
	-nessic	-icness	-oful	-iful

All suffix combinations used in this study as well as their type frequency were additionally verified using the OneLook Dictionary database (available at: https://www.onelook.com/), which contains more than 19 million words (more than 1000 online dictionaries are indexed by the OneLook search engine). OneLook can be used as a reverse dictionary and allows for wildcard search. All searches for evaluation of productivity of suffix combinations were manually verified by a native English-speaking linguist.

3. Two psycholinguistic experiments

To probe the status of suffix combinations in the mental lexicon we ran two experiments: one with native and one with non-native speakers of English. The experiments consisted of an identification task similar to the lexical decision task (Meyer & Schvaneveldt 1971); however, the participants did not see whole words but two-suffix combinations such as *-mentary* (formed from *-ment* + *-ary*), *-ageous* (from *-age* + *-ous*), etc. Appendix 1 contains the list of stimuli in the order in which they were presented to the participants. The subjects had to decide as soon as possible whether the form they saw was the termination of an English word or not.

3.1 Participants

Participants in the study comprised 45 native and 30 advanced non-native speakers of English, who self-identified as having a CEFR level of C1 ($n = 16$) or C2 ($n = 14$). Of the native English speakers, 11 were male and 34 were female, median age 27. Thirty-three were citizens of the US, 10 of the UK, and one participant each of Australia and Malta. Three listed a second native language of either Spanish or French; the rest only listed English as a native language. Twenty-seven listed only English as a spoken language, while 18 listed one or more additional non-native spoken languages with at least intermediate (B1 or higher) proficiency. Of the non-native English speakers, 12 were male and 18 were female, median age 27. Their native languages and nationalities varied, with a majority ($n = 25$) listing either a Romance (French, Spanish, Romanian) or Germanic (Dutch or German) native language, while the remaining five participants listed a Slavic (Serbian, Russian, Slovenian), Uralic (Hungarian) or Mongolic (Mongolian) native language. Most non-native English-speaking participants ($n = 25$) listed another spoken non-native language with at least intermediate proficiency in addition to English and their native language.

3.2 Stimuli

Sixty suffix combinations were presented to participants, 30 of which exist in English and 30 of which do not (see Tables 2 and 3, and Appendix 1). Of the existing combinations, 15 were productive (as defined in § 2) and 15 unproductive. Of the non-existing combinations (shown in Table 3), 15 were created from a permutation of an existing combination (reversing the order of the two suffixes such that the combination was not possible in English), and 15 were created through a spelling manipulation of an existing combination (changing one letter from an existing combination such that the new form does not exist in English). No non-existing combinations included any phonological and/or orthographical impossibilities in English.

3.3 Testing method

The task was presented via an online questionnaire developed using Google Forms that was completed on a voluntary basis. The questionnaire contained three sections: the first was a series of general demographic questions, including age, gender, nationality, native language(s), English fluency level, other languages spoken, level of education, and experience in a linguistic or other language-related field. The second was a practice test to familiarize participants with the nature of the main task and ensure that they understood it properly. The training included 10 example suffix combinations, five of which were productive existing combinations and five of which were non-existing combinations. These suffix combinations were not from the test stimuli. The third section was the main task, in which the 60 suffix combinations described above were presented in a randomized order, and participants were then asked to identify which of the combinations exist and which do not exist as word terminations in English. Participants were given a 10-minute time limit to complete the main task.

3.4 Data analysis

We used independent t-tests to consider possible significance of overall scores for individual groups and between native and non-native speakers, as well as for each stimulus type (i.e., existing vs. non-existing, productive vs. unproductive, and manipulation vs. permutation). As is typical practice in psycholinguistics, participants that were above or below 2.5 standard deviations from the mean score for all participants were excluded from the analysis. This exclusion is based on the assumption that if a participant makes mistakes above 2.5 SD, s/he was not focused on the task. One native English-speaking participant was eliminated from the data for this reason. The data of three other native English-speaking participants were not considered in the analysis due to exceeding the allotted time limit, leaving 41 native English speakers whose data were analyzed. Likewise for stimuli, if the accuracy score of an individual stimulus was more than 2.5 SD below the mean for all stimuli, this indicates that there could be something wrong with that stimulus. The unproductive existing combination *-ianness* (as in *Christ-ian-ness*) was the only stimulus eliminated from the data for this reason. *-ianness* had an accuracy score of more than 2.5 SD below the mean for all stimuli in both native and non-native speaker groups.

Additional possible affective variables for which data was gathered via the demographic questions (e.g., age, gender, education level, socioeconomic status, and linguistics experience) were also tested for possible significance using independent t-tests. The results of these tests are reported below.

3.5 Results

Both native and non-native speaker groups showed a strong ability to recognize existing suffix combinations and differentiate them from non-existing combinations, with an overall mean score of 46.29 (79%) for the native speaker group and 44.73 (76%) for the non-native speaker group. Although native speakers showed slightly higher overall scores, no significant difference was found between the native and non-native speaker groups (p-value = 0.19). Similar to previous iterations of this study using other test languages, productive combinations were the most accurately recognized of the four types (Figure 1), with both groups showing a mean score of 92% for identifying this type of stimulus. The other three types had more evenly distributed accuracy levels, with the only significant differences being found between productive and unproductive combinations in both groups (p-value < 0.001).

One question concerning how participants come to an answer refers to the time required to complete the test. While we were not able to measure reaction times to individual stimuli, we did obtain information about the duration of the main task: participants indicated the time immediately before beginning the main task and this was compared with the time of submission of the questionnaire. As was noted in § 3.4, three native-speaker participants were eliminated from the analysis because their total time exceeded the 10-minute time limit for the task. While no significance was found between time and overall accuracy for native speakers, non-native speakers tended to perform slightly better when taking longer to complete the test (p-value = 0.04). The results are summarized in Figure 2.

As can be seen from Figure 2, the majority of the participants needed about four minutes to complete the task (identification of 60 existing and non-existing suffix combinations), which means an average of 4s per suffix combination. We interpret this fact as evidence that the participants responded intuitively and did not relate the stimuli to whole words.

Finally, non-native speakers with a Romance or Germanic first language tended to perform better than those from other language groups. All other variables considered showed no significant differences in either group.

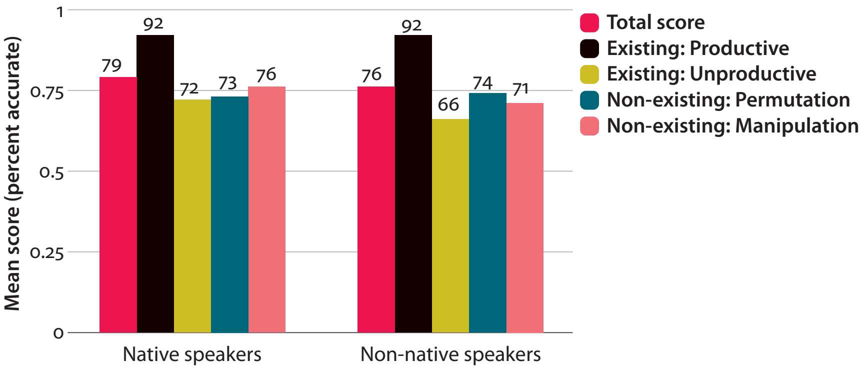


Figure 1. Mean scores for native and non-native speakers

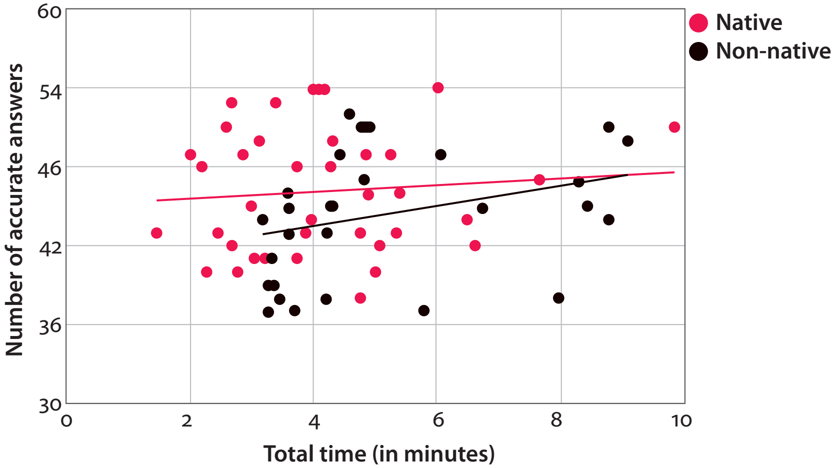


Figure 2. Individual scores and time required to reply to all stimuli

4. Discussion

Situated in the psycho- and neurolinguistic literature, our research relates to pioneer work such as the 'affix stripping model' of Taft and colleagues (Taft & Forster 1975, 1976; Taft 1979) who provided evidence that not only stems but also affixes (prefixes in their case) are stored in the mental lexicon; recent neurolinguistic studies also support this finding: Fruchter, Stockall & Marantz (2013), Fruchter & Marantz (2015) demonstrate that speakers obligatorily decompose the (visual) stimulus into morphemes, look these up in the mental lexicon, and recombine them. However, of all recent research on the topic, the closest to our experiments seems to be Lázaro, Illera & Sainz (2016) who investigate the effect of suffix priming on the lexical decision of suffixed (ero-JORNALERO) and pseudosuffixed (ero-CORDERO) Spanish words, as well as the effect of orthographic priming on nonsuffixed words (eba-PRUEBA). Their results show that in the case of suffixed and pseudosuffixed words, related primes significantly accelerate response latencies in comparison to unrelated primes (ista-JORNALERO; ura-CORDERO). Importantly, there was no facilitation effect of the orthographically related prime for simple words in comparison to the unrelated prime (eba-PRUEBA; afo-PRUEBA). Thus, the study concludes that the priming effect of suffixes is not orthographic but morphological, i.e., the effect holds only for derived and pseudo-derived words (such as *corn:er* in English; *corner* is a pseudo-derivative because *corn* is not its root). In our experiments we went further: we presented the participants with only suffix combinations (purely morphologically-derived items) and encouraged them to identify the stimuli as existing and non-existing pieces of word structure without relating them to whole words. The participants recognized existing and non-existing suffix combinations with very high accuracy, which we interpret as evidence that not only affixes but also affix combinations are stored in the mental lexicon.

As a rule, English language pedagogy, both for native and non-native speakers, does not include information about the combinability of the English derivational suffixes. In other words, the participants in our experiments have never learned English suffix combinations in a systematic way. Also, theories of morphology do not allow suffix combinations to have a life of their own, that is, according to grammatical theory all derivations in a language start with either a root, stem or word to which then affixes are attached. Yet, in an experiment trial, native and advanced non-native speakers of English do not need to see lexical bases to decide which suffix combinations exist and which do not. The only plausible explanation of why the participants in our experiments successfully accomplished a task they are not supposed to be able to solve is that they have extracted the necessary knowledge about the combinability of English suffixes subconsciously.

In the processing of suffix combinations, productivity seems to play a major role and productive combinations are more readily recognized than unproductive ones. The effect of productivity is well known in the psycholinguistic literature and we will not pay special attention to this issue but to how we established the productivity of a suffix. Almost all approaches to productivity of derivational suffixes control for token frequency because the more a subject sees or hears a token, the easier its recognition. However, a reliable estimation of token frequency requires linguistic resources that contain a large amount of (appropriately annotated) data (tokens), which is problematic for some languages. In this study, we defined productivity in an alternative way, based on type frequency (as discussed in § 2). We assumed that suffix combinations that derive more than 10 types are productive, while those that derive 10 or fewer types are unproductive (see the distribution of the suffixes in Table 2) and we found that the difference in processing of productive and unproductive combinations is statistically significant (p-value < 0.001). We see this result as a confirmation of our productivity criterion and believe to have suggested a useful and easy way to estimate productivity of a suffix in word formation.

This study is part of a series of experiments on the processing of suffix combinability in languages belonging to different language families such as Slavic, Germanic and Romance, and it confirms what has been established so far for the processing of suffix combinations by native speakers of Slavic (Polish and Slovene) and Romance (Italian) languages, who also recognized existing and non-existing combinations surprisingly accurately (Manova, ms); in addition, productivity (as defined in this study) plays a role in suffix combination processing. Thus, although the results of the present study indicate that non-native speakers with a Romance or Germanic first language tend to perform better in the recognition of English suffix combinations than those with a Slavic first language, all participants seem to know by heart a significant portion of the suffix combinations of their native languages. Taken together, the results of the experiments reported here and the results for Polish, Slovene and Italian obtained in previous experiments suggest that native and native-like language competence presupposes knowledge of suffix combinability with a morphomic access to suffix combinations in the mental lexicon.

Theoretical morphology research on rule conflation (Stump 2019) provides indirect support to what we observed for the organization of the mental lexicons of native and non-native speakers of English with respect to suffix combinability. Table 4 is a comparison of rule conflation and suffix combination:

Table 4. Rule conflation versus suffix combination (Data adapted from Stump 2019)

Base	**Rule conflation**	**Suffix combination**
whimsy, nonsense	*whimsic, *nonsensic	*cycle* → *cycl-ic*
	whims-ical, nonsens-ical	*cycl-ic-al*
probable, simple	*probabilist, *simplist	*national* → *national-ist*
	probabil-istic, simpl-istic	*national-ist-ic*
beauty, mort	*beautic, *mortic	*academy* → *academ-ic*
	beaut-ician, mort-ician	*academ-ic-ian*

As can be seen from Table 4, in the case of suffix combination, the two suffixes attach in two steps, while in the case of rule conflation, we deal with a single suffix that we see as attaching directly. Since we are interested in suffix combinations, our Table 2, the list of all existing combinations used in the psycholinguistic experiments, does not include rule conflation patterns. Nevertheless, we see rule conflation as important evidence for memorization of two-suffix combinations and for their use as ready-made structures, which are defining properties of the morphome.

Further, morphomic patterns appear highly relevant to computational linguistics (computational extraction of grammatical rules from (unannotated) raw text, e.g., within Unsupervised Learning of Morphology (ULM)) and Natural Language Processing (NLP) – including machine translation – based on neural networks, i.e., inspired by the organization of the human brain (see, e.g., Wu et al. 2016). ULM and NLP seem to employ strategies similar to the morphomic one we identified in native and advanced non-native speakers of English in relation to suffix combinability. ULM and neural network approaches take large amounts of raw text data as input and attempt to induce the patterns and rules of the input language or to translate one language into another. The procedure is entirely form-based (thus morphomic) and relies on sequence-to-sequence comparisons (sequences can be of any length) and weighting of substring frequencies in recurrent formations. An example of ULM logic: the frequency of the final substring *-ing* in English will be much greater than that of a random substring of the same length, and most of the words that have *-ing* will also get *-ed* much more often than chance (for an overview of computational approaches to morphology, see Hammarström & Borin 2011). The same logic applies to the identification of stems; as well as to suffix combinations: the frequency of, for example, *-ifier* (*-ify* + *-er*) will be much greater than that of a random substring of the same form, and words with *-ifier* will also combine with *-s* (to form plural) more often than chance. Facilitation and accuracy of rule extraction can be achieved by the addition of portions of annotated data to the raw data. For example, for morphological purposes, pairs such as <"bags", bag-PL> have been used (Kann & Schütze 2016 and Chrupala 2008, Chapter 6). Rules based on such annotations serve to make guesses when analyzing previously unseen data. For suffix combinations, one may use annotations such as <"-ifier", -ifier-NOUN from -ify-VERB> and <"-ifiers", -ifier-PL>; such annotations can be used, for example, to prepare corpora for word-formation research. As for semantics, it has been assumed that representations can also be extracted in an unsupervised manner through standard techniques of context-occurrence analysis (e.g., Deerwester et al. 1990; Mikolov et al. 2013), i.e., forms that are semantically related tend to co-occur.

Finally, in this study we did not measure reaction times to individual stimuli (only the duration of the completion of the task) and cannot provide fine-grade analyses of the response latencies. We are aware that our results will be regarded as only preliminary by some scholars and that more precise testing should be done. Psycholinguistic experiments with reaction time measuring are planned for future research, but we also hope that the findings of this paper will encourage other scholars to carry out similar or replication experiments. Additionally, to establish the exact relation between morphomic access of suffix combinations in the mental lexicon and language competence, non-native speakers with CEFR A and B (beginning and intermediate) levels of language skills should also be tested, and, of course, not only with English data.

5. Conclusion

Based on theoretical observations about the uniqueness of suffix combinations, we ran two online psycholinguistic experiments as part of a series of experiments on the processing of morphological structure in languages from different families

(Slavic, Germanic and Romance). To probe the status of suffix combinations in the mental lexicon, in the first experiment only native speakers of English were tested, while the second experiment was with advanced (C1 and C2 level) non-native speakers of English. Although knowledge of suffix combinability is not systematically acquired at school, and linguistic theories recognize only derivations that start from either a root (DM and lexical and incremental approaches to morphology) or from roots/stems/words (PFM and inferential approaches to morphology), in the experimental trials, native and advanced non-native speakers did not need to see lexical bases to differentiate between existing and non-existing suffix combinations. In the processing of suffix combinations, productivity plays a role and productive combinations were more readily recognized than unproductive ones. Since the task to identify suffix combinations as either existing or non-existing pieces of word structure was without any semantic cues and the average time for processing a suffix combination (approx. 4s) was too short to allow associations with whole words, we see two-suffix combinations as morphological structures listed in the mental lexicon in terms of purely morphomic patterns. Two-suffix combinations also attach as single suffixes, i.e., as readymade pieces of morphological structure, in rule conflation patterns, which provides further evidence for listedness in the mental lexicon and resembles the repurposing of morphomes in inflectional morphology. Computational approaches to grammar such as ULM and NLP models based on neural networks seem to employ a morphomic procedure similar to the one reported in this study for humans to induce (grammatical) patterns and rules and for machine translation.

The results of this research show that not only words, stems, roots and affixes but also affix combinations are of importance to the mental lexicon and natural language processing. Moreover, morphomic knowledge of suffix combinability seems to correlate with high language competence in humans and can thus be used for educational purposes, e.g., to facilitate foreign language learning, as well as for testing language proficiency.

Acknowledgements

We dedicate this chapter to the *morphomeman*, Mark Aronoff, and hope that it is original enough for the occasion.

We would also like to thank the editors for the invitation to contribute to this volume and for all the organisational work done, as well as the anonymous reviewers for accepting the paper as it was.

References

Allen, Margaret. 1978. Morphological investigations. Storrs, CT: University of Connecticut dissertation.

Aronoff, Mark. 1994. *Morphology by itself: Stems and inflectional classes*. Cambridge, MA: MIT Press.

Aronoff, Mark. 2012. Morphological stems: What William of Ockham really said. *Word Structure* 5. 28–51. https://doi.org/10.3366/word.2012.0018

Aronoff, Mark. 2016. Unnatural kinds. In Ana Luís & Ricardo Bermúdez-Otero (eds.), *The morphome debate*, 11–32. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198702108.003.0002

Aronoff, Mark & Nanna Fuhrhop. 2002. Restricting suffix combinations in German and English: Closing suffixes and the monosuffix constraint. *Natural Language & Linguistic Theory* 20. 451–490. https://doi.org/10.1023/A:1015858920912

Baayen, R. Harald. 1991. Quantitative aspects of morphological productivity. In Geert E. Booij & Jaap van Marie (eds.), *Yearbook of morphology* 1991, 109–149. Dordrecht: Kluwer.

Baayen, R. Harald. 1993. On frequency, transparency and productivity. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 1992*, 181–208. Dordrecht: Springer. https://doi.org/10.1007/978-94-017-3710-4_7

Baayen, R. Harald. 2001. *Word frequency distributions*. Dordrecht: Springer. https://doi.org/10.1007/978-94-010-0844-0

Baayen, R. Harald & Rochelle Lieber. 1991. Productivity and English derivation: A corpus based study. *Linguistics* 29. 801–843. https://doi.org/10.1515/ling.1991.29.5.801

Baayen, R. Harald, Richard Piepenbrock & Leon Gulikers. 1995. *CELEX2* LDC96L14. Web Download. Philadelphia: Linguistic Data Consortium.

Baerman, Matthew, Greville G. Corbett, Dustan Brown & Andrew Hippisley (eds.). 2007. *Deponency and other morphological mismatches*. Oxford: Oxford University Press. https://doi.org/10.5871/bacad/9780197264102.001.0001

Bauer, Laurie. 2001. *Morphological productivity*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486210

Chrupala, Grzegorz. 2008. Towards a machine-learning architecture for lexical functional grammar parsing. Dublin, Ireland: Dublin City University dissertation.

Deerwester, Scott, Susan T. Dumais, George W. Furnas, Thomas K. Landauer & Richard Harshman. 1990. Indexing by latent semantic analysis. *Journal of the American Society for Information Science* 41(6). 391–407. https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9

Fabb, Nigel. 1988. English suffixation is constrained only by selectional restrictions. *Natural Language and Linguistic Theory* 6. 527–539. https://doi.org/10.1007/BF00134491

Fruchter, Joseph & Alec Marantz. 2015. Decomposition, lookup, and recombination: MEG evidence for the Full Decomposition model of complex visual word recognition. *Brain and Language* 143. 81–96. https://doi.org/10.1016/j.bandl.2015.03.001

Fruchter, Joseph, Linnaea Stockall & Alec Marantz. 2013. MEG masked priming evidence for form-based decomposition of irregular verbs. *Frontiers in Human Neuroscience* 7. 1–16. https://doi.org/10.3389/fnhum.2013.00798

Giegerich, Heinz. 1999. *Lexical strata in English: Morphological causes, phonological effects*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486470

Halle, Morris & Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In Keyser Hale & Samuel Jay Keyser (eds.), *The view from Building 20*, 111–176. Cambridge, MA: MIT Press.

Hammarström, Harald & Lars Borin. 2011. Unsupervised learning of morphology. *Computational Linguistics* 37(2). 309–350. https://doi.org/10.1162/COLI_a_00050

Hay, Jennifer. 2001. Lexical frequency in morphology: Is everything relative? *Linguistics* 39. 1041–1070. https://doi.org/10.1515/ling.2001.041

Hay, Jennifer. 2002. From speech perception to morphology: Affix ordering revisited. *Language* 78. 527–555. https://doi.org/10.1353/lan.2002.0159

Hay, Jennifer. 2003. *Causes and consequences of word structure*. London: Routledge.

Hay, Jennifer & Ingo Plag. 2004. What constrains possible suffix combinations? On the interaction of grammatical and processing restrictions in derivational morphology. *Natural Language & Linguistic Theory* 22. 565–596. https://doi.org/10.1023/B:NALA.0000027679.63308.89

Kann, Katharina & Hinrich Schütze. 2016. MED: The LMU system for the SIGMORPHON 2016 shared task on morphological reinflection. In *Proceedings of the 14th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, 62–70. Association for Computational Linguistics. https://doi.org/10.18653/v1/W16-2010

Kiparsky, Paul. 1982. Lexical morphology and phonology. In In-Seok Yang (ed.), *Linguistics in the morning calm: Selected papers from SICOL-1981* (Volume 1), 3–91. Seoul: Hanshin.

Lázaro, Miguel, Victor Illera & Javier Sainz. 2016. The suffix priming effect: Further evidence for an early morpho-orthographic segmentation process independent of its semantic content. *Quarterly Journal of Experimental Psychology (Hove)* 69(1). 197–208. https://doi.org/10.1080/17470218.2015.1031146

Luís, Ana R. & Ricardo Bermúdez-Otero (eds.). 2016. *The morphome debate*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198702108.001.0001

Maiden, Martin. 2016. Some lessons from history: Morphomes in diachrony. In Ana R. Luís & Ricardo Bermúdez-Otero (eds.), *The Morphome Debate*, 33–63. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198702108.003.0003

Manova, Stela. 2010. Suffix combinations in Bulgarian: Parsability and hierarchy-based ordering. *Morphology* 20. 267–296. https://doi.org/10.1007/s11525-010-9148-3

Manova, Stela. 2011. A cognitive approach to SUFF1-SUFF2 combinations: A tribute to Carl Friedrich Gauss. *Word Structure* 4. 272–300. https://doi.org/10.3366/word.2011.0014

Manova, Stela. 2015. Affix order and the structure of the Slavic word. In Stela Manova (ed.), *Affix ordering across languages and frameworks*, 205–230. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190210434.003.0009

Manova, Stela. Manuscript. *Suffix combinations in English, Italian, Polish and Slovene: Theory and processing*. University of Vienna.

Manova, Stela & Mark Aronoff. 2010. Modeling affix order. *Morphology* 20. 109–131. https://doi.org/10.1007/s11525-010-9153-6

Manova, Stela & Luigi Talamo. 2015. On the significance of the corpus size in affix-order research. *SKASE Journal of theoretical linguistics* 12(3). 369–397.

Manova, Stela, Harald Hammarström, Itamar Kastner & Yining Nie. 2020. What is in a morpheme? Theoretical, experimental and computational approaches to the relation of meaning and form in morphology. *Word Structure* 13(1). 1–21. https://doi.org/10.3366/word.2020.0157

Meyer, David & Roger Schvaneveldt. 1971. Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology* 90. 227–234. https://doi.org/10.1037/h0031564

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado & Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In Chris J. C. Burges, Léon Bottou, Zoubin Ghahramani, & Kilian Q. Weinberger (eds.), *Advances in neural information processing systems 26 (NIPS 2013)*, 3111–3119. Neural Information Processing Systems.

Mohanan, Karuvannur Puthanveettil. 1986. *The theory of lexical phonology*. Dordrecht: Reidel.

O'Neill, Paul. 2014. The morphome in constructive and abstractive models of morphology. *Morphology* 24. 25–70. https://doi.org/10.1007/s11525-014-9232-1

Plag, Ingo. 1996. Selectional restrictions in English suffixation revisited. A Reply to Fabb (1988). *Linguistics* 34. 769–798. https://doi.org/10.1515/ling.1996.34.4.769

Plag, Ingo. 1999. *Morphological productivity: Structural constraints in English derivation*. Berlin: Mouton De Gruyter.

Plag, Ingo. 2002. The Role of selectional restrictions, phonotactics and parsing in constraining suffix ordering in English. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 2001*, 285–314. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-017-3726-5_11

Plag, Ingo & Harald Baayen. 2009. Suffix ordering and morphological processing. *Language* 85. 109–152. https://doi.org/10.1353/lan.0.0087

Selkirk, Elisabeth. 1982. *The syntax of words*. Cambridge, MA: MIT Press.

Siegel, Dorothy. 1974. *Topics in English morphology*. Cambridge, MA: MIT Press.

Stump, Gregory T. 2001. *Inflectional morphology*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486333

Stump, Gregory T. 2019. Evidence for rule conflation. Paper presented at The 4th American international Morphology Meeting (AIMM4), Stony Brook, New York, May 3–5.

Taft, Marcus. 1979. Recognition of affixed words and the word frequency effect. *Memory & Cognition* 7(4). 263–272. https://doi.org/10.3758/BF03197599

Taft, Marcus & Kenneth Forster. 1975. Lexical storage and the retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior* 14. 630–647. https://doi.org/10.1016/S0022-5371(75)80051-X

Taft, Marcus & Kenneth Forster. 1976. Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior* 15. 607–620. https://doi.org/10.1016/0022-5371(76)90054-2

Talamo, Luigi. 2015. Suffix combinations in Italian: Selectional restrictions and processing constraints. In Stela Manova (ed.), *Affix ordering across languages and frameworks*, 175–204. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190210434.003.0008

Wu, Yonghui, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun et al. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. Technical Report, arXiv:1609.08144v2 [cs.CL]

Appendix 1. List of suffix combinations (existing and non-existing) used as stimuli

Stimulus	Existing	Non-existing	Stimulus	Existing	Non-existing
-ority	prod(uctive)		**-oryst**		manip(ulation)
-ership	prod		**-tsful**		manip
-izic		perm(utation)	**-ative**	prod	
-lessness	prod		**-ionier**		manip
-adive		manip	**-ariment**		perm
-ishness	prod		**-ivist**	prod	
-erish	unprod		**-iless**	unprod	
-ifilation		manip	**-iarship**		manip
-lessdom		perm	**-mentary**	prod	
-itior		perm	**-ianness**	unprod	
-liness	prod		**-mentous**	unprod	
-iunism		manip	**-lihood**	unprod	
-ifiable	prod		**-istal**		perm
-ingless	unprod		**-domless**	unprod	
-thful	unprod		**-ishor**		perm
-eeship	unprod		**-mentish**		perm
-aguous		manip	**-ionalate**		perm
-izive		perm	**-ausness**		manip
-istion		perm	**-osity**	prod	
-nessic		perm	**-istage**		perm
-ationize		perm	**-eeism**	unprod	
-ivate	prod		**-ablify**		perm
-osify		manip	**-isty**	unprod	
-ional	prod		**-ieship**		manip
-orish	unprod		**-emment**		manip
-ageous	unprod		**-ization**	prod	
-enment	unprod		**-atee**	unprod	
-liress		manip	**-ousment**		perm
-ifier	prod		**-icism**	prod	
-isiable		manip	**-oful**		manip

PART V

Interfaces

CHAPTER 18

A short history of phonology in America

Plus c'est la même chose, plus ça change

Stephen R. Anderson
Yale University

Although awareness of a difference between the study of the sound patterns of particular languages and the study of the language-independent capacity of humans to produce and perceive sound existed in European and American thought in the early years of the twentieth century, a clear enunciation of the distinction between phonology and phonetics is due to Otto Jespersen in 1924. As phonology became established as a coherent object of inquiry, two themes can be identified in theorizing about it: first, the question of whether phonological structure is in the mind, as an aspect of human cognition, or only a set of facts about the external data of language; and second, the question of whether there are valid universals of phonological structure. These issues are traced across the past century in the work of American linguists. An additional factor identifiable in historical shifts in theoretical perspective is somewhat less principled: as discussion of fundamental issues becomes more technical and relevant data harder to identify, students and scholars looking for productive research topics tend to abandon previous frameworks for others in a search for lower hanging fruit without necessarily having resolved the earlier questions.

Keywords: cognitive science, constraints, generative phonology, phonemics, rules, universals

Although the serious study of language goes back millennia, 'Phonology' as a distinct subdiscipline of linguistics is an idea with less than a century of history.[1] We

1. I note at the outset that after some initial comments, I will focus below primarily on the history of phonology in recent American linguistics, not because other traditions have nothing to say about the matters under discussion, but rather because they have too much, and a more general treatment would risk losing the thread of a particular interesting story. Also, although phonological notions are applicable to signed as well as spoken languages, I limit myself here to the latter as constituting the center of the developments described.

https://doi.org/10.1075/cilt.353.18and

can begin with the distinction underlying the emergence of this sub-field from the more general study of sound in natural languages.

Phonetics, on the one hand, I take to be the study of sound related phenomena as these play a role in natural language, but from a point of view that is independent of the properties of any particular language. It is the study of the range of properties of utterances that distinguish one possible human utterance from another, independent of any particular language. Phonetics is thus concerned with the range of human articulatory capabilities, the acoustic consequences of articulatory gestures, and the way acoustic events affect the auditory system. In contrast, phonology is the study of the way sound properties and distinctions function within the systems of individual languages: the study of their sound patterns as an aspect of their particular grammars.

Section 1 below briefly surveys the understanding of 'phonology' in the years prior to the rise of generative views in the late 1950s. Section 2 then discusses the emergence of generative phonology and the associations between that point of view and two important issues: the notion of linguistic structure as an aspect of the mind, and the existence of universals of phonological structure. Section 3 then surveys some of the reactions to Chomsky & Halle (1968) and their parallels to developments in the philosophy of mathematics, and the eventual replacement of the rule-based views of that work with theories based on simultaneously applicable constraints. Over many years, the replacement of one theoretical emphasis by another is seen as motivated not exclusively by the search for solutions to continuing problems, but also by the needs of emerging scholars for immediately significant results. Section 4 concludes with the suggestion that this motivation may still be at play today.

1. The early history

The word 'phonology' itself shows up in literature well before what we think of as the period of modern linguistics: the first citation in the OED is from 1798. In general, though, it does not have a precise meaning beyond something to do with the study of sound in language. Confusingly, de Saussure (1916) uses the word (or its French equivalent), but not in the modern sense: he uses *phonologie* as essentially what we call phonetics, and *phonétique* for the historical study of sounds: he had no term for the systematic study of language-particular synchronic sound systems.

It is in the writings of Otto Jespersen that we first find a proposal for a clear delineation of the sort we assume today: "It would, perhaps, be advisable to restrict the word 'phonetics' to universal or general phonetics and to use the word *phonology* of the phenomena peculiar to a particular language (e.g., 'English Phonology')", (Jespersen 1924: 35).

Regardless of the words used to describe this difference, however, it was not unfamiliar to serious students of language, and there was quite a lot of interesting work in the European tradition. This might be seen as beginning with Jan Baudouin de Courtenay (1895 [1972]), whose distinction between 'anthropophonics' – essentially physical phonetics – and 'psychophonetics' comes quite close to our contemporary understanding. But this tradition was little known and quite without influence in American linguistics at the beginning of the 20th century. Ferdinand de Saussure is generally thought of as more central to the development of the field, but he too was not at all well known in America in the 1920s, and in any event his ideas of phonological structure are somewhat harder to discern than is sometimes supposed (Anderson 1985: 33–55).

In America, Franz Boas represents a major strand in distinctively American thought about language in these early years, but as far as sound was concerned, he was mostly concerned with accurate phonetic transcription, rather than with things like sound patterns (Anderson 1985: 204–216). Efforts have been made to find support in Boas' practice for a notion of phonological representation distinct from surface phonetics (Postal 1964), but this interpretation is hard to sustain. Indeed, it is interesting to note that when his principal Kwakw'ala consultant George Hunt sent him textual material transcribed in a way that abstracted away from certain phonetic differences predictable within the language, Boas 'corrected' this so as to restore surface phonetic accuracy (Anderson 1985: 206). Boas' descriptive practice does reveal a certain amount of interest in potentially universal properties of sound structure, a topic which will occupy us below in § 2.2, but these ideas were quite inexplicit.

The first significant piece of work in America dealing with phonology as we understand it was Sapir's (1925) paper "Sound Patterns in Language," which appeared in the first issue of *Language*. Sapir argued not only that such structure is quite apart from the concrete properties of speech sounds, but that it is a fundamentally psychological notion, in the sense we would today call 'cognitive'. Elements have their place in such a structure on the basis of complex patterns of associations and not on the basis of external, objective properties. We can see in this paper the beginning of the notion that the study of language, and phonology in particular, is a part of Cognitive Science (Anderson 2001).

Reaction to that conception was not long in coming, however: the very next paper in the same inaugural issue of *Language* was Albert P. Weiss's (1925) article "Linguistics and Psychology," which resolutely rejects the significance of a notion of 'mind' and insists that all 'psychological' study of language must be grounded in (and limited to) externally observable properties and phenomena. Weiss's call was enthusiastically taken up by his new colleague at Ohio State, Leonard Bloomfield, and in the scientific atmosphere of the time, Sapir's conception was displaced for the next several decades by a very different (and non-cognitive) notion of phonemics – a notion that was at the heart of nearly all American theorizing about language during this period, which tended to take phonemics as its model for the study of other components of linguistic structure.

In theorizing about sound structure on a path following Sapir's early ideas, we see that the Weiss/Bloomfield conception overwhelmingly held sway in the development of phonemics. Bloomfield's (1933) Language represents the classical period of structuralist thinking about sound structure. This is a picture focused on surface phonetic form, with the abstract notion of the phoneme best thought of not as a cognitive reality in itself, as Sapir had urged, but rather as a set of purely phonetic segments, united by statements of their respective distributions. This way of thinking predominated up until around 1960, when it was fairly abruptly replaced by quite a different point of view.

2. The rise of generative phonology

A new way of approaching sound structure which we can see as that of 'phonology' in today's sense was ushered in by the publication of Morris Halle's (1959) *Sound Pattern of Russian* and "Phonology in Generative Grammar" (Halle 1962). These works challenged the fundamental assumptions of structuralist phonology, and the approach they advocated quickly took over the field. Following a heated exchange between Fred Householder (1965) and Chomsky & Halle (1965), the ascendancy of the generative point of view was essentially complete by the time its codification in Chomsky & Halle (1968) appeared.

What provoked the fundamental re-orientation of phonological thinking associated with the replacement of structuralist phonemics by generative phonology and its descendants? The usual picture is that in the late 1950s, Halle (1957) discussed some interesting facts about voicing in Russian obstruents. In this language, voicing is distinctive for some obstruents but not for others, but voicing assimilation is quite general across all obstruents in clusters. Halle argued that a description interposing a phonemic representation between a morphophonemic one and the surface phonetic form necessarily loses the unitary nature of the generalization about voicing assimilation. Presented with that evidence, the story goes, the field rapidly converged on the view that structuralist phonemic representations were a bad idea, and replaced them with the more abstract representations we have since come to know and (mostly) love.

This story leaves out some important history, though (Anderson 2000). In fact, examples just like Halle's had been discussed before, and taken to lead to very different conclusions. Bloch (1941), for example, discussed essentially similar facts in the distribution of vowel length in some dialects of American English, but instead of concluding that phonemic representations were a bad idea, used the facts to argue that the rigor of phonemic theory could save the linguist from being misled by a seductive apparent regularity.

The force of Halle's argument, in fact, lay not in the nature of the facts, but in the use Halle made of them. He argued essentially that a description of a language's phonology had to do more than simply get the forms right – it also had to capture the regularities that govern the distribution of those forms: the 'rules' as well as the 'representations'. Of course, this line of reasoning only makes sense if we assume that the rules are really part of the language, and that in turn only makes sense if we assume that our description has to encompass the knowledge speakers have. And that, of course, is just what phonemic theory was opposed to, based as it was on essentially behaviorist notions according to which only external observables, and not ineffable notions like 'the mind' were the province of genuine science.

A key component in the striking success of the generative point of view on phonology, then, was its appeal to linguistics as what would later be called Cognitive Science, and in the resurgence of talk of language as something in the mind, not just out there in the products of speech. But this was not the whole story.

Another contributor to this development was surely the trend in science observed by Max Planck (1949: 33f.): "A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it." That is, scientific progress generally occurs not horizontally, by transmission from one established figure to another, but rather vertically, by the influence of an innovator on students. And this was definitely a factor in the rise of generative views. As Searle (1972) put it, "All of this facilitated a rapid dissemination of new ideas and a dramatic and visible clash of conflicting views. Chomsky did not convince the established leaders of the field but he did something more important, he convinced their graduate students." Searle refers here to Chomsky's influence in the emergence of new views on syntax, but the point is also applicable to Halle's role in the development of phonology.

Planck's comment on scientific progress is generally taken to require the actual death of proponents of an earlier view: it is commonly paraphrased as "Science advances one funeral at a time." Generational change of the implied sort can be precipitated without this, though. In the specific case of phonemic theory, we can observe that by the 1950s, the range of available research questions had become quite constrained. The overall theory of the phoneme was considered a settled matter, and the theoretical thrust of descriptive studies restricted at best. The limitations of the research horizon in phonology were somewhat exacerbated by the methodological rigor and strict procedural bias of the main structuralist phonologists (Harris 1951), together with their insistence that phonemic analyses must be arrived at without reference to any other aspect of language. Significant scholars who argued otherwise, such as Kenneth Pike (1947a,b) with his insistence that all areas of language were to be investigated in their mutual relation, were essentially marginalized in discourse about phonemic theory.

Some remaining controversies, such as the correct analysis of suprasegmental properties such as levels of stress in English, seemed quite intractable, and the structuralist approach to these phenomena in terms of phonemes was seriously undermined by work in what would become the new theoretical framework (Chomsky, Halle & Lukoff 1956). As a result of all of these factors, graduate students searching for dissertation topics were more than ready to have opened before them a quite distinct view that, by discarding the specific limitations of classical phonemics, made available a rich array of new research questions. Some of the success of generative phonology, then, must undoubtedly be attributed to a degree of stagnation in structuralist theorizing.

2.1 Phonological structure as an aspect of the mind

In intellectual terms, we can identify two 'big picture' issues important to the rise of generative phonology on which positions have shifted over the years. One of these is the question of whether phonological structure and regularity resides in the mind or in the external productions of speakers.[2]

Sapir and his students, together with most adherents of generative phonology and its various descendants, would maintain that the description of a phonological system is the description of a fundamentally cognitive reality. In contrast, following the lines of behaviorists, American structuralists sought such structure in the sounds themselves. And indeed some recent views, becoming more prominent even as I write, represent in some ways a return to such a concern with the external manifestations of language, and a rejection (or at least considerable weakening) of the view that phonological structure is an aspect of the mind and the biological language faculty.

2. For a vastly more comprehensive study of the role of notions of the mind and cognition in the fields of philosophy, psychology and linguistics, including the relations among these fields in the years leading up to the emergence of generative phonology and syntax, see Goldsmith & Laks (2019).

When we look at the history of notions in the field of psychology that are relevant to language, we can date the emergence of an important perspective to the publication of John B. Watson's (1924) book *Behaviorism*. This point of view, which we saw already in Weiss's (1925) article in *Language*, became the dominant paradigm in psychology for quite a number of years (especially as later taken up by B. F. Skinner), almost completely eclipsing Sapir's approach among linguists.

Matters changed significantly with Chomsky's (1959) very influential review of Skinner's (1957) book *Verbal Behavior*, which pointed out various ways in which the behaviorist perspective was unable to deal with the observed facts of how language is acquired and used. This review had an effect not only on thinking about language but more broadly, and it led to the rapid decline of 'behaviorist' psychology more generally and the rise of 'cognitivist' views. That in turn prepared the way for the revival of mentalist notions specifically in the study of language, as represented by the emergence of generative phonology.

Actually, the decline of behaviorism in psychology was in some ways more apparent than real: a remarkable number of psychologists today, even if they reject being labeled as "behaviorists", nonetheless see the primary object of inquiry in psychology as an understanding of behavior, something to which talk of 'minds' may contribute, but which is primarily about understanding external manifestations. But an analysis of these matters falls outside the scope of the present chapter.

2.2 The place of universals

A second intellectual theme, characterized by a similar polarity, is the matter of universals: whether these are real or just epiphenomena resulting from the limited range of languages we have to explore; and if they are real, what their status is. Do the regularities we uncover across languages follow from deeply grounded principles of Universal Grammar, or are they simply the natural outcome of more external events?

While Boas is historically associated with the notion that every language must be approached on its own terms, and thus with a rejection of universals, this is probably an over-simplification. Actually Boas had a rather strong, if implicit, view of universal constraints determining the content of phonological systems (Anderson 1985: 202ff.), but that was never really expressed as a theoretical principle, and at any rate had no particular theoretical status in the years when phonology was first coming to be seen as distinct from phonetics.

Sapir, however, already in his little book *Language* (Sapir 1921), rejected the notion that there are real, substantive universals of language. Sapir draws this in part from the Boas tradition of emphasizing the study of languages on their own terms (while neglecting Boas' rather strong views about typological frameworks for linguistic description, which imply a strong set of substantive universals), and in part from the liberal politics of the time, which emphasized 'nurture' at the expense of 'nature' in accounting for the properties of individuals. Unlike his views on language as an aspect of the mind, Sapir was more in tune with the way the field would develop in rejecting language universals and attributing observed cross-linguistic regularities to factors of culture and historical transmission. This would be true at least until the 1960s, when the re-emergence of nativist views of innate structure resulted in the ascendance of strong theories of linguistic universals.

In the rise of generative phonology, we can see a resurgence of the idea that there are strong universals governing linguistic structure (and phonology in particular), principles that are grounded in the biological nature of the human language faculty. The embedding of facts of particular languages in a framework of universals of linguistic structure is certainly the approach associated with almost all strains of generative phonology, up to and including Optimality Theory.

More recently, though, we see some scholars staking out positions that are more skeptical about the status of apparent universals: for example, Juliette Blevins's (2004) arguments that these are just the consequences of independent principles of linguistic change, whose outcomes show regularities that are not due to any innate faculty of UG or the like. Similarly, Jeff Mielke (2008) has argued that even the basic descriptive framework of the feature system is not grounded in biologically determined universals, but just follows from the range of classes that arise in phonological regularities as a consequence of change and related factors; see also the papers in Clements & Ridouane (2011). These and other writers assert that an adequate understanding of observed cross-linguistic regularities can be obtained on the basis of statistical analyses of surface phenomena without an appeal to a rich human faculty of language whose characteristics determine and constrain the substantive content of linguistic systems, in phonology as in other areas of grammar.

In contrast, the most recent major innovation in phonological theory, Optimality Theory (as discussed below in § 3.3) takes a strong position on universals. The collection of markedness and faithfulness constraints that constitute the core of a grammar in this system are asserted to be universal in nature, with only their relative ranking subject to language-particular specification. The effectiveness of this claim is open to question, but the commitment to it on the part of theoreticians is clear, indicating that the issue of phonological universals is far from a settled matter.

3. Phonological theory after Chomsky & Halle (1968)

The appearance in print of Chomsky & Halle (1968) after many years of circulation in *samizdat*' form was a major landmark in the development of phonological theory. By no means simply a description of the "sound pattern of English", the work presented a more or less definitive formulation of the theory of generative phonology as this had emerged over the preceding decade or so. As such, it offered a clearer target for criticism of the underlying assumptions of that theory than had previously been available, and the field took up the challenge with enthusiasm.

3.1 The logicism of SPE and reaction to it[3]

Chomsky and Halle's work presented a theory of phonological structure in the form of a notation for the expression of phonological regularities, together with an architecture (based on sequential application of rules) within which the relation between underlying phonological form and surface phonetics could be computed. Following the classic discussion in Chomsky (1964), the theory aspired to success at three levels.

Minimally, the level of 'observational adequacy' could be reached to the extent descriptions formulated within the theory could accurately reproduce the empirical data on which they were based. Secondly, these descriptions could be characterized as 'descriptively adequate' to the extent they could be shown to reflect accurately not only the data from which they were constructed, but the underlying knowledge on the part of speakers that gave rise to those data. This could best be shown by demonstrating that the grammar correctly predicts novel data not present in its original basis, but which are confirmed by the intuitions of speakers.

There was little controversy over the claim that descriptively (and *a fortiori*, observationally) adequate grammars could be constructed within the SPE framework. Much less clear, however, was the claim that the theory itself could be shown to be 'explanatorily' adequate, in the sense of providing a mechanical procedure by which to determine which of multiple candidate descriptions of a given set of facts, each observationally adequate, was is in fact descriptively adequate.

Explanatory adequacy was to be achieved by incorporating into the theory an evaluation measure: an algorithm for computing, for each candidate grammar, its overall complexity, with the claim that the least complex observationally adequate grammar was in fact the descriptively adequate one. This evaluation was to

3. For a fuller presentation of the points discussed in this section, see Anderson (1980) and Anderson (1985: 328– 347).

consist in a counting of the feature specifications in each candidate grammar, with fewer features corresponding to less complexity. In order to make the description's complexity correspond to its likelihood of being 'correct' in the relevant sense, a variety of notational conventions were proposed, each supposedly corresponding to a type of linguistically significant generalization to be favored, and allowing the description to be shortened to reflect this. These notational conventions, far from being purely matters of aesthetics, were in fact a crucial, central component of the theory in its goal of attaining explanatory adequacy, each constituting an empirical claim about what sorts of generalization were in fact 'linguistically significant'.

The resulting program intended to allow all problems associated with the discovery of a descriptively adequate account of sound structure in a given language to be reduced to the mechanical manipulation of expressions in a fully explicit notational system. It was not claimed that SPE completely accomplished this goal, but that was nonetheless the program of the theory. As a result, the research issues that it presented for study centered on matters of the correct choice of notational conventions and the elaboration of the architecture within which the rules were to be applied.

The goals of *SPE* were in fact strikingly similar to those of another fundamental work of 20th century thought, Whitehead & Russell's (1910–1913) *Principia Mathematica*. That work also took as its project the reduction of all of the intellectual content of a field (mathematics) to the formal manipulation of expressions in a logistic system by means of fully explicit mechanical operations. Such an attempt to express all of the content of a field in terms subject to purely formal manipulation by explicit rules came to be known as 'logicism' in the philosophy of mathematics, a term that can also be appropriately applied to the program of SPE.

As is well known, the initial successes of Whitehead & Russell's work soon gave rise to discontent, as it became clear that there were fundamental challenges to it. These included for instance the fact that it apparently gave rise to a number of unresolvable paradoxes. Russell attempted to resolve these through the addition of a theory of 'types', but this in turn led to counterintuitive and overly restrictive emendations of the system, and the logicist program for mathematics gradually came to be abandoned.

As summarily described by Kleene (1952: Chapter 3), the problems posed by logicism gave rise to alternative views. Simplifying grossly, David Hilbert and others pursued on the one hand what is called 'formalism', attempting to pursue as much mathematical content as possible by logical manipulation without claiming that all of mathematics has this character. On the other hand, L. E. J. Brouwer attempted a resolution under the name of 'intuitionism' by rejecting all expressions purporting to refer to objects that cannot in fact be fully constructed (such as infinite sets). This has the consequence that the paradoxes arising in Whitehead & Russell's system are avoided, since the problematic classes cannot be constructed within the limits of an intuitionist logic. It also has the consequence, however, that a great deal of mathematics must fall outside the scope of the program. In the end, both formalism and intuitionism have generally been seen as inadequate bases for an understanding of mathematics.

Interestingly, just as the logicist program for mathematics gave rise to apparent problems and reactions, the same was true for Chomsky & Halle's program as expressed in *SPE*. One of these was already recognized in the concluding chapter of that work: the fact that in principle, the theory excludes all matters of concrete phonetic content in favor of purely formal representation. This they attempted to remedy by the addition of a theory of 'markedness' intended to play a role in reducing the substantive content of phonological representations and rules to matters of pure form. Just as the theory of types failed to find acceptance as an adequate resolution of the problems that arise within Whitehead & Russell's system, however, the theory of markedness advanced in *SPE* failed to attract significant adherents.

A second problematic aspect of the SPE theory was noted by Paul Kiparsky in papers circulated in the 1960s, later published in fuller form as Kiparsky (1973). The issue here was the fact that descriptions in classical generative phonology can be constructed that are unrealistically abstract, in that they posit representations and sequences of rules that are probably inaccessible to native speakers. Again, a variety of restrictive additions to the theory of SPE were proposed to exclude such analyses, but no fully satisfactory resolution was achieved.

One attempt to avoid the problem of excessive abstractness was made by Theo Vennemann and his students at UCLA in the early 1970s, and most fully presented in the work of Joan [Bybee] Hooper (1976) under the label of Natural Generative Phonology. This theory took as its basic premise the requirement that phonological statements should be confined to ones that are literally true of surface phonetic forms. Such a principle is strikingly reminiscent of the intuitionist requirement that mathematical statements should not be allowed to refer to objects that cannot be explicitly constructed; like intuitionism in mathematics, Natural Generative Phonology allows for an account of a limited subset of its intended domain (here phonology, as opposed to mathematics), but in both cases the limitations are sufficiently severe that neither achieved broad acceptance.

Another effort to address the issue of phonetic content in phonological expressions can be found in the work of David Stampe (1973) under the label Natural Phonology. This view completely rejects the purely formal nature of phonological regularities as expressed by rules in *SPE*. It posits a rich inventory of substantive 'natural processes', all presumed to be innate and active at the outset of language acquisition. On this view, acquiring the phonology of a particular language consists in learning to suppress the activity of some processes and to subordinate others, while also adding a limited set of arbitrary and language particular rules. The resulting system has some resonance with important aspects of Optimality Theory (§ 3.3 below), but in its basic form it attracted only limited attention, due in part to the fact that a much more important role must apparently be ascribed to the rules that fall outside the set of natural processes than can easily be accommodated in the system.

Within a few years after the publication of *SPE*, then, there was enough discontent with its program to encourage students to look elsewhere. On the one hand, the program itself offered few obvious topics for significant advances: the problems raised by matters of notational conventions and rule ordering were highly technical, and difficult to address in terms of readily accessible empirical evidence – difficulties similar to those presented by phonemic theory a decade or so earlier. On the other hand, the attempts to deal with limitations of the SPE program within its general 'Weltanschauung' all seemed unappealing for reasons of their own. The climate was ripe for a more serious re-orientation of theoretical attention.

3.2 A focus on representations

This found its expression in the re-orientation of phonological research over the next two decades from the study of phonological rule systems to the study of representations. SPE and related theories were built on the assumption that phonological (and phonetic) representations took the form of sequences of segmental units, each composed of values for features taken from a universally available set. The result could be represented as a matrix whose rows correspond to the features and whose columns to the successive segments. A variety of challenges arose to this formally simple mode of representation, each presenting the possibility of new research problems for investigation and new solutions to old problems. In each case the novel vista provided a certain amount of low hanging fruit for exploitation by a generation of graduate students and others; once that had been gathered, the tendency was to look for other novelties, and the field was quite prepared to provide these.

One of the earliest successes of (what would become) the generative approach to phonology was the elegant analysis of English stress provided by Chomsky et al. (1956), an account which evolved into that of Chomsky & Halle (1968). The apparatus necessary to support this picture, however, presented some problems. Since features in the SPE framework were associated with individual columns (segments) in the representation, this resulted in stress values being associated only with a single vowel, rather than with an entire syllable – and indeed, syllables had no status at all in this theory. Furthermore, the feature [Stress], unlike others, was required to take a range of numeric values, rather than a binary choice between '+' and '–'; and a rather unwieldy convention of stress reduction had to be posited such that assignment of [1Stress] anywhere within a form resulted in the demotion of all other values from [*n*Stress] to [*n*+1Stress].

A resolution to these difficulties was provided by Mark Liberman's (1975) MIT dissertation. There it was proposed that instead of a single homogeneous matrix of features, a phonological representation should be regarded as a binary branching tree whose terminal elements were prosodic constituents: syllables (whose phonological role was recognized in another MIT thesis, Kahn 1976). The hierarchical representation of the relations among these, then, could be annotated as a relation of strong to weak, and this organization could be interpreted as relative stress, instead of treating stress as a feature like others.

At one shot, this theoretical move (which came to be known as the theory of Metrical Phonology) elegantly resolved the problematic aspects of stress in the purely featural account. It also, however, opened the door to other innovations based on attributing more structure to phonological representations than that of a single uniform matrix of features. Metrical phonology was an immediate success. Analyses of a variety of languages in these terms were produced, and something of a consensus about the range of possible stress systems in the languages of the world emerged in Bruce Hayes' (1980) dissertation a few years later.

A variant of metrical representations already foreshadowed in its presentation by Liberman & Prince (1977) was the treatment of rhythmic phenomena in terms of a grid, rather than hierarchical constituent structure. This was developed and extended to other properties by Prince (1983), though work in this framework has generally been quite limited.

In contrast, the recognition of syllables and higher metrical constituents such as feet and prosodic words as structurally significant units was taken up widely and quickly became part of the basic vocabulary of phonological description. The theory of these categories was organized in work such as Nespor & Vogel (1986), and the relation of hierarchical prosodic structure to similar structures in syntax was explored by Selkirk (1984).

The new perspective on stress and other prosodic systems provided by the richer notions of structure in Metrical and Prosodic Phonology encouraged students to look for other areas in which comparable moves would provide better accounts of problematic phenomena. Such a domain was the analysis of tonal systems: like stress, tonal features seemed to be associated with phonological content in ways that were not satisfactorily represented by features of individual segments. To accommodate this, John Goldsmith (1979) proposed a view known as Autosegmental Phonology, on which rather than all being present in unitary columns of the representation, individual features were linked to one another by association lines, such that a single specification of one feature could be linked to one or more specification of others (thus allowing for a tonal value to take a number of segments as its unitary domain) or multiple values of a single feature could be linked to a single specification of some other (thus allowing for the description of contour tones as sequences of levels associated with a single vowel).

While initially motivated by phenomena of tone, the apparatus of Autosegmental Phonology was soon pressed into service for a variety of segmental processes. Assimilation became reassociation, for example, as the classical picture of a matrix of rows and columns was replaced by a generalized notion of association among features, such that traditional segmental structure is simply the limiting case where all associations are one-to-one. As work of this sort developed, the question was raised of which features tend to associate together, and which independently. This in turn led to the proposal that the features themselves are organized into a sort of tree structure, such that for instance a node [Place] dominates a number of features specifying place of articulation, and can associate as a unit (in, e.g., nasal assimilation from a following obstruent). The proposal of such organization produced the program of Feature Geometry as introduced by Clements (1985), Sagey (1990) and others. The research agenda set by this view was to uncover a single uniform organization of features into higher-level categories valid across languages. Despite considerable effort, however (e.g., McCarthy 1988), such an organization did not emerge, and the focus on work in this program declined rapidly.

Each of the theoretical innovations just surveyed resulted in an initial burst of enthusiasm and proliferation of research results. As the work became more standardized, however, phonologists sought out new topics, sometimes leaving problems of the previous round of innovation unresolved – just as the move to richer theories of representation had left behind unresolved problems of the SPE theory, such as the relation of notational conventions to special principles of application, the nature and generality of rule ordering, and others. In some instances, innovations left their traces in the general view, as with the acceptance of much of the apparatus of Metrical, Prosodic and Autosegmental theory in forming a broadly accepted view of phonological structure. By the early 1990s, however, the field – which had become accustomed to a rapid turnover of ideas and research topics – was in need of a new infusion of both, and had begun to develop a sense of stagnation. What happened then represented a more extreme change than anything since the appearance of SPE.

3.3 The rise of optimality theory[4]

During the 1980s, a great deal of artificial intelligence research within the computer science community was focused on the development of the architecture of neural networks (Rumelhart et al. 1986), 'Connectionist' systems that were claimed to be able to learn, on the basis of exemplars as training data, complex associations between inputs and outputs without explicit instruction in the nature of the relation involved. One application of this work was in the analysis of natural language, and an influential paper in that framework (Rumelhart & McClelland 1986) claimed to document such a system that acquired a significant segment of the morphology of English (past tense forms of verbs) without direct instruction apart from a training set. The assumptions and adequacy of this model were strongly criticized by Pinker & Prince (1988), and linguists and cognitive scientists generally were not impressed with the promise this approach might hold for their fields.

Among those in the computer science community engaged in the exploration of neural network models, Paul Smolensky had especially broad interests in cognition more generally, and in particular in the nature of the representations that could be attributed to these models and ways in which symbolic processes could be modeled. When he as a prominent Connectionist, and Alan Prince as a prominent critic of that approach, met and began to work together, there was actually a substantial amount of common ground for them to explore.

After several years of collaboration, with occasional presentations to other phonologists (e.g., in a workshop at the 1991 Linguistic Institute at UC Santa Cruz), Prince & Smolensky (1993; later published as Prince & Smolensky 2004) appeared as a photocopied manuscript that was widely disseminated to large numbers of phonologists. This set out a radically new approach to the description of phonological systems, dispensing entirely with language-particular rules functioning in a serial derivation.

The new framework was based centrally on generalizations about surface phonetic forms, represented as constraints taken from a universal set. These were of two sorts: 'faithfulness' constraints, which require that the output phonetic form resemble the phonological input in various ways, and 'markedness' constraints, requiring the output form to conform to various universal conditions of phonetic naturalness. These are typically in conflict with one another, and thus must be ranked with respect to their importance: a grammar then consists precisely of a ranking of the members of the universal set of constraints. A given constraint is

4. For a more detailed account than can be provided here of the rise of Optimality Theory and of the theoretical proposals that preceded and anticipated it, see Burzio (1995), Griffiths (2019) and van Oostendorp (forthcoming).

allowed to be violated in the surface form precisely to the extent this is required by other higher-ranking constraints. A derivation consists in taking a phonological input form, allowing a component of the system ("GEN") to generate a potentially unlimited range of possible corresponding outputs, and then evaluating in parallel each of these against the ranked constraint set ("EVAL"). The candidate form whose constraint violations are least serious (the 'optimal' candidate) is selected as the output.

It is not necessary for our purposes here to go into more of the details of how such an Optimality Theoretic (OT) grammar works. What is most important is the fact that it provides a radically different account of phonological organization from that of the SPE model, even supplemented with all of the representational innovations discussed above. The absence of language particular rules, or any sort of serial derivational structure, combined with the focus on surface-oriented constraints, made this quite unlike anything that had gone before.

It is true that the potential importance of regularities over surface phonetic forms had been brought into discussion previously. Kisseberth (1970) had noted in the early days of work within the SPE model that such regularities sometimes did not find any expression in such a grammar. Multiple rules might for instance 'conspire' to have the effect that stress in a given language never falls on a weak syllable (sometimes moving the stress, sometimes lengthening a vowel, sometimes deleting a syllable, etc.), but nowhere in the grammar is that stated in a unitary way, as for instance by a constraint preventing stressed weak syllables. While such 'conspiracies' often seem to constitute quite real aspects of a language's structure, the theory provided no effective way to incorporate that observation into the description, and so it had to go unstated. OT, in contrast, provided a clear status for such generalizations.

It is also important to observe (as Prince & Smolensky do) the resemblance between OT's positing a universal inventory of constraints – especially markedness constraints – and the system of natural processes in Stampe's (1973) Natural Phonology. In both cases it is presumed that the intrinsic nature of the system implementing speech has an important role to play in language, and that the effects of this have to be ranked with respect to one another and to the need to maintain distinct signals for distinct content (a primary role of the faithfulness constraints).

The universal nature of constraints assumed in OT is not a self-evident property of the theory. In particular, as analyses of individual languages have proliferated within this framework, it has become increasingly clear that the constraints that need to be posited for individual languages can in fact be quite specific. It is not obvious that the assumption of a universal set of constraints is other than a placeholder for a system by which the constraints active in a given language could be learned – a result that would change the cognitive commitments of the theory in important ways.

OT was also not the first phonological descriptive framework to focus on constraints as a formal method. Paradis & LaCharité (1993) survey three such theories as they existed at the time OT first appeared on the scene, but none of the others caught the attention of the field in the way OT did. For whatever reasons, OT came along just at the right time (when the field was eager for something new), and was aggressively promoted by its originators. Within a remarkably short time, most new work in phonology was being produced in this framework.

Initially, on the basis of the illustrative analyses provided by Prince & Smolensky (1993), OT appeared to be primarily useful for the description of prosodic properties, including syllable structure and related effects. Soon, however, research had pushed the techniques of the framework into essentially all areas of phonological structure, and its victory was to all intents and purposes complete. Some important linguists (including, notably, both Chomsky and Halle) were unconvinced, and argued against OT, but others came on board, and most importantly, students rushed to adopt the new approach in formulating dissertation topics. By the end of the 1990s, rule-based serial derivations were only to be seen in the work of a few outliers, a state of affairs that continued well into the new millennium.

Although differing in many fundamental ways from preceding phonological theories, OT can be seen as essentially developing a basic view of the architecture of grammar very similar to that of various forms of Generative Phonology. Strongly contrasting with this conception is that of the Laboratory Phonology movement, an approach originating in the 1980s in work of a number of phoneticians. The essential goal of this program was to remake phonology, deriving all of the regularities of sound structure in natural language from phonetic phenomena observable in the laboratory with little or no appeal to higher-level cognitive principles.[5] Although much work in this tradition proceeds with strongly expressed hostility to traditional generative views, there has in fact been very little actual interaction between the two approaches. Given the tenuous connections between this view and others surveyed above, a serious consideration of the Laboratory Phonology literature and related theories such as that of Articulatory Phonology (Browman & Goldstein 1989 and subsequent publications) is beyond the scope of the present chapter.

5. For a review of early work in Laboratory Phonology and an assessment of its relation to the phonological tradition, see Dobrovolsky (1994).

4. Conclusion

In the developments reviewed above over nearly a century, we can find a considerable amount of progress in ideas that has, overall, led to a richer and more substantial view of the sound systems of natural languages. By no means all of what has been seen as progress, however, can be attributed solely to the superiority of new approaches over old ones in terms of their ideas. In some cases, we see that when the problems arising within a theoretical perspective become sufficiently complex, and the data necessary to resolve them too elusive, the result is a search for a new research agenda that would allow students to achieve results in quite another direction. Science does, undeniably, make progress, but at least some of this progress results as much from a need for novelty as from the resolution of old problems.

As I write, such a change may be taking place in phonology once again, though no landmark work as significant as that of Bloomfield (1933), Harris (1951), Chomsky & Halle (1968) or Prince & Smolensky (2004) has appeared to incarnate it. While still without doubt the most widespread and influential theoretical position among phonologists, Optimality Theory too has begun to lose its force. As was the case for earlier ascendant views, the theory has reached a point where the outstanding matters of controversy are somewhat obscure and hard to resolve. In addition, the important phenomenon of opacity – generalizations that are importantly true, but true in a way not susceptible of formulation in terms of surface form – has not received a satisfactory resolution, despite serious effort and attempts to incorporate aspects of previous theories such as serial derivation (McCarthy 2007).

Partially in reaction to such accumulating problems, phonologists (along with some other linguists) have turned away from traditional methodologies to seek solutions in computational analyses, statistical inferences and the data-mining study of increasingly available large corpora of language materials. It is far too early to say whether this approach will succeed in replacing linguists' conceptions of phonological structure with something quite different. While it may be possible that long unresolved questions of phonological organization will yield to these methods, the temptation to see in this turn yet another search for low-hanging fruit is hard (for the present author) to resist.

Acknowledgements

This chapter has its origins in a talk I intended to present to the LSA at its Annual Meeting in 2014 commemorating 90 years of the Society. I prepared the talk, but at the last minute was prevented by transportation problems from attending the meeting. I passed on my slides and notes to Mark, and he gave the talk in my stead. As a result, this chapter is as much his as mine, and I am pleased to dedicate it to him. Thanks also to Martin Haspelmath for comments on a previous version.

References

Anderson, Stephen R. 1980. Notes on the development of phonological theory. *Language and Speech* 23. 115–123. https://doi.org/10.1177/002383098002300111
Anderson, Stephen R. 1985. *Phonology in the twentieth century*. Chicago: University of Chicago Press.
Anderson, Stephen R. 2000. Reflections on 'On the phonetic rules of Russian'. *Folia Linguistica* 34. 1–17. https://doi.org/10.1515/flin.2000.34.1-2.11
Anderson, Stephen R. 2001. Why linguistics needs the cognitive scientist. *Psychiatry* 64(1). 11–13. https://doi.org/10.1521/psyc.64.1.11.18237
Baudouin de Courtenay, Jan. 1895 [1972]. An attempt at a theory of phonetic alternations. In Edward Stankiewicz (ed.), *A Baudouin de Courtenay anthology*, 144–212. Bloomington, IN: Indiana University Press.
Blevins, Juliette. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486357
Bloch, Bernard. 1941. Phonemic overlapping. *American Speech* 16. 278–284. https://doi.org/10.2307/486567
Bloomfield, Leonard. 1933. *Language*. New York: Holt.
Browman, Catherine P. & Louis Goldstein. 1989. Articulatory gestures as phonological units. *Phonology* 6. 201–251. https://doi.org/10.1017/S0952675700001019
Burzio, Luigi. 1995. The Rise of Optimality Theory. *GLOT International* 1(6). 3–7.
Chomsky, Noam. 1959. Review of B. F. Skinner, *Verbal behavior. Language* 35. 547–578.
Chomsky, Noam. 1964. *Current issues in linguistic theory*. The Hague: Mouton & Co.
Chomsky, Noam & Morris Halle. 1965. Some controversial questions in phonological theory. *Journal of Linguistics* 1. 97–138. https://doi.org/10.1017/S0022226700001134
Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.
Chomsky, Noam, Morris Halle & Fred Lukoff. 1956. On accent and juncture in English. In Morris Halle, Horace Lunt & Hugh MacLean (eds.), *For Roman Jakobson*, 65–80. The Hague: Mouton & Co.
Clements, George N. 1985. The geometry of phonological features. *Phonology Yearbook* 2. 223–252. https://doi.org/10.1017/S0952675700000440
Clements, George N., & Rachid Ridouane (eds.). 2011. *Where do features come from? Cognitive, physical and developmental bases of distinctive speech categories*. Amsterdam: John Benjamins. https://doi.org/10.1075/lfab.6
Dobrovolsky, Michael. 1994. Review article: Measuring the Interface. *Canadian Journal of Linguistics* 39. 113–140. https://doi.org/10.1017/S0008413100015012
Goldsmith, John. 1979. *Autosegmental phonology*. New York: Garland Press.
Goldsmith, John & Bernard Laks. 2019. *Battle in the mind fields*. Chicago: University of Chicago Press.
Griffiths, Joshua M. 2019. On the rapid expansion of optimality theory at the end of the twentieth century. *Historiographia Linguistica* 46. 133–162. https://doi.org/10.1075/hl.00041.gri
Halle, Morris. 1957. On the phonetic rules of Russian. Unpublished paper read to the Annual Meeting of the Linguistic Society of America, Chicago.
Halle, Morris. 1959. *The sound pattern of Russian*. The Hague: Mouton.
Halle, Morris. 1962. Phonology in generative grammar. *Word* 18. 54–72. https://doi.org/10.1080/00437956.1962.11659765

Harris, Zellig. 1951. *Methods in structural linguistics*. Chicago: University of Chicago Press.
Hayes, Bruce. 1980. A metrical theory of stress rules. Cambridge, MA: MIT dissertation. [Circulated by the Indiana University Linguistics club; later published by Garland Press.].
Hooper, Joan. 1976. *An introduction to natural generative phonology*. New York: Academic Press.
Householder, Fred W. 1965. On some recent claims in phonological theory. *Journal of Linguistics* 1. 13–34. https://doi.org/10.1017/S0022226700000992
Jespersen, Otto. 1924. *The philosophy of grammar*. New York: Henry Holt and Company.
Kahn, Daniel. 1976. *Syllable based generalizations in English phonology*. Bloomington, IN: Indiana University Linguistics Club.
Kiparsky, Paul. 1973. Phonological representations. In Osamu Fujimura (ed.), *Three dimensions of linguistic theory*, 1–136. Tokyo: TEC Company.
Kisseberth, Charles. 1970. On the functional unity of phonological rules. *Linguistic Inquiry* 1. 291–306.
Kleene, Stephen C. 1952. *Introduction to meta-mathematics*. New York: Van Nostrand.
Liberman, Mark Y. 1975. The intonational system of English. Cambridge, MA: MIT dissertation. Published 1979 by Garland Press.
Liberman, Mark Y. & Alan S. Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8. 249–336.
McCarthy, John J. 1988. Feature geometry and dependency: A review. *Phonetica* 45. 84–108. https://doi.org/10.1159/000261820
McCarthy, John J. 2007. *Hidden generalizations: Phonological opacity in optimality theory*. London: Equinox Press.
Mielke, Jeff. 2008. *The emergence of distinctive features*. Oxford: Oxford University Press.
Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris.
van Oostendorp, Marc. to appear. History of phonology: Optimality Theory. In Elan Dresher & Harry van der Hulst (eds.), *Oxford handbook of the history of phonology*. Oxford: Oxford University Press.
Paradis, Carole & Darlene LaCharité. 1993. Introduction: The emergence of constraints in generative phonology and a comparison of three constraint-based models. *Canadian Journal of Linguistics* 38. 127–153. https://doi.org/10.1017/S0008413100014717
Pike, Kenneth L. 1947a. Grammatical prerequisites to phonemic analysis. *Word* 3. 155–172. https://doi.org/10.1080/00437956.1947.11659314
Pike, Kenneth L. 1947b. *Phonemics: A technique for reducing languages to writing*. Ann Arbor, MI: University of Michigan Press.
Pinker, Steven & Alan Prince. 1988. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28. 73–193. https://doi.org/10.1016/0010-0277(88)90032-7
Planck, Max. 1949. *Scientific autobiography, and other papers*. New York: Philosophical Library. Translated from German by Frank Gaynor.
Postal, Paul. 1964. Boas and the development of phonology: Some comments based on Iroquoian. *International Journal of American Linguistics* 30. 269–280. https://doi.org/10.1086/464784
Prince, Alan. 1983. Relating to the grid. *Linguistic Inquiry* 14. 19–100.
Prince, Alan & Paul Smolensky. 1993. Optimality theory: Constraint interaction in generative grammar. Manuscript, Rutgers University and University of Colorado; published as Prince & Smolensky (2004).
Prince, Alan & Paul Smolensky. 2004. *Optimality theory: Constraint interaction in generative grammar*. Oxford: Blackwell. https://doi.org/10.1002/9780470759400

Rumelhart, David E. & James L. McClelland. 1986. On learning the past tenses of English verbs. In Rumelhart et al. (1986), 216–271.

Rumelhart, David E., James L. McClelland & The PDP Research Group (eds.). 1986. *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/5236.001.0001

Sagey, Elizabeth. 1990. *The representation of features in nonlinear phonology: The articulator node*. New York: Garland.

Sapir, Edward. 1921. *Language*. New York: Harcourt, Brace & World.

Sapir, Edward. 1925. Sound patterns in language. *Language* 1. 37–51. https://doi.org/10.2307/409004

de Saussure, Ferdinand. 1916. *Cours de linguistiique générale*. Lausanne: Librairie Payot. Edited by Charles Bally & Albert Sechehaye, with the collaboration of Albert Riedlinger.

Searle, John R. 1972. Chomsky's revolution in linguistics. *New York Review of Books* 18(12). 16–24.

Selkirk, Elizabeth. 1984. *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

Skinner, B. F. 1957. *Verbal behavior*. New York: Appleton-Century-Crofts. https://doi.org/10.1037/11256-000

Stampe, David. 1973. A dissertation on natural phonology. Chicago: University of Chicago dissertation.

Watson, John B. 1924. *Behaviorism*. New York: The People's Institute Publishing Company.

Weiss, Albert Paul. 1925. Linguistics and psychology. *Language* 1. 52–57. https://doi.org/10.2307/409005

Whitehead, Alfred North & Bertrand Russell. 1910–1913. *Principia mathematica*, Volumes 1–3. Cambridge: Cambridge University Press.

CHAPTER 19

Realization Optimality Theory

A constraint-based theory of morphology

Zheng Xu
Shanghai Jiao Tong University

This chapter discusses the framework of Realization Optimality Theory, a constraint-based theory of morphology. I provide an overview of its theoretical assumptions and components such as realization constraints, markedness constraints, morphotactic constraints, inputs, outputs, and the function Gen. Previous accounts of several morphological phenomena based on Realization Optimality Theory are introduced. I also present some new thoughts on this framework with regard to affix ordering, stem formation and stem selection in inflectional morphology as well as Realization Optimality Theory approaches to derivational morphology, compounding, and clitics.

Keywords: Optimality Theory, blocking, extended exponence, affix ordering, syncretism, inflectional morphology, interaction of morphology and phonology, derivational morphology, compounding, clitics

1. Introduction

Realization Optimality Theory (Realization OT hereafter) is an inferential-realizational constraint-based framework of morphology (Xu 2007, 2011, 2016, 2019; Aronoff & Xu 2010; Xu & Aronoff 2011a, b), originally designed to handle issues on inflectional morphology. Its major components include inputs, outputs, the function Gen, realization constraints, markedness constraints, and morphotactic constraints.

In a Realization OT approach to inflectional morphology, an input usually consists of a lexical stem plus morphosyntactic features. It is assumed that these features remain unchanged in their output (cf. Grimshaw 1997). An output usually consists of a lexical stem plus morphosyntactic features that are spelled out by phonological forms. The function Gen generates an infinite list of phonological forms that realize the morphosyntactic features. Inflectional exponents are usually

https://doi.org/10.1075/cilt.353.19xu

encoded in realization constraints, which associate morphosyntactic features with phonological forms.

Markedness constraints, which usually encode typological generalizations, play a crucial role in OT and likewise in Realization OT. For example, *Feature Split is a markedness constraint forbidding a morphosyntactic feature to be realized by more than one exponent. Realization OT adopts the idea of Natural Morphology (Wurzel 1989) that in an ideal or unmarked situation one morphosyntactic feature corresponds to one exponent, which characterizes agglutinative languages. But many languages are not agglutinative, which suggests the violability of this unmarked situation. To see how this works in Realization OT, in (1), assume that an input consists of the lexeme ox, its phonological form *ox*, its inflectional class feature <ox>, and {pl}. The realization constraint {<ox>, pl}: -*en* states that both the inflectional class feature <ox> and {pl} should be realized by -*en*. It outranks the realization constraint {pl}: -*s* (i.e., the plural should be realized with -*s*) because of the specificity condition, which is indicated by a solid line. It is equally ranked with *Feature Split, which is indicated by a dotted line. Candidate b is ruled out because both the inflectional class feature <ox> and {pl} are not realized by -*en*. Candidate c is ruled out because {pl} is realized by both -*en* and -*s*, hence violating *Feature Split. Candidate a wins even if it violates the lower ranked realization constraint.

(1) oxen

Input: ox, <ox>, pl ox	*Feature Split	{<ox>, pl}: -en	{pl}: -s
☞ a. ox, <ox>, pl ox -en			*
b. ox, <ox>, pl ox -s		*!	
c. ox, <ox>, pl ox -en -s	*!		

Morphotactic constraints determine the order of exponents and include precedence constraints that require one exponent to precede another (Paster 2005; Aronoff & Xu 2010; Caballero 2010), as is the case with affix order in the West African language Fuuta Tooro (Paster 2005: 184). A morphotactic constraint could also be a pairwise order constraint that not only orders two markers but also requires them to be adjacent, as in affix order in Tagalog (Ryan 2010).

In this chapter I will review major mechanisms of Realization OT as it applies to several issues in morphology and word-formation and will introduce other mechanisms when necessary. Section 2 reviews Realization OT approaches to morphological phenomena such as blocking and extended exponence, affix ordering, and syncretism. Section 3 elaborates on stem formation and selection in Realization OT. Section 4 discusses the interaction of morphology and phonology in Realization OT. Section 5 introduces Realization OT approaches to derivational morphology, compounding, and clitics. Section 6 concludes and summarizes the chapter.

2. Several morphological phenomena

Because every morphological theory should be able to handle blocking and extended exponence, affix ordering, and syncretism, this section illustrates Realization OT approaches to these morphological phenomena.

2.1 Blocking and extended exponence

Blocking in inflectional morphology refers to a phenomenon in which a rule or affix prevents or 'bleeds' (Kiparsky 1968) the application of another rule or affix that expresses the same morphosyntactic feature value as the bleeding rule or affix. Blocking thus prevents the occurrence of multiple exponents of a single morphosyntactic feature value. On the other hand, extended exponence refers to cases in which a morphosyntactic feature value is realized by more than one exponent. Natural languages exhibit both blocking and extended exponence, so any theory of morphology must accommodate both.

Realization OT provides a unified account of blocking and extended exponence. If two constraints realizing the same morphosyntactic feature value outrank *Feature Split, we will observe extended exponence. Otherwise, we will observe blocking. (See also Harris 2017 on extended exponence and different OT approaches to it.) In Table 1, either {2} or {3, F} is realized by *t*-. The suffix -*ii* is an exponent of {2, F, SG}. The form *t-aktub-ii* shows a case of extended exponence, in which {2} is realized by both *t*- and -*ii*. The form *y-aktub-na* shows a case of blocking. The suffix -*na*, which realizes {F, PL}, blocks the prefix *t*-, which realizes {3, F}, given that both -*na* and *t*- realize {F}.

Table 1. Classical Arabic partial conjugation (Noyer 1997: 4–5)

	SBJV		
	SG	DU	PL
1	ʔ-aktub-a	n-aktub-a	n-aktub-a
2.M	t-aktub-a	t-aktub-aa	t-aktub-uu
2.F	**t-aktub-ii**	t-aktub-aa	t-aktub-na
3.M	y-aktub-a	y-aktub-aa	y-aktub-uu
3.F	t-aktub-a	t-aktub-aa	**y-aktub-na**

In (2), assume that an input consists of the stem *aktub* and {2, F, SG}. The constraint {2, F, SG}: *-ii* requires {2, F, SG} to be realized by *-ii*. The constraint {2}: *t-* requires {2} to be realized by *t-*. These two constraints are not equally ranked because of the specificity condition. Candidates b and c are ruled out by either of the two higher ranked realization constraints, respectively. Candidate a violates *FEATURE SPLIT because {2} is realized by both *t-* and *-ii*. It wins because *FEATURE SPLIT is lower ranked than the two realization constraints.

(2) Classical Arabic t-aktub-ii {2, F, SG}

aktub, 2, F, SG	{2, F, SG}: -ii	{2}: t-	*FEATURE SPLIT	{F, PL}: -na	{3, F}: t-	{3}: y-
☞ a. 2, F, SG t-aktub-ii			*			
b. 2, F, SG t-aktub	*!					
c. 2, F, SG aktub-ii		*!				

In (3), candidate b is ruled out by *FEATURE SPLIT because {F} is realized by both *t-* and *-na*. Candidate c is ruled out by {F, PL}: *-na*, which requires {F, PL} to be realized by *-na*.

(3) Classical Arabic y-aktub-na {3, F, PL}

aktub, 3, F, PL	{2, F, SG}: -ii	{2}: t-	*FEATURE SPLIT	{F, PL}: -na	{3, F}: t-	{3}: y-
☞ a. 3, F, PL y-aktub-na					*	
b. 3, F, PL t-aktub-na			*!			*
c. 3, F, PL t-aktub				*!		*

2.2 Affix ordering

This subsection presents Realization OT approaches to variable morphotactics, a phenomenon in which the position of an exponent is either flexible or unexpected. In Crysmann & Bonami's (2016) cross-linguistic study of variable morphotactics, almost all of their data can be accounted for under Realization OT, as illustrated in the data of Nepali, Murrinh-Patha, and Chintang.

Table 2 presents a partial paradigm of the Nepali verb *birsanu* 'forget'. The future tense exponent *lā* follows both the {1} exponent *aũ* and the {3.LOW} exponent *au*. By contrast, it precedes both the {2.LOW} exponent -*s* and the {3.MID} exponent -*n*. (LOW and MID refer to levels of formality.)

Table 2. Masculine singular forms of the Nepali verb *birsanu* 'forget' (Crysmann & Bonami 2016: 316)

	FUT
1	birse-aũ-lā
2.LOW	birse-lā-s
2.MID	birse-lā
3.LOW	birse-au-lā
3.MID	birse-lā-n

The position of *lā* in the words in Table 2 can be accounted for with four morphotactic constraints (4), all of which are satisfied. The illicit form *birse-lā-aũ violates constraint (4a). The illicit form *birse-s-lā violates constraint (4c). The illicit form *birse-lā-au violates constraint (4b). The illicit form *birse-n-lā violates constraint (4d).

(4) a. {1} > {FUT}: A {1} exponent should precede a {FUT} exponent.
b. {3.LOW} > {FUT}: A {3.LOW} exponent should precede a {FUT} exponent.
c. {FUT} > {2.LOW}: A {FUT} exponent should precede a {2.LOW} exponent.
d. {FUT} > {3.MID}: A {FUT} exponent should precede a {3.MID} exponent.

In Murrinh-Patha, the position of the {DU, F} exponent *ngintha* is flexible (Nordlinger 2010, 2015), preceding its verbal stem when no object marker is present (5a) but following its verbal stem when an object marker is present (5b).

(5) a. ba-**ngintha**-ngkardu-nu
SBJ3SG.SEE(13).FUT-DU.F-see-FUT
'They two (female non-siblings) will see it.'
b. ba-nhi-ngkardu-nu-**ngintha**
SBJ3SG.SEE(13).FUT-OBJ2SG-see-FUT-DU.F
'They two (female non-siblings) will see you.'

The position of *ngintha* with respect to its stem can be accounted for with the constraints in (6). Constraint (6b) outranks constraint (6a).

(6) a. {DU, F} > STEM: An exponent of {DU, F} should precede its stem.
b. {FUT} > {DU, F} / OBJ: An exponent of {FUT} should precede an exponent of {DU, F} when an {OBJ} marker is present.

In (7), candidate b is ruled out because the {DU, F} exponent *ngintha* follows its stem.

(7) Murrinh-Patha ba-**ngintha**-ngkardu-nu

	{FUT} > {DU, F} / OBJ	{DU, F} > STEM
☞ a. ba-**ngintha**-ngkardu-nu SBJ3SG.SEE(13).FUT-**DU.F**-see-FUT		
b. ba-ngkardu-nu-**ngintha** SBJ3SG.SEE(13).FUT-see-FUT-**DU.F**		*!

In (8), candidate b is ruled out because the {FUT} exponent -*nu* follows the {DU, F} exponent *ngintha* even if the object marker *nhi*- is present.

(8) Murrinh-Patha ba-nhi-ngkardu-nu-**ngintha**

	{FUT} > {DU, F} / OBJ	{DU, F} > STEM
☞ a. ba-nhi-ngkardu-nu-**ngintha** SBJ3SG.SEE(13).FUT-OBJ2SG-see-FUT-**DU.F**		*
b. ba-nhi-**ngintha**-ngkardu-nu SBJ3SG.SEE(13).FUT-OBJ2SG-**DU.F**-see-FUT	*!	

In Chintang, a language of Nepal, the order of the {SBJ, 3, NONSG} exponent *u*-, the {OBJ, 1, NONSG} exponent *kha*-, and the {NEG} marker *ma*- is not fixed and varies freely (Bickel et al. 2007), as shown in Table 3. Realization OT accounts for this by having no morphotactic constraint on the ordering of these elements.

Table 3. Chintang verb prefixes (Bickel et al. 2007: 44)

u SBJ.3.NONSG	kha OBJ.1.NONSG	ma NEG	cop see	yokt NEG	e PST	'They didn't see us.'
u	kha	ma	cop	yokt	e	'They didn't see us.'
u	ma	kha	cop	yokt	e	'They didn't see us.'
kha	u	ma	cop	yokt	e	'They didn't see us.'
kha	ma	u	cop	yokt	e	'They didn't see us.'
ma	u	kha	cop	yokt	e	'They didn't see us.'
ma	kha	u	cop	yokt	e	'They didn't see us.'

2.3 Syncretism

Stump (2001: 212) remarks that: "[i]n instances of syncretism, two or more cells within a lexeme's paradigm are occupied by the same form." In Realization OT, syncretism is mainly accounted for by either an underspecification of morphosyntactic features or output-to-output correspondence constraints.

In Hupa, an Athabaskan language (Golla 1970, cited in Embick & Halle 2005: 40), both the markers of {1, PL, OBJ} and {2, PL, OBJ} share the form *noh*-, as shown in Table 4. This syncretism can be explained by an underspecification of the person feature. It can be formally described with the constraint {PL, OBJ}: *noh*-.

Table 4. Hupa subject / object markers

	SBJ	OBJ
1.SG	W-	Wɨ-
2.SG	n-	nɨ-
1.PL	dɨ-	**noh**-
2.PL	oh-	**noh**-

In German, the marker of {2, PL} is -*t*. In indicative present, the form of {3, SG} refers to that of {2, PL}, as shown in Table 5 (adapted from Duden 2006: 441–442).

Table 5. German weak verb lachen 'laugh'

	IND.PRS	IND.PRET
1.SG	lache	lachte
2.SG	lachst	lachtest
3.SG	**lacht**	lachte
1.PL	lachen	lachten
2.PL	**lacht**	lachtet
3.PL	lachen	lachten

This directional syncretism can be accounted for with an output-to-output correspondence constraint (9).

(9) MAX ({2, PL}, {3, SG}) / {IND, PRS}: In indicative present, every segment of an exponent of {2, PL} should have a correspondent in an exponent of {3, SG}.

In (10), assume that an input consists of both the stem *lach* and the morphosyntactic feature value set {3, SG, IND, PRS}. Candidate b is ruled out because -*t*, an exponent of {2, PL}, does not occur to realize {3, SG}.

(10) German lacht

Input: lach, 3, SG, IND, PRS Output: 2, PL -t	MAX ({2, PL}, {3, SG}) / {IND, PRS}	{2, PL}: -*t*
☞ a. 3, SG, IND, PRS lach -t		
b. lach, 3, SG, IND, PRS	*!	

Stump (2001: 216) remarks that "there are instances of syncretism in which there is no discernible directionality but the property sets of the syncretized forms do not constitute any kind of natural class." In Hua, a language of New Guinea, 2.SG forms and 1.PL forms of inflected verbs always share the same termination. In Table 6, 2.SG and 1.PL forms in non-future-tense interrogative carry the terminal suffix *-pe*. This non-directional syncretism can be accounted for with constraint (11).

Table 6. Non-future-tense interrogative forms of three Hua verbs (Haiman 1980: 47f, cited in Stump 2001: 216)

	Type I	**Type II**	**Type III**
	HU **'do'**	DO **'eat'**	MI **'give'**
1.SG	hu-ve	do-ve	mu-ve
2.SG	ha-pe	da-pe	mi-pe
3.SG	hi-ve	de-ve	mi-ve
1.DU	hu-'-ve	do-'-ve	mu-'-ve
2/3.DU	ha-'-ve	da-'-ve	mi-'-ve
1.PL	hu-pe	do-pe	mu-pe
2/3.PL	ha-ve	da-ve	mi-ve

(11) {INT, 2, SG} = {INT, 1, PL}: *-pe*: Both {INT, 2, SG} and {INT, 1, PL} should be realized by *-pe*.

3. Stem formation and stem selection in inflectional morphology

Because stem formation and stem selection have not been discussed in detail in previous literature on Realization OT, this section presents Realization OT approaches to stem formation and stem selection using data of Sanskrit and Breton, based on some of the insights from Stump (2001). Stem formation refers to a process in which a stem realizes a lexeme. Stem selection refers to a process in which a stem is chosen in the context of morphosyntactic features. Stem formation and stem selection in general precede processes of realizing morphosyntactic features by affixation.

3.1 Stem formation

In Table 7, the masculine forms of the Sanskrit possessive adjective BHAGAVANT 'fortunate' have two stems, a Strong stem and a Weak stem. "Strong" and "Weak" are labels or features of morphomic categories (Aronoff 1994). The Strong stem is *bhágavant-* and the Weak stem is *bhágavat-*.

Table 7. Masculine forms of the Sanskrit possessive adjective BHAGAVANT 'fortunate' (Strong stem bhágavant-; Weak stem bhágavat-) (Stump 2001: 170)

	SG	DU	PL	
NOM	bhágavān	bhágavant-āu	bhágavant-as	based on the Strong stem
ACC	bhágavant-am	bhágavant-āu	bhágavat-as	based on the Weak stem
INS	bhágavat-ā	bhágavad-bhyām	bhágavad-bhis	
DAT	bhágavat-e	bhágavad-bhyām	bhágavad-bhyas	
ABL	bhágavat-as	bhágavad-bhyām	bhágavad-bhyas	
GEN	bhágavat-as	bhágavat-os	bhágavat-ām	
LOC	bhágavat-i	bhágavat-os	bhágavat-su	

The formation of both the Strong and Weak stems of BHAGAVANT 'fortunate' can be formally described with the constraints in (12).

(12) a. BHAGAVANT, <STRONG>: *bhágavant-*
b. BHAGAVANT, <WEAK>: *bhágavat-*

In (13), assume that an input consists of the lexeme BHAGAVANT and its lexically stipulated morphomic category feature <STRONG>. Candidate b is ruled out because it violates the constraint that requires both BHAGAVANT and <STRONG> to be realized by *bhágavant-*.

(13) Sanskrit bhágavant-

BHAGAVANT, <STRONG>	BHAGAVANT, <STRONG>: bhágavant-	BHAGAVANT, <WEAK>: bhágavat-
☞ a. BHAGAVANT, <STRONG> bhágavant		
b. BHAGAVANT, <STRONG> bhágavat-	*!	

Let us look at another case. Table 8 presents a partial paradigm of the Breton verb lexeme SKRIVAÑ 'write'. Stump (2001: 175) notes five stems for this verb, listed in (14):

(14) a. *e*-stem *skrive-* (in: all imperfect forms; third person imperative; second person singular and impersonal indicative)
b. *o*-stem *skrivo* (in: future: third singular, second plural, impersonal; present indicative: first and third plural)
c. *i*-stem *skrivi-* (in: other future forms; second plural present indicative)
d. *a*-stem *skriva-* (in: first singular present indicative)
e. radical stem *skriv-* (basis for other stems; also in: third singular present indicative, second singular imperative)

Table 8. Partial inflectional paradigm of Breton SKRIVAÑ 'write' (Stump 2001: 176)

	PRS IND	IMPF	FUT	IMP
1.SG	skriv-a-n	skriv-e-n	skriv-i-n	–
2.SG	skriv-e-z	skriv-e-s	skriv-i	skriv
3.SG	skriv	skriv-e	skriv-o	skriv-e-t
1.PL	skriv-o-m	skriv-e-m	skriv-i-m	skriv-o-m (=PRS IND)
2.PL	skriv-i-t	skriv-e-c'h	skriv-o-t	skriv-i-t (=PRS IND)
3.PL	skriv-o-nt	skriv-e-nt	skriv-i-nt	skriv-e-nt
IMPERSONAL	skriv-e-r	skriv-e-d	skriv-o-r	–

The realization of the lexeme SKRIVAÑ 'write' by one of the five verbal stems can be formally described with the constraints in (15). <A-STEM>, <E-STEM>, <I-STEM>, <O-STEM>, and <RADICAL STEM> are features of morphomic categories.

(15) a. VERB, <A-STEM>: verb root-*a*: The *a*-stem of a verb is its root plus -*a*.
b. VERB, <E-STEM>: verb root-*e*: The *e*-stem of a verb is its root plus -*e*.
c. VERB, <I-STEM>: verb root-*i*: The *i*-stem of a verb is its root plus -*i*.
d. VERB, <O-STEM>: verb root-*o*: The *o*-stem of a verb is its root plus -*o*.
e. VERB, <RADICAL STEM>: verb root: The radical stem of a verb is its root.

In (16), assume that an input consists of the lexeme SKRIVAÑ 'write', its root *skriv*, and its lexically stipulated morphomic category feature <A-STEM>. Candidate a wins because it satisfies the constraint that requires the *a*-stem of a verb to be realized by its verb root followed by -*a*.

(16) Breton skriv-a

SKRIVAÑ, <A-STEM> skriv	VERB, <A-STEM>: verb root-a	VERB, <E-STEM>: verb root-e	VERB, <I-STEM>: verb root-i	VERB, <O-STEM>: verb root-o	VERB, <RADICAL STEM>: verb root
☞ a. SKRIVAÑ, <A-STEM> skriva					
b. SKRIVAÑ, <A-STEM> skrive	*!				
c. SKRIVAÑ, <A-STEM> skrivi	*!				
d. SKRIVAÑ, <A-STEM> skrivo	*!				
e. SKRIVAÑ, <A-STEM> skriv	*!				

3.2 Stem selection

Having accounted for stem formation, we need to be able to explain stem selection. In Table 7, which presented a masculine paradigm, the Sanskrit Strong stem *bhágavant-* occurs in nominative, accusative singular, and accusative dual forms. The Weak stem *bhágavat-* occurs elsewhere. The selection of either stem can be described with the two output-to-output correspondence constraints in (17).

(17) a. MAX ({BHAGAVANT, <STRONG>}: {BHAGAVANT}) / {M, NOM}/{M, ACC, SG}/{M, ACC, DU}: In the context of {M, NOM} or {M, ACC, SG} or {M, ACC, DU}, every segment of the Strong stem of BHAGAVANT should have a correspondent in the form of BHAGAVANT.

b. DEP ({BHAGAVANT, <WEAK>}: {BHAGAVANT}) / {M, CASE}: In the context of any case in masculine, every segment of the form of BHAGAVANT should have a correspondent in the Weak stem of BHAGAVANT.

In (18), assume that an input consists of the lexeme BHAGAVANT and a morphosyntactic feature set {M, NOM, DU}. A list of output stems, including both the Strong stem *bhágavant-* and the Weak stem *bhágavat-*, is derived from previous stem-formation processes. Candidate a wins even if it violates the lower ranked DEP constraint. It violates this constraint because it contains *n*, which does not occur in the Weak stem. Candidate b is ruled out by the MAX constraint because it lacks *n*, which occurs in the Strong stem.

(18) Sanskrit bhágavant-, M, NOM, DU

Input: BHAGAVANT, M, NOM, DU Output 1: BHAGAVANT, <STRONG> bhágavant- Output 2: BHAGAVANT, <WEAK> bhágavat-	MAX ({BHAGAVANT, <STRONG>}: {BHAGAVANT}) / {M, NOM}/{M, ACC, SG}/{M, ACC, DU}	DEP ({BHAGAVANT, <WEAK>}: {BHAGAVANT}) / {M, CASE}
☞ a. BHAGAVANT, M, NOM, DU bhágavant-		*
b. BHAGAVANT, M, NOM, DU bhágavat-	*!	

In (19), assume that an input consists of the lexeme BHAGAVANT and a morphosyntactic feature set {M, ACC, PL}. Candidate b vacuously satisfies the MAX constraint because it does not apply in the context of {M, ACC, PL}.

(19) Sanskrit bhágavat-, M, ACC, PL

Input: BHAGAVANT, M, ACC, PL Output 1: BHAGAVANT, <STRONG> bhágavant- Output 2: BHAGAVANT, <WEAK> bhágavat-	MAX ({BHAGAVANT, <STRONG>}: {BHAGAVANT}) / {M, NOM}/{M, ACC, SG}/{M, ACC, DU}	DEP ({BHAGAVANT, <WEAK>}: {BHAGAVANT}) / {M, CASE}
a. BHAGAVANT, M, ACC, PL bhágavant-		*!
☞ b. BHAGAVANT, M, ACC, PL bhágavat-		

The selection of each of the five stems of the Breton verb skrivañ 'write' (Table 8) can be accounted for with the output-to-output correspondence constraints in (20).

(20) a. Max ({verb, <a-stem>}: {verb} / {1, sg, prs, ind}): In the context of {1, sg, prs, ind}, every segment of the *a*-stem of a verb should occur in the verb. (abbreviated as Max a-stem)

b. Ident ({verb, <a-stem>}: {verb} / {1, sg, prs, ind}): In the context of {1, sg, prs, ind}, every segment of the *a*-stem of a verb should be identical with its corresponding segment in the verb in terms of its phonological features. (abbreviated as Ident a-stem)

c. Max ({verb, <e-stem>}: {verb} / {impf}/{3, imp}/{2, sg, prs, ind}/{impersonal, prs, ind}): In the context of {impf} or {3, imp} or {2, sg, prs, ind} or {impersonal, prs, ind}, every segment of the *e*-stem of a verb should occur in the verb. (abbreviated as Max e-stem)

d. Ident ({verb, <e-stem>}: {verb} / {impf}/{3, imp}/{2, sg, prs, ind}/{impersonal, prs, ind}): In the context of {impf} or {3, imp} or {2, sg, prs, ind} or {impersonal, prs, ind}, every segment of the *e*-stem of a verb should be identical with its corresponding segment in the verb in terms of its phonological features. (abbreviated as Ident e-stem)

e. Max ({verb, <i-stem>}: {verb} / {2, pl, prs, ind}/{1, sg, fut}/{2, sg, fut}/{1, pl, fut}/{3, pl, fut}/{2, pl, imp}): In the context of {2, pl, prs, ind} or {1, sg, fut} or {2, sg, fut} or {1, pl, fut} or {3, pl, fut} or {2, pl, imp}, every segment of the *i*-stem of a verb should occur in the verb. (abbreviated as Max i-stem)

f. Ident ({verb, <i-stem>}: {verb} / {2, pl, prs, ind}/{1, sg, fut}/{2, sg, fut}/{1, pl, fut}/{3, pl, fut}/{2, pl, imp}): In the context of {2, pl, prs, ind} or {1, sg, fut} or {2, sg, fut} or {1, pl, fut} or {3, pl, fut} or {2, pl, imp}, every segment of the *i*-stem of a verb should be identical with its corresponding segment in the verb in terms of its phonological features. (abbreviated as Ident i-stem)

g. Max ({verb, <o-stem>}: {verb} / {1, pl, prs, ind}/{3, pl, prs, ind}/{3, sg, fut}/{2, pl, fut}/{impersonal, fut}/{1, pl, imp}): In the context of {1, pl, prs, ind} or {3, pl, prs, ind} or {3, sg, fut} or {2, pl, fut} or {impersonal, fut} or {1, pl, imp}, every segment of the *o*-stem of a verb should occur in the verb. (abbreviated as Max o-stem)

h. Ident ({verb, <o-stem>}: {verb} / {1, pl, prs, ind}/{3, pl, prs, ind}/{3, sg, fut}/{2, pl, fut}/{impersonal, fut}/{1, pl, imp}): In the context of {1, pl, prs, ind} or {3, pl, prs, ind} or {3, sg, fut} or {2, pl, fut} or {impersonal, fut} or {1, pl, imp}, every segment of the *o*-stem of a verb should be identical with its corresponding segment in the verb in terms of its phonological features. (abbreviated as Ident o-stem)

i. Dep ({verb, <radical stem>}: {verb} / {3, sg, prs, ind}/{2, sg, imp}): In the context of {3, sg, prs, ind} or {2, sg, imp}, every segment of a verb should have a correspondent in its radical stem. (abbreviated as Dep radical stem)

Admittedly, the constraints in (20) are language-specific and are formulated in a seemingly complicated manner, but this is necessitated by the language-specific nature of morphological realization. Additionally, a Realization OT grammar consisting of constraints such as Max, Dep, and Ident can not only select a correct stem but also rule out unacceptable ones in a formally explicit manner.

In (21), assume that an input consists of the lexeme skrivañ 'write' and a morphosyntactic feature set {1, sg, prs, ind}. Output stems are derived from previous processes of stem-formation. Candidates (b)–(d) violate Ident a-stem because none of them ends in *a*. The *a*-stem *skriva* does not have a corresponding final vowel in candidate e, which is therefore ruled out.

(21) Breton skriva, 1, sg, prs, ind

Input: skrivañ, 1, sg, prs, ind Outputs: skriva, skrive, skrivi, skrivo, skriv	Max a-stem	Ident a-stem	Max e-stem	Ident e-stem	Max i-stem	Ident i-stem	Max o-stem	Ident o-stem	Dep radical stem
☞ a. skrivañ, 1, sg, prs, ind skriva									
b. skrivañ, 1, sg, prs, ind skrive		*!							
c. skrivañ, 1, sg, prs, ind skrivi		*!							
d. skrivañ, 1, sg, prs, ind skrivo		*!							
e. skrivañ, 1, sg, prs, ind skriv	*!								

In (22), assume that an input consists of the lexeme skrivañ 'write' and a morphosyntactic feature set {3, sg, prs, ind}. Candidates (a)–(d) are ruled out because they all end in a vowel which does not have a correspondent in the radical stem *skriv*.

(22) Breton skriv, 3, SG, PRS, IND

Input: SKRIVAÑ, 3, SG, PRS, IND Outputs: skriva, skrive, skrivi, skrivo, skriv	MAX A-STEM	IDENT A-STEM	MAX E-STEM	IDENT E-STEM	MAX I-STEM	IDENT I-STEM	MAX O-STEM	IDENT O-STEM	DEP RADICAL STEM
a. SKRIVAÑ, 3, SG, PRS, IND | skriva									*!
b. SKRIVAÑ, 3, SG, PRS, IND | skrive									*!
c. SKRIVAÑ, 3, SG, PRS, IND | skrivi									*!
d. SKRIVAÑ, 3, SG, PRS, IND | skrivo									*!
☞ e. SKRIVAÑ, 3, SG, PRS, IND | skriv									

4. The interaction of morphology and phonology

In Realization OT, it is assumed that, by default, morphological processes such as the realization of morphosyntactic features by affixation precede phonological processes such as vowel assimilation and consonant deletion, which accompany the affixation. Phonologically conditioned morphological processes refer to phenomena in which phonological information or constraints affect morphological processes (see, for example, Inkelas 2011, 2014). As this section shows, we can account for phonologically conditioned selection of competing exponents that is optimizing if both phonological constraints and realization constraints are ranked together. Competing exponents can realize either the same set of morphosyntactic feature values or not. Competing exponents can be suppletive allomorphs, which are not derived from a single underlying phonological representation.

Hungarian, for example, presents a case of the emergence of an unmarked exponent due to phonological requirements. Hungarian has both a definite conjugation, used when there is a direct object, and an indefinite conjugation used elsewhere (Rounds 2001: 23). The exponent of {2, SG, INDF, PRS, IND} is *-sz* ([s]) (e.g., *ír-sz* '(you, SG) write (something (INDF))'). When *-sz* is expected to follow a verbal stem ending in a strident, it is replaced by *-Vl* (subject to vowel harmony) (e.g., *olvas-ol* '(you, sg) read (something (indf))', **olvas-(o)sz*) (Carstairs 1988, 1990; Rounds 2001; Vago 1980). The suffix -*Vl* is a default marker for {2, sg, indf} compared to -*sz* because -*Vl* not only marks {prs, ind} but also {pst}, {cond}, and {sbjv} (Vago 1980).

A Realization OT approach accounts for the competition between -*sz* and -*Vl* for realizing {2, sg, indf, prs, ind} with the constraints in (23) (Xu 2011: 474). Constraints (23a)–(b) are realization constraints; constraint (23c) is a phonological constraint; constraint (23d) is a markedness constraint.

(23) a. {2, sg, indf, prs, ind}: -*sz*: {2, sg, indf, prs, ind} is realized by -*sz*. (-sz)
b. {2, sg, indf}: -*Vl*: {2, sg, indf} is realized by -*Vl*. (-Vl)
c. OCP (strident): No adjacent stridents are allowed.
d. *Empty Morph: A morph must associate with a feature value.

In (24), assume that an input consists of the verbal stem *olvas* 'read' and {2, sg, indf, prs, ind}.[1] Candidate b contains two adjacent stridents, hence violating OCP (strident). Candidate c contains an empty morph -*o*, hence violating *Empty Morph.

(24) Competition between Hungarian -sz and -Vl

olvas, 2, sg, indf, prs, ind	OCP (strident)	*Empty Morph	-sz	-Vl
☞ a. 2, sg, indf, prs, ind olvas -ol			*	
b. 2, sg, indf, prs, ind olvas -sz	*!			*
c. 2, sg, indf, prs, ind olvas -o -sz		*!		*

Korean presents a case of phonologically conditioned suppletive allomorphy that is optimizing. Its accusative suffix has two allomorphs, -*lul* and -*ul*. The suffix -*lul* occurs after stems ending in a vowel while the suffix -*ul* occurs after stems ending in a consonant (Lapointe 2001, Bonet & Harbour 2012). Realization OT can resolve the competition between -*lul* and -*ul* for realizing {acc} (Xu 2016). In (25), assume that an input consists of the stem *cho* and {acc}. The grammar in (25) consists of two realization constraints and two phonological constraints. The constraint Onset

1. I omit discussion of vowel harmony in Hungarian. Readers are referred to Vago (1980), Kenesei, Vago & Fenyvesi (1997) and Rounds (2001) for relevant discussion.

requires a syllable to begin with a consonant. The constraint No Coda forbids a syllable to end in a consonant.

(25) Korean cho-lul

Input: cho, ACC	{ACC}: -lul	{ACC}: -ul	Onset	No Coda
☞ a. ACC cho.lul		*		*
b. ACC cho.ul	*		*!	*
c. cho, ACC	*	*!		

Similarly, the correct form *Kim-ul* can be selected with the same grammar. In (26), assume that an input consists of the stem *Kim* and {ACC}.

(26) Korean Kim-ul

Input: Kim, ACC	{ACC}: -lul	{ACC}: -ul	Onset	No Coda
☞ a. ACC Ki.mul	*			*
b. ACC Kim.lul		*		**!
c. Kim, ACC	*	*!		*

5. Realization Optimality Theory approaches to derivational morphology, compounding, and clitics

Although it was originally designed to account for inflectional morphology, Realization OT can be applied to derivational morphology, compounding, and clitics. We can account for competitions in word formation if constructional schemas with derivational affixes or compound components are converted into realization constraints. Additionally, Realization OT can be easily reconciled with Anderson's (2005) OT accounts of the positioning of special clitics.

5.1 Competition in derivational morphology

Booij (2010) presents the framework of Construction Morphology, which focuses on both derivational morphology and compounding. A major task of Construction Morphology is to abstract constructional schemas from their specific instantiations and organize them into inheritance hierarchies. For example, a constructional schema with the variable *x* can be abstracted over its specific words, (27).[2] A more abstract constructional schema (with more open slots or variables) usually stands higher in a hierarchy.

(27) $[[x]_V\ er]_N$ 'one who Vs habitually, professionally'

buyer eater shouter walker …

Construction Morphology lacks a formal mechanism, though, to account for competition between words for expressing the same meaning. For example, *cook* means 'a person who cooks habitually, professionally'. By contrast, *cooker* cannot be used for a person at all, only for an appliance (Mark Aronoff p.c.). The word *thief* means 'a person who steals habitually'. By contrast, the word *stealer* is not used in isolation for this meaning in modern English (Mark Aronoff p.c.).

Realization OT can resolve the problem with constructional schemas being converted into realization constraints. In (28), constraints (28a) and (28b) are both markedness constraints, with (28a) resembling *Feature Split and (28b) conforming to the specificity condition. Constraint (28c) is a realization constraint.

(28) a. *Multiple Realization: A meaning cannot be multiply expressed. That is, if a meaning or its part is expressed by one form, this meaning (part) should not be expressed by another form at the same time.
b. Specificity: A lexical item which expresses a meaning should have priority to be spelled out over another lexical item which expresses part of this meaning.
c. 'one who Vs habitually, professionally': $[[x]_V\ er]_N$: The meaning 'one who Vs habitually, professionally' should be expressed by $[[x]_V\ er]_N$. (abbreviated as $[[x]_V\ er]_N$)

In (29), assume that an input contains the meaning 'a person who steals habitually'. The lexicon supplies two relevant lexical items *thief* and *steal*. Candidate b is ruled

2. Booij (2010: 4) remarks that the use of a constructional schema like the one in (27) looks similar to the use of word formation rules, as proposed in Aronoff 1976, e.g., $[x]_V \rightarrow [[x]_V\ er]_N$ Semantics: 'one who Vs habitually, professionally.'

out because the input meaning is realized by both *thief* and *stealer*. Candidate c violates Specificity because *thief* expresses 'a person who steals habitually' while *steal* expresses part of this meaning so that *thief* should have priority to be spelled out. Candidate d violates *Multiple Realization because *thief* expresses the whole input meaning so that *-er* is not needed to express the same meaning. It also violates the realization constraint because the suffix *-er* is supposed to attach to a verb instead of a noun.

(29) thief

'person who steals habitually' $[\text{thief}]_N$ 'person who steals habitually'; $[\text{steal}]_V$ 'steal'	*Multiple Realization	Specificity	$[[x]_V \text{ er}]_N$
☞ a. $[\text{thief}]_N$			*
b. $[\text{thief}]_N$-$[[\text{steal}]_V \text{ er}]_N$	*!		
c. $[[\text{steal}]_V \text{ er}]_N$		*!	
d. $[[\text{thief}]_N \text{ er}]_N$	*!		*
e. $[\text{steal}]_V$		*!	*

5.2 Competition in compounding

Realization OT can also account for cases of competition that involve compounds, with constructional schemas being converted into realization constraints. In Chinese, for example, adjective-noun combinations ([A N] hereafter), which are compounds, can be used interchangeably with [A *de* N], which are phrases (30).[3]

(30) huang (de) chenshan
yellow DE shirt
'yellow shirt'

In many contexts only one construction, either [A N] or [A *de* N], is selected. For example, [A *de* N] tends not to occur immediately after *de* (31).

(31) Lisi de huang (*de) chenshan
Lisi DE yellow DE shirt
'Lisi's yellow shirts'

3. See Xu (2018) for arguments for the word status of Chinese [A N].

[A *de* N] cannot name entities while [A N] can because the former are phrases while the latter are words (32).

(32) Zhe jiao huang (*de) chenshan
this call yellow DE shirt
'This is called "yellow shirt".'

[A *de* N] tends to be used if the adjective is complex (33).

(33) hong-tongtong *(de) lian
red-MODIFYING SUFFIX DE face
'reddish face'

[A *de* N] is used if a selectional restriction forbids the adjective to directly combine with the noun. For example, *duan* 'short' cannot directly modify nouns expressing large objects (34a). By contrast, it can directly modify nouns expressing small objects (34b–c).

(34) a. duan *(de) he
short DE river
'short river'
b. duan (de) chi
short DE ruler
'short ruler'
c. duan (de) dao
short DE knife
'short knife'

The constraints shown in (35) account for (30)–(34) (Xu 2019).

(35) a. $[A\ N]_N$: 'N with the property of being A' should be expressed by $[A\ N]_N$.
b. $[A\ \textit{de}\ N]_{NP}$: 'N with the property of being A' should be expressed by $[A\ \textit{de}\ N]_{NP}$.
c. OCP (*de*): Sequences of *de* are not allowed.
d. NoPhraseNaming: A phrase cannot name an entity.
e. $[A\ N]_N$ ($A = \mu$): The adjective in an [A N] compound should be monomorphemic.
f. $*[\text{duan}\ N]_N$ (N = large objects): *duan* 'short' cannot directly combine with nouns which express large objects described in terms of length.

In (36), assuming that the input contains meanings to be realized, both the adjective *huang* 'yellow' and the noun *chenshan* 'shirt' are supposed to occur in the output (Xu 2019). Both candidates a and b are correct outputs. Candidate c, a null output candidate, is ruled out because it violates both of the realization constraints. Hence, (30) is accounted for.

(36) Chinese huang (de) chenshan 'yellow shirt'

Input: 'shirt with the property of being yellow' 'shirt': chenshan; 'yellow': huang	[A N]$_N$	[A de N]$_{NP}$	OCP (de)	No Phrase Naming	[A N]$_N$ (A = μ)	*[duan N]$_N$ (N = large objects)
☞ a. huang chenshan yellow shirt		*				
☞ b. huang de chenshan yellow DE shirt	*					
c. Ø	*	*!				

Constraint (35c) is used to account for (31). It rules out candidate b in (37).

(37) Chinese Lisi de huang (*de) chenshan 'Lisi's yellow shirt'

Input: 'Lisi's shirt with the property of being yellow' 'shirt': chenshan; 'yellow': huang	[A N]$_N$	[A de N]$_{NP}$	OCP (de)	No Phrase Naming	[A N]$_N$ (A = μ)	*[duan N]$_N$ (N = large objects)
☞ a. Lisi de huang chenshan Lisi DE yellow shirt		*				
b. Lisi de huang de chenshan Lisi DE yellow DE shirt	*		*!			

Constraint (35d) is used to account for (32). It rules out candidate b in (38).

(38) Chinese Zhe jiao huang (*de) chenshan. 'This is called "yellow shirt".'

Input: 'the type of shirt with the property of being yellow' 'shirt': chenshan; 'yellow': huang	[A N]$_{N}$	[A de N]$_{NP}$	OCP (de)	No Phrase Naming	[A N]$_{N}$ (A = μ)	*[duan N]$_{N}$ (N = large objects)
☞ a. 'the type of shirt with the property of being yellow': huang chenshan (yellow shirt)		*				
b. 'the type of shirt with the property of being yellow': huang de chenshan (yellow DE shirt)	*			*!		

Constraint (35e) is used to account for (33). It rules out candidate a in (39).

(39) Chinese hong-tongtong *(de) lian 'reddish face'

Input: 'face with the property of being reddish' 'face': lian; 'reddish': hong-tongtong (red-MODIFYING SUFFIX)	[A N]$_{N}$	[A de N]$_{NP}$	OCP (de)	No Phrase Naming	[A N]$_{N}$ (A = μ)	*[duan N]$_{N}$ (N = large objects)
a. hong-tongtong lian red-MODIFYING SUFFIX face		*			*!	
☞ b. hong-tongtong de lian red-MODIFYING SUFFIX DE face	*					

Constraint (35f) is used to account for (34a). It rules out candidate a in (40).

(40) Chinese duan *(de) he 'short river'

Input: 'river with the property of being short' 'river': he; 'short': duan	[A N]$_{N}$	[A de N]$_{NP}$	OCP (de)	No Phrase Naming	[A N]$_{N}$ (A = μ)	*[duan N]$_{N}$ (N = large objects)
a. duan he		*				*!
short river						
☞ b. duan de he short DE river	*					

5.3 The positioning of special clitics

Realization OT can be easily reconciled with Anderson's (2005) OT accounts of the positioning of special clitics,[4] which Anderson considered phrasal affixes. In contrast to affixes inside words, special clitics attach promiscuously to their hosts (Spencer & Luís 2012). According to Anderson (2005), the positioning of special clitics is sensitive to either the first or the last element of a domain, which could be CP, IP, DP, etc. Anderson (2005) proposes constraints such as those in (41).

(41) a. NonInitial (*cl*$_i$, CP): The clitic *cl*$_i$ should not be initial in a CP that includes it.
     b. LeftMost (*cl*$_i$, IP): The clitic *cl*$_i$ should be at the left edge of an IP that includes it.

In Czech, for example, the reflexive clitic *se* occurs at the left edge of an IP that includes it. This clitic does not occur in the initial position of a CP that includes it. See (42) (Richardson 1997; Anderson 2005).

(42) Helena říkala, [$_{CP}$že [$_{IP}$=**se** Petr odstěhoval]]
     Helen said that REFL Peter moved
     'Helen said that Peter had moved.'

Anderson (2005) claims that the position of *se* in (42) should be accounted for with the constraints in (41), but he does not provide any tableaux for his analysis. The constraints in (41) can be easily incorporated into a Realization OT grammar, which also uses constraint (43) to account for the realization of the reflexive feature.

(43) {REFL}: *se*$_{cli}$: The reflexive feature should be realized by the clitic *se*.

In terms of the reflexive feature in (44), assume that {REFL} occurs in both the input and output. Candidate b is unacceptable because it occurs in the initial position of the CP. Additionally, it does not occur at the left edge of the IP. Candidate c is ruled out because {REFL} is not realized by the clitic *se*.

4. See also Kasak (2019) for discussion of reconciling Realization OT with Anderson (2005).

(44) Czech Helena říkala, že se Petr odstěhoval 'Helen said that Peter had moved'

	{REFL}: se_{cli}	NONINITIAL (cl_i, CP)	LEFTMOST (cl_i, IP)
☞ a. Helena říkala, [$_{CP}$že [$_{IP}$=**se** Petr odstěhoval]] Helen said that REFL Peter moved			
b. Helena říkala, [$_{CP}$**se**= že [$_{IP}$Petr odstěhoval]] Helen said REFL that Peter moved		*!	*
c. Helena říkala, [$_{CP}$že [$_{IP}$ Ø Petr odstěhoval]] Helen said that REFL Peter moved	*!		

6. Conclusion

Realization OT can provide a unified account of blocking and extended morphological exponence by ranking *FEATURE SPLIT with realization constraints expressing the same morphosyntactic feature value. Likewise, variable morphotactics can be well accounted for under Realization OT, as can all the types of syncretism in Stump (2001). Stem formation and stem selection in inflectional morphology can be described with stem formation constraints and stem selection constraints, which are essentially output-to-output correspondence constraints. Phonologically conditioned selection of competing exponents that is optimizing can be accounted for by ranking phonological constraints with realization constraints. By converting constructional schemas into realization constraints, we can account for cases of competition in word formation. Special clitics can be encoded in realization constraints without specifying their position. Realization OT can be easily reconciled with Anderson (2005) by ranking these realization constraints with his constraints on the positioning of special clitics.

Realization OT provides a useful and promising framework of morphology and word-formation. Realization OT combines the ability to account for language-specific complexities through realization constraints with the prioritization of universal violable constraints through Optimality Theory. This approach can thus provide a more unified and coherent account of some morphological phenomena than previous approaches were able to do. Given that previous works on this framework have only covered a limited quantity of language data, the examination of more language phenomena and data can further demonstrate the advantages, disadvantages, and limitations of this framework when it is compared with alternative approaches.

Acknowledgements

It is my great pleasure to dedicate this chapter to Mark Aronoff, who has been my mentor and advisor since I studied at Stony Brook University in 2000.

Funding

This work was supported by grant 15BYY004 to the author from the National Social Science Foundation of China.

This chapter uses the abbreviations

1, 2, 3	first, second, and third person	LOC	locative
ABL	ablative	M	masculine
ACC	accusative	NEG	negation
COND	conditional	NOM	nominative
DAT	dative	OBJ	object
DU	dual	PL	plural
F	feminine	PRET	preterite
FUT	future	PRS	present
GEN	genitive	PST	past
IMP	imperative	REFL	reflexive
IMPF	imperfect	REP	repetitive
IND	indicative	SBJ	subject
INDF	indefinite	SBJV	subjunctive
INS	instrumental	SG	singular
INT	interrogative		

References

Anderson, Stephen R. 2005. *Aspects of the theory of clitics*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199279906.001.0001

Aronoff, Mark. 1976. *Word formation in generative grammar*. Cambridge, MA: MIT Press.

Aronoff, Mark. 1994. *Morphology by itself: Stems and inflectional classes*. Cambridge, MA: MIT Press.

Aronoff, Mark & Zheng Xu. 2010. A Realization Optimality-Theoretic approach to affix order. *Morphology* 20(2). 381–411. https://doi.org/10.1007/s11525-010-9181-2

Bickel, Balthasar, Goma Banjade, Martin Gaenzle, Elena Lieven, Netra Prasad Paudya, Ichchha Purna Rai, Rai Manoj, Novel Kishore Rai & Sabine Stoll. 2007. Free prefix ordering in Chintang. *Language* 83. 43–73. https://doi.org/10.1353/lan.2007.0002

Bonet, Eulàlia & Daniel Harbour. 2012. Contextual allomorphy. In Jochen Trommer (ed.), *The morphology and phonology of exponence*, 195–235. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199573721.003.0007
Booij, Geert. 2010. *Construction Morphology*. Oxford: Oxford University Press.
Caballero, Gabriella. 2010. Scope, phonology and morphology in an agglutinating language: Choguita Rarámuri (Tarahumara) variable suffix ordering. *Morphology* 20. 165–204. https://doi.org/10.1007/s11525-010-9147-4
Carstairs, Andrew. 1988. Some implications of phonologically conditioned suppletion. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology* 1988, 67–94. Dordrecht: Foris.
Carstairs, Andrew. 1990. Phonologically conditioned suppletion. In Wolfgang. U. Dressler, Hans C. Luschützky, Oskar E. Pfeiffer & John R. Rennison (eds.), *Contemporary morphology*, 17–23. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110874082.17
Crysmann, Berthold & Olivier Bonami. 2016. Variable morphotactics in information-based morphology. *Journal of Linguistics* 52. 311–374. https://doi.org/10.1017/S0022226715000018
Duden. 2006. *Die Grammatik* (Band 4). Mannheim: Dudenverlag.
Embick, David & Morris Halle. 2005. On the status of stems in morphological theory. In Twan Geerts, Ivo van Ginnekan & Haike Jacobs (eds.), *Romance languages and linguistic theory 2003*, 37–62. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.270.03emb
Golla, Victor. 1970. Hupa grammar. Berkeley, CA: University of California dissertation.
Grimshaw, Jane. 1997. Projection, heads, and optimality. *Linguistic Inquiry* 28(3). 373–422.
Haiman, John. 1980. *Hua: A Papuan language of the Eastern Highlands of New Guinea*. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.5
Harris, Alice C. 2017. *Multiple exponence*. Oxford: Oxford University Press.
Inkelas, Sharon. 2011. The interaction between morphology and phonology. In John Goldsmith, Jason Riggle & Alan C. L. Yu (eds.), *The handbook of phonological theory*, 68–102. Oxford: Blackwell. https://doi.org/10.1002/9781444343069.ch3
Inkelas, Sharon. 2014. *The interplay of morphology and phonology*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199280476.001.0001
Kasak, Ryan. 2019. Affix ordering and templatic morphology in Mandan. New Haven, CT: Yale University dissertation.
Kenesei, István, Robert M. Vago & Anna Fenyvesi. 1997. *Hungarian*. London: Routledge.
Kiparsky, Paul. 1968. Linguistic universals and linguistic change. In Emmon Bach & Robert T. Harms (eds.), *Universals in linguistic theory*, 170–210. New York: Holt, Rinehart, and Winston.
Lapointe, Steven G. 2001. Stem selection and OT. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 1999*, 263–297. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-017-3722-7_10
Nordlinger, Rachel. 2010. Verbal morphology in Murrinh-Patha: Evidence for templates. *Morphology* 20. 321–341. https://doi.org/10.1007/s11525-010-9184-z
Nordlinger, Rachel. 2015. Inflection in Murrinh-Patha. In Matthew Baerman (ed.), *The Oxford handbook of inflection*, 491–519. Oxford: Oxford University Press.
Noyer, Rolf. 1997. *Features, positions, and affixes in autonomous morphological structure*. New York: Garland.
Paster, Mary. 2005. Pulaar verbal extensions and phonologically driven affix order. In Geert Booij and Jaap van Marle (eds.), *Yearbook of morphology* 2005, 155–199. Dordrecht: Springer.

Richardson, Matthew. 1997. Czech clitics: 2P or not 2P, that is the question. In Lizanne Kaiser (ed.), *Yale a-morphous linguistics essays*, 131–150. New Haven: Department of Linguistics, Yale University.

Rounds, Carol. 2001. *Hungarian: An essential grammar*. London: Routledge.

Ryan, Kevin. 2010. Variable affix order: Grammar and learning. *Language* 86.758–791. https://doi.org/10.1353/lan.2010.0032

Spencer, Andrew & Ana R. Luís. 2012. *Clitics: An introduction*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139033763

Stump, Gregory T. 2001. *Inflectional morphology: A theory of paradigm structure*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486333

Vago, Robert M. 1980. *The sound pattern of Hungarian*. Washington, D.C.: Georgetown University Press.

Wurzel, Wolfgang Ullrich. 1989. *Inflectional morphology and naturalness*. Dordrecht: Kluwer.

Xu, Zheng. 2007. Inflectional morphology in Optimality Theory. Stony Brook, NY: Stony Brook University dissertation.

Xu, Zheng. 2011. Optimality Theory and morphology. *Language and Linguistics Compass* 5(7). 466–484. https://doi.org/10.1111/j.1749-818X.2011.00290.x

Xu, Zheng. 2016. The role of morphology in Optimality Theory. In Andrew Hippisley & Gregory T. Stump (eds.), *The Cambridge handbook of morphology*, 550–587. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781139814720.020

Xu, Zheng. 2018. The word status of Chinese adjective-noun combinations. *Linguistics* 56(1). 207–256. https://doi.org/10.1515/ling-2017-0035

Xu, Zheng. 2019. Chinese adjective-noun combinations. In Franz Rainer, Francesco Gardani, Wolfgang U. Dressler & Hans Christian Luschützky (eds.), *Competition in inflection and word-formation* (Studies in Morphology 5), 307–334. Cham: Springer. https://doi.org/10.1007/978-3-030-02550-2_12

Xu, Zheng & Mark Aronoff. 2011a. A Realization Optimality Theory approach to blocking and extended morphological exponence. *Journal of Linguistics* 47(3). 673–707. https://doi.org/10.1017/S002222671100003X

Xu, Zheng & Mark Aronoff. 2011b. A Realization Optimality-Theoretic approach to full and partial identity of forms. In Martin Maiden, John Charles Smith, Maria Goldbach & Marc-Olivier Hinzelin (eds.), *Morphological autonomy: Perspectives from Romance inflectional morphology*, 257–286. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199589982.003.0013

CHAPTER 20

A-prefixing in the ex-slave narratives

Janie Rees-Miller
Professor Emerita, Marietta College

The speech of elderly African-Americans collected in the ex-slave narratives during the 1930s can enhance our understanding of both form and meaning of a-prefixing. Drawing on a selection of interviews that includes approximately 400 tokens of a-prefixing from over 100 individuals, this study will examine the types of words on which a-prefixing occurs, the phonological constraints on it, and the morpho-syntactic patterns in which it is more likely. From these interviews we can infer a semantic meaning of intensification for a-prefixing, especially in narrative contexts involving emotionally-charged events. This semantic meaning of a-prefixing may be related to the development of the progressive.

Keywords: a-prefixing, ex-slave narratives, AAVE, progressive

1. Introduction

A-prefixing refers to a relic form in nonstandard American English in which *a-* is prefixed to a word, most typically a verb ending in *-ing*, e.g., The wind was *a-blowin'* and the boat was *a-rockin'*. Previous studies of a-prefixing have focused almost exclusively on its use by whites. For example, the various articles by Wolfram (1976, 1982, 1988) are based on recorded interviews with white speakers in southern West Virginia. Feagin's (1979) work on a-prefixing forms part of a larger scale investigation of variation and change in the speech of the white community in Anniston, Alabama, which is located within sight of the southern edge of the Appalachian mountains. The work by Christian, Wolfram & Dube (1987) also places a-prefixing within the larger context of variation and change in the white southern mountain speech of the Ozarks of Missouri and Arkansas and in the speech of Appalachia, drawing on the same West Virginia interviews as Wolfram (1976). Again dealing with white speakers in the mountain south, Montgomery's (2009) article is based on a corpus of English used in the Smoky Mountain area of the Tennessee-North Carolina border. Similarly, McQuaid's (2017) data is also drawn from eastern Tennessee. As a result of these various studies, a-prefixing has

https://doi.org/10.1075/cilt.353.20ree

been most associated with rural whites in the mountain south (Matyiku, McCoy & Martin 2018). However, the use of a-prefixing a century ago was much more widespread both geographically and across racial lines, as demonstrated in the ex-slave narratives collected during the Great Depression of the 1930s. Although the ex-slave narratives have certain limitations as a reliable source of precise linguistic data, if used with care, they can provide evidence in general of a-prefixing.

Some previous work on a-prefixing, notably by Wolfram (1976, 1988), focused on form-based analysis. Wolfram suggested morpho-syntactic and phonological constraints that operate on a-prefixing and downplayed its semantic significance. Some of Wolfram's constraints have been challenged (e.g., Montgomery, 2009), and these form-based constraints can be further tested in the current study of ex-slave narratives, which provide over 400 tokens of a-prefixing from 115 individuals. On the one hand, the form-based analysis of a-prefixing in this chapter shows that it generally follows the phonological and syntactic constraints posited by Wolfram (1976) with some exceptions. On the other hand, the ex-slave narratives provide a rich source of tokens for a meaning-based analysis of a-prefixing and its role in discourse. Not only is the form itself old, but so too is the intensive meaning that a-prefixing encodes when used in emotionally-charged narratives. In this context, a-prefixing in the ex-slave narratives preserves evidence of its association with the development of the progressive.

2. Who uses a-prefixing?

In twenty-first century English in the US, it does seem to be the case that a-prefixing is geographically limited and is actively used only in vernacular contexts by the oldest generation of white rural speakers. However, as nineteenth-century documents reveal, this was not always the case. And although Feagin (1979: 117) stated she had no knowledge of a-prefixing among African-Americans and suggested that its use might be limited to white speech, there are abundant examples of it from earlier African American English.

2.1 Distribution of a-prefixing in time and space

A-prefixing was commonly in use throughout the United States in the nineteenth and first half of the twentieth centuries. Examples have been noted in the Lewis and Clark journals from the early 1800s (Davis, Houck & Upton 2000: 145; Montgomery 2009: 7), and its use has been studied in Mark Twain's *Huckleberry*

Finn (Antieau 2001). By the time Mark Twain was writing in the late nineteenth century, a-prefixing was associated with informality and rusticity (Montgomery 2009: 7). In the first half of the twentieth century, a-prefixing was documented in use in New England, the mid-Atlantic states, the Midwest, the West, the South, and the mountain South (Wentworth 1944: 2–3; Atwood 1953: 34–35). Although it was becoming much less common by the later twentieth century, a-prefixing was still in common enough use to be studied in the South (Feagin 1979), in the Smoky Mountains (Montgomery 2009; McQuaid 2017), and in Appalachia and the Ozarks (Wolfram 1976; Christian et al. 1987). Some use has also been noted among older speakers in Illinois (Frazer 1990) and in Colorado (Antieau 2003: 397–398). As a native speaker of an a-prefixing dialect, I am particularly interested in the use of a-prefixing in other dialects.

2.2 African-American use of a-prefixing

Kautzsch (2008: 540) notes that a-prefixing is indeed very rare in modern-day African-American English; however, a-prefixing was a "fairly stable feature" of earlier African-American Vernacular English. As a matter of fact, documents from the nineteenth and early twentieth centuries give evidence of the use of a-prefixing by African Americans both before and after the Civil War. For example, Solomon Bayley's first person account of his enslavement, published in 1825 (partially reprinted in Carbado & Weise 2012), includes a cluster of a-prefixing in his story of his near-capture during escape, a story which he introduces as "the greatest trial I ever met with in all my distress" (pp. 120–122). Furthermore, songs collected by Union soldiers from freed slaves in South Carolina during the Civil War (Allen, Ware & Garrison 1867) contain multiple examples of a-prefixing in the lyrics, as illustrated in (1):

(1) a. "Day is *a-breakin*'" (Song 6, p. 4)
b. "Dere's no whips *a-crackin*'" (Song 61, p. 46)
c. " … she see a man *a-comin*'" (Song 73, p. 54)
d. "I'm *a-going* home." (Song 105, p. 84)

Later African-American writers, such as the poet Paul Laurence Dunbar (1872–1906; complete poems available at http://www.gutenberg.org/ebooks/18338), also used a-prefixing when writing in 'Negro dialect'. And eleven recordings made between 1935–1974 of former slaves describing their experiences contain 17 tokens of a-prefixing used by seven of the individuals (Bailey, Maynor & Cukor-Avila 1991).

3. The ex-slave narratives

Given that a-prefixing was indeed in use by African-Americans, it is not surprising that there are examples in the ex-slave narratives collected during the 1930s as part of the Federal Writers' Project in the Works Progress Administration. The program enlisted unemployed writers, anthropologists, and other educated professionals to interview former slaves and write narrative accounts based on their reminiscences about slavery, the Civil War, and African-American cultural beliefs and practices. Some 2000 of these narratives are housed today in the Library of Congress (available online at https://www.loc.gov/collections/slave-narratives-from-the-federal-writers-project-1936-to-1938/about-this-collection/), while a number of documents still reside in university and state libraries (e.g., Perdue, Barden & Phillips 1976: xxviii and xxxiv for Virginia; Federal Writers' Project 1937–1938 for Ohio).

3.1 The ex-slave narratives as linguistic data

Using these narratives as linguistic data is not without problems (see chapters by Montgomery and by Rickford in Bailey et al. 1991; also Myhill, 1995). With the exception of the eleven recordings of interviews transcribed in Bailey et al. (1991), all the other extant narratives were written down after the fact and could be subject to multiple levels of editing (Montgomery 1991; Perdue et al. 1976: xxii-xxvi). In addition, there are the confounding effects of an elderly, often desperately poor and uneducated former slave talking with a previously unknown interviewer of a considerably higher social status – confounding effects that were magnified if the interviewer was white.

Although the linguistic data in the ex-slave narratives is not without problems, it can still provide useful insights (Kautzsch 2008: 536). The interviewers by and large would have been familiar with the uses of a-prefixing and therefore likely to get it right, perhaps even using it themselves (as was the case with two interviewers in Bailey et al. 1991). Although most interviewers were white, in Virginia at least, there were a number of African-American interviewers, some quite skilled as ethnographers. These interviewers developed effective note-taking techniques, and a few interviews were even recorded on aluminum disks and transcribed by the interviewers (Perdue et al. 1976: xxxvi–xxxviii, xlii; also, Kautzsch 2008: 536). While this chapter will not offer precise counts, statistical analyses, or definitive conclusions about the history of African-American Vernacular English based on the data in the ex-slave narratives, the narratives *can* be used to indicate trends and draw conclusions about how a-prefixing was used in general and to test hypotheses proposed in previous studies. Furthermore, the ex-slave narratives provide a rich source of vivid accounts of life under slavery, with a-prefixing associated with strong emotional content for the informants.

3.2 Description of corpus

In selecting material for my corpus, I examined all narratives from Kansas, Kentucky, Maryland, and Ohio in the Library of Congress website (Federal Writers' Project 1936–1938, Volumes 6, 7, 8, 12), as well as additional interviews from Ohio (Federal Writers' Project 1937–1938, housed in the Ohio Historical Society) and the first 20 interviews from Georgia as reprinted by the Applewood Press (Federal Writers' Project 1936/2006). I also utilized some published collections of narratives (Hurmence 1997; Berlin, Favreau & Miller 1998; and Yetman 2002). From these various sources, I found 73 individuals who used a-prefixing at least once; the total number of tokens of a-prefixing was approximately 300 from these sources.

The sources mentioned in the previous paragraph are subject to all the caveats discussed in § 3.1 about reliability of the ex-slave narratives. Two other collections (Bailey et al. 1991; Perdue et al. 1976), though, stand in contrast because of their greater likelihood of representing the actual speech of the ex-slaves. The eleven ex-slave recordings published in Bailey et al. (1991) were transcribed and checked for accuracy by a team of linguists. The Perdue et al. (1976) collection is particularly important because not only does it contain a wealth of interviews from ex-slaves in Virginia, which is poorly represented by manuscripts in the Library of Congress, but the majority of these interviews were conducted by skilled African-American interviewers. From these two collections, 42 individuals used a-prefixing for a total of over 130 tokens.

Thus, a total of 115 individuals are represented in this study of selected ex-slave narratives with approximately 430 tokens of a-prefixing. The individuals represent enslavement in a variety of states, as shown in Table 1. Although the geographical distribution is far from even, it is not the purpose of this chapter to investigate regional variation in African-American use of a-prefixing.

Also, in contrast with Feagin (1979: 101), who observed that a-prefixing was much more common among women than men in the white speech community of Anniston, Alabama, in the ex-slave narratives examined here the distribution was relatively even among male and female informants: 55 males and 60 females used a-prefixing.

Table 1. Number of individuals from each state

State	Individuals (n) = No. males + No. females		
Virginia	49	24	25
Kentucky	14	7	7
Texas	13	5	8
Georgia	9	5	4
North Carolina	9	4	5
Mississippi	6	3	3
Alabama	3	1	2
West Virginia	3	1	2
Arkansas	2	1	1
Missouri	2	1	1
Louisiana	1	1	–
Maryland	1	1	–
Oklahoma	1	–	1
South Carolina	1	1	–
Tennessee	1	–	1
Total	**115**	**55**	**60**

4. Form-based analysis

Much previous work on a-prefixing has concerned itself with form. Feagin (1979: 107–109) investigated the types of words to which a-prefixing attaches, as well as exploring meaning. Other researchers (e.g., Wolfram 1976; Montgomery 2009; Antieau 2001) have concentrated on form with examinations of phonological constraints and analyses of the morpho-syntactic distribution of a-prefixing

4.1 Forms to which a-prefixing attaches

In the ex-slave narrative data here, approximately 95% of the words to which the *a-* prefix attaches are verbs ending in *-ing,* comprising 172 verb lexemes. Of these verb lexemes, 85% can be classed in one of three general categories: approximately 37% are action verbs with an animate subject (e.g., *work*, *hunt*, *carry*), followed by 26% movement or rest verbs (e.g., *go*, *dance*, *trot*, *limp*, *lay*), and 22% having to do with speech or other sounds (e.g., *yell*, *groan*, *tell*, *preach*, *knock*). Similar to the verbs noted by Feagin (1979: 107–109), these are generally verbs which can be used in the progressive (as opposed to stative verbs, which cannot).

In addition to present participles, there are a few examples of other types of words with a-prefixing. In a few cases, the simple present or past tense verb has a-prefixing: *a-start, a-took, a-dyed*. There is one example of a-prefixing on a compound verb, which occurs when an informant refuses to tell his interviewer any "sad stories 'bout beatin' an' *a-slave drivin*'" (Samuel Sutton, enslaved in Kentucky and interviewed in Ohio, quoted in Federal Writers' Project 1936–1938, Vol. 12: 96). There are also a couple of examples of a-prefixed adjectives: *a-loose* and *a-scared*.

4.2 Phonological constraints

In Wolfram's discussion of phonological constraints on a-prefixing (1976: 50–51), he posits that a-prefixing is prohibited on words that are not stressed on the first syllable and on words that begin with a vowel. As for the first constraint, the ex-slave narratives provide no counter examples; all a-prefixed words in the ex-slave narratives under study are stressed on the first syllable. Indeed, almost 90% of the words with a-prefixing are monosyllables; the few two-syllable words to which a- is prefixed (e.g., *holler, happen, listen*) all have initial stress. As for Wolfram's second constraint – prohibiting a-prefixing on words starting with a vowel – the a-prefixed words in these ex-slave narratives almost all begin with a consonant as predicted. However, there are two examples each of *a-eatin*' and *a-askin*', coincidentally the two starred example verbs used by Wolfram to illustrate the constraint against a-prefixing on vowel-initial words (1976: 51). In his article, however, Wolfram (1976: 51 n. 4) noted counterexamples stating, "… it is possible that this constraint is variable rather than categorical for some speakers." For those speakers, including speakers of African-American Vernacular English, the general constraint against two adjacent vowels is weaker than is the case in other dialects.

4.3 Morpho-syntactic distribution of a-prefixing

The common morpho-syntactic uses of a-prefixing in the ex-slave narratives correspond fairly closely to those documented in Appalachian speech (Wolfram 1976: 47–48), in Smoky Mountain speech (Montgomery 2009: 18), and in *Huckleberry Finn* (Antieau 2001: 151). According to these sources, a-prefixing commonly occurs: with a verb in the progressive; as an adverbial complement to verbs of movement, perception, starting, or continuing; in other adverbial constructions; and in reduced relative clauses.

For Wolfram's West Virginia data (1976: 47) and the corpus in *Huckleberry Finn* (Antieau 2001: 151), the most common position for an a-prefixed verb + *-ing* is in the progressive. In *Huckleberry Finn*, 41% of the a-prefixed verbs are in the progressive (Antieau 2001: 151). Likewise, in the ex-slave narratives, just over 40% of all tokens of a-prefixing are used in the progressive aspect, with or without the auxiliary *be*, as illustrated in (2):

(2) a. [Reporting her mother's words during a meteor shower]: "Just look up at what is *a-happenin'*."
(Sarah Gudger, ex-slave in North Carolina, quoted in Yetman, 2002: 68)
b. [Describing his brother's beating]: "His lips wuz *a-quiverin'* and his body wuz *a-shakin'*." (William Colbert, former slave in Georgia, quoted in Berlin et al. 1998: 26)
c. "He didn' get up, he jus' *a-praying*."
(Laura Smalley of Texas, in Bailey et al. 1991: 70)
d. "… dem Yankees be *a-sittin'* along de road wid dey blue coats on."
(Candis Goodwin of Virginia, quoted in Perdue et al. 1976: 108)

As the examples in (2) indicate, a-prefixing could be used on the progressive in the present tense as in (a), in the past – with or without the auxiliary – as in (b) and (c), and even as in (d) with invariant *be* before V-*ing,* a construction that Kautzsch (2008: 540) identifies as a post-Civil War innovation in African-American Vernacular English.

A-prefixing is also relatively common as an adverbial complement after verbs of movement, perception, starting, and keeping. As for verbs of movement, the phrase *come a-runnin'* is used by at least six different individuals, all originally from different states in the South. Other examples of adverbial complements are illustrated in (3):

(3) a. [Reporting his escape to Yankee lines]: "Heered the hounds *a-howling*, getting ready to chase us."
(Boston Blackwell, former slave in Arkansas, quoted in Yetman 2002: 11)
b. [Reporting a gale from the north in Texas]: "When dat nawther [norther] struck, weeds and leaves would jes' staht [start] *a-rollin'* along de land."
(Martha Spence of Texas, quoted in Berlin et al. 1998: 142)
c. [Describing a "conjurer" approaching her house]: "Her voice kep' *a-gittin'* nearer." (Matilda Herrietta "Sweet Ma" Perry of Virginia, quoted in Perdue et al. 1976: 223)

The use of a-prefixing in other adverbial constructions is relatively common in the ex-slave narratives under study here, as illustrated in (4):

(4) a. "Durin' de big war, all the white folkses was off *a-fightin'* …"
(Jasper Battle of Georgia, quoted in Federal Writers' Project 1936/2006: 67)
b. [During church]: "We waited on de white folks, *a-totin'* water an' seein' bout de horses and buggies, an' *a-tendin'* to de chillun."
(Vinnie Busby of Mississippi, quoted in Berlin et al. 1998: 16)

One relatively common use of a-prefixing in Smoky Mountain English, as reported in Montgomery (2009), is in a reduced relative clause. This use also occurs in the ex-slave narratives, as shown in (5)

(5) a. "… there come a whole troop of Yankees, all riding horses, big guns *a-hanging* on in there …"
(Uncle Billy McCrea of Texas, in Bailey et al. 1991: 42)

b. "De fields would be full o' slaves *a-wukin'* [a-working] hard."
(Manda Boggan of Mississippi, quoted in Berlin et al. 1998: 41)

Although Wolfram (1976: 49) asserts that a-prefixing is not found with nominals, some few examples from the ex-slave narratives seem very noun-like. According to Wolfram, use of a-prefixing is prohibited immediately after a preposition, although Montgomery (2009: 45) cites a few counter-examples from Smoky Mountain English. Likewise, in the ex-slave narratives studied here, there are a handful of examples in which an a-prefixed V-*ing* form immediately follows a true preposition as its object, as exemplified in (6):

(6) a. "… ah [I] sho [sure] do believe in *a-havin'* church …"
(Samuel Sutton, formerly enslaved in Kentucky and interviewed in Ohio, quoted in Federal Writers' Project 1936–1938, Vol. 12: 95)

b. [Reporting advice she gave to a landowner who was considering killing a worker]: "I wouldn't have the name of *a-killing* my hands on the place."
(Laura Smalley of Texas, in Bailey et al. 1991: 76)

c. "… she 'sisted [insisted] on *a-comin'* in the house."
(William Yager of Virginia, quoted in Perdue et al. 1976: 340)

One noun-like structure in the form shown in (7) occurs in the narratives of eight separate individuals, as illustrated in (8).

(7)

$$\ldots\text{such}\quad a\text{-}V\text{-}ing \begin{bmatrix} \text{I} \\ \\ \text{you} \end{bmatrix} \quad \text{never} \begin{bmatrix} \text{saw} \\ \\ \text{heard} \end{bmatrix}$$

(8) a. [Describing the wake for a cruel white man laid out on his bier] "… and all of a sudden Mister Jim rolled off'n the coolin' board, and sich [such] *a-runnin'* and gittin' out'n that room you never saw."
(Julia Brown of Georgia in Federal Writers' Project 1936/2006: 145)

b. "…you would see gangs of slaves chained together [to be sold] un, un, setch [such] *a-cryin'* an' screamin' you ain't nebber [never] heard like dem pitiful cries of dem po' slaves."
(Samuel Walter Chilton of Virginia quoted in Perdue et al. 1976: 71)

c. [Describing a memorable party given by the master] "Sech [such] *a-dancin'* you never seed befo' …"
(Nancy Williams of Virginia, quoted in Perdue et al. 1976: 318)

In all these cases, I would argue that the a-prefixed V-*ing* is acting as a nominal, and that the *a*- is not a determiner (which would not be required by the structure in any case) but a true a-prefix, used on a gerund for a semantic reason, namely to intensify the meaning.

5. Meaning-based analysis

Wolfram (1976: 54–55) found "no evidence for a distinct semantic category of a-prefixing" since there was no obligatory context in which a-prefixing was required. McQuaid (2017: 151–153) neatly addresses this argument by interpreting a-prefixing within the framework of Optimality Theory; a-prefixing could be required in certain contexts but is blocked by higher-ranking phonological constraints. She posits that the *a*- prefix encodes the pragmatic feature [MIRATIVE], used for surprise occasioned by a sudden discovery or revelation or by something counter to expectations. In the ex-slave narratives, the meaning of a-prefixing for something unexpected is well-illustrated in an interview with Virginia Hayes Shepherd, a former slave in Virginia, explaining the effects of a yellow fever epidemic (Perdue et al. 1976: 258): "Everywhere you went the slaves were sitting on the front porches just *a-rockin'* – white folks all dead." This exemplifies the element of surprise; during the time of slavery, the last thing one would expect to see was slaves taking their ease in rocking chairs on the front porch.

However, a-prefixing in the ex-slave narratives was not limited to unexpected events; indeed, it was frequently used for describing beating and other punishments commonly meted out to slaves. In these instances, rather than encoding surprise or the unexpected, the *a*- prefix encodes intense emotive meaning more generally and is often accompanied by multiple intensifiers.

5.1 Intensive meaning

In studies of modern a-prefixing, Christian et al. (1987: 61) conclude that it may serve a nonunique discourse function of intensification in Appalachian and Ozark English. Feagin (1979: 113) notes in her corpus from Anniston Alabama that a-prefixing is often accompanied by *just* and adds intensity and vividness to a narrative.

Of the over 400 tokens of a-prefixing in the ex-slave narratives under study here, just over 100 tokens – nearly 25% – were accompanied by some other forms of intensification as well, indicating a context that was highly charged with emotion. For example, W. L. Bost, an ex-slave from North Carolina, remembered a woman being sold at auction, traditionally held on January 1 (Yetman 2002: 16): "The poor thing stand on the block *a-shiverin'* and *a-shakin'* nearly froze to death." Typically the slave on sale was stripped of clothing, and the intensity of the cold is emphasized through a-prefixing on the coordinate structure *a-shiverin' and a-shakin'* along with the phrase *nearly froze to death* to reinforce the harsh effect.

Likewise, markers of intensification such as *just*, *keep*, and *still* are often used with a-prefixing. For example, in the account of the ultimately fatal beating administered over several hours to a slave on a neighboring plantation, Fanny Berry, a former slave in Virginia, uses six tokens of a-prefixing and multiple other intensifiers:

> [At the start of day, we] heard the most awful screamin' an' *a-hollerin'*… At noon dey was *still a-beatin'* him. … the overseer was *jus' a-layin'* et [it] on … [The slave] was *jus' a-moanin' an' a-groanin'* and beggin' … but dat ole white man *kep' a-layin'* it on. (Perdue et al. 1976: 43)

Not only does she use a cluster of a-prefixed verbs, but she strengthens the emotive force of the a-prefixed verbs with intensifiers *keep* (which serves to stress the continuous nature), *still*, and *just*. She intensifies the meaning further with coordinate structures and semantically intense vocabulary: *screamin' an' a-hollerin'* and *a-moanin', and a-groanin' and beggin'*.

5.2 A-prefixing in narratives

It is consistent with the tendency for a-prefixing to mark intensity that we would find a-prefixing in narratives, particularly those associated with strong emotion. For example, in *Huckleberry Finn*, there is some clustering of a-prefixing in narratives of the shipwreck and of Pap going to court, both very emotional scenes for Huck (Antieau 2001: 153). Wolfram (1988: 249) noted that approximately 2/3 of the tokens of a-prefixing in his corpus of Appalachian English from West Virginia occurred in narratives. And Frazer (1990) noted that his examples of a-prefixing in Illinois seemed to be associated with "narratives of unusual drama or excitement."

The ex-slave narratives certainly provide multiple opportunities for informants to tell "narratives of unusual drama or excitement," and approximately 3/4 of individuals who used a-prefixing used one or more tokens of it in the context of a narrative, with a total of 129 separate narratives containing a-prefixing. Tokens in these narratives accounted for approximately 2/3 of the tokens of a-prefixing in the corpus, a figure that corresponds with Wolfram's findings. Of the 129 narratives, 36% concerned the life of enslavement and its aftermath – punishments, slave auctions, escape, and restrictions on movement. For example, West Turner, a former slave in Virginia, told a colorful story replete with a-prefixing and other forms of intensification about his successful attempt to divert and head off the patrollers (called *paddyrollers*), whose mission was to round up and punish any slaves found off their plantation at night (Perdue et al. 1976: 290):

> … de paddyrollers come *a-chasin'* arter [after] me, *jus' a gallopin'* down de lane *to beat de band*. [He hides by quickly ducking off the road into the woods] Course the paddyrollers couldn't stop so quick … an' den dere came *a-screamin'* an' cryin' dat make you think dat *hell done bust loose* … [The patrollers became entangled in grapevines that had been strung across the road.] An' some done landed mighty hard, cause dey was *a-limpin'* … and callin' fo' de slaves to come an' help dem, but dem slaves got plenty o' sense. Dey lay in the bushes an' *hole dere sides a-laughin'* …

War-related subjects accounted for 25% of those narratives containing a-prefixing, including descriptions of battles, the coming of soldiers, and learning of emancipation. An example from Felix Haywood, a former slave on a Texas ranch, tells of the end of the war (Berlin et al. 1998: 265): "Soldiers, all of a sudden, was everywhere – comin' in bunches, crossin' and ridin'. Everyone was *a-singin'*. We was all walkin' on golden clouds."

A further 12% of narratives concern beliefs and practices, either in the form of superstitions and ghost stories or of religious conversion and beliefs. A-prefixing abounds in the narrative Fannie Moore, an ex-slave from North Carolina, tells of her mother's conversion: "… she sta't [start] singin' an' *a-shoutin'* an' *a-whoopin'* an' *a-hollowin* [a-hollerin']' …" (Federal Writers' Project 1936–1938, Vol. 11, Part 2: 130).

Approximately 9% of narratives have to do with family and life passages – birth, marriage, death. The narrative told by Patience Avery, a former slave in Virginia, uses a cluster of a-prefixing to convey her horror as a child being introduced to her father and discovering that he was white (Perdue et al. 1976: 15):

> I woke up jes' cryin' an' *a-rubbin'* my eyes an' *a-ketchin'* my breath between sobs. … I put one han' up to my face an' was rubbin' my eyes an' *a-cryin'*, an' *a-snifflin'*, an' *a-sayin'*, "I ain't got no father … He no father o' mine! He white!"

Table 2. Narratives containing a-prefixing

Topic of narrative	Number (percentage) of narratives	Number (percentage) of tokens
Life of enslavement and after effects	47 (36%)	94 (32%)
War-related stories	32 (25%)	64 (22%)
Religious or superstitious beliefs/practices	15 (12%)	37 (13%)
Family/ life passages	12 (9%)	22 (7%)
Other (e.g. memorable dream, job after war, unusual weather, etc.)	23 (18%)	78 (26%)
Totals:	**129**	**295**

Table 2 summarizes the numbers and subjects of narratives containing a-prefixing. In all these narratives, the effect of the a-prefixing is to add to the intensity of emotion called forth by the events. The origins of this effect may be related to the development of the progressive, which will be taken up in the next section

6. Coalescence of form and meaning: A-prefixing and the progressive

The development of the progressive form and aspect can shed light on the semantic and discourse functions of a-prefixing. As mentioned in § 4.3, over 40% of the tokens of a-prefixing in the ex-slave narratives under examination here were used as part of a progressive verb form. In addition, as Feagin (1979: 107–109) noted in her corpus and as is true in the ex-slave narratives, the verbs to which a-prefixing attaches are typically those which can also appear in the progressive. Citing the work of Geoffrey Leech, Feagin (1979: 104) also noted that – in addition to expressing duration or incompleteness – the progressive could also lend immediacy and vividness to the discourse, and Biber et al. (1999: 471–473) note that the progressive today is more common in conversation and fiction than in academic or other formal writing. These descriptions of the progressive sound very like some of the uses of a-prefixing in the ex-slave narratives, where it is commonly used in dramatic narrative contexts. Although the exact origin of the progressive is still under debate, one proposal is that the progressive developed from a form with a-prefixing (van Gelderen 2006: 216; Bybee & Dahl 1989: 79). A second proposal holds that the progressive developed from a structure in Old English using *beon* or *wesan* plus the present participle of the verb (Kranich 2013: 21 n. 13 states this as definite; Sairio 2018: 181 is less certain). A third suggestion splits the difference: the progressive arose in early Modern English from a merger of the a-prefixed form (originally a preposition plus gerund) with Old English *beon* or *wesan* plus present participle (Elsness 1994; Nuñez Pertejo 1996). Whichever the case may be, a-prefixing is closely intertwined with the development of the progressive.

Bybee & Dahl (1989: 81) argue that the English progressive today has rich nuances of meaning because of its original semantics, and Kranich (2013) argues that because of the history of the progressive, it has acquired a layered meaning. I contend that this likewise applies to a-prefixing. In Old English and Middle English, the progressive form could be used for two purposes: (1) the aspectual purpose: to highlight the durative nature of a situation (a use which was not predominant until hundreds of years later in the nineteenth century); and (2) the subjective purpose: to highlight a remarkable situation (a use which still lingers today despite being overtaken in frequency by the aspectual use) (Kranich 2013: 21–22, 26; Sairio 2018: 182). Indeed, the subjective use was predominant in Old and Middle English and found in narratives for dramatic events (Kranich 2013: 25–26). Because of its use in narrative, which tends to be less formal, the progressive form was used for more colloquial genres by the eighteenth century (Sairio 2018: 191–193), and that usage continued into the nineteenth century (Sairio 2018: 181). As Kranich (2013) argues, although the aspectual (durative) use of the progressive is predominant today, the subjective meaning of the progressive lingers from an earlier time.

The association of a-prefixing with the progressive would account for the fact noted by Montgomery (2009: 7) that a-prefixing became viewed as informal, even colloquial, in the eighteenth century and eventually as rustic in the nineteenth century. In the ex-slave narratives, whether a-prefixing is used on the progressive verb or in other structures, it preserves the earlier subjective purpose of the progressive, which was to highlight a remarkable situation. I would argue that for those who have a-prefixing in their native grammar, including the ex-slaves, a-prefixing can be chosen to express that subjective meaning as distinct from the purely aspectual role of the unprefixed progressive verb.

The choice of a-prefixing for subjective purposes can be illustrated by the narrative of Elizabeth Sparks, an ex-slave interviewed in Virginia (in Perdue et al. 1976: 273–277). She exploits different nuances of expressions of the future, including a-prefixing on the progressive *be going to* form, as she decides how much to tell her interviewer about slavery. At the beginning of the interview, after greeting her African-American interviewer and hearing his request to tell about slavery days, she says:

> Well, I kin [can] tell yer [you], but I ain't. … 's too awful to tell anyway. Yer're too young to know all that talk anyway. Well, I'll tell yer some to put in yer book, but I ain't*a goin'* tell yer the worse. (quoted in Perdue et al. 1976: 273)

She makes clear through her use of *kin* [*can*] that she has the ability – the experience and the memory – to tell brutal stories, but chooses not to do so in order to spare her younger listeners and herself the pain of reliving the trauma. Using the simple future with *will*, indicating a decision at the time of speaking, she grudgingly offers to provide some information, but her use of *be going to* – a progressive form with future meaning – denotes planned intent. The a-prefixing on *ain'ta goin' tell yer the worse* makes it a definitive refusal.

Then after offering some information about her former young owner, who was relatively good to her, Elizabeth Sparks once again categorically refuses to describe the young mistress's "rough" parents:

> Kin I tell yer about her parents? Lord, yes. … But I ain't*a goin'* tell yer nuffin [nothing]. No, I ain't. T'ain't no sense for yer ta know 'bout all those mean white folks. Dey all daid [dead] now. … Well, I'll tell yer some, but I ain't*a goin'* tell yer much more. (Perdue et al. 1976: 273–274)

Once again, using simple future with *will*, she decides at the moment of speaking to provide some limited information, but her considered intent of refusal expressed through twice-repeated negative *be going to* is intensified and made more definite with a-prefixing both times.

That the a-prefixing in this narrative was a choice determined by the speaker's intended meaning is confirmed by other uses of the *be going to* future without a-prefixing in the same narrative. For example, in her reminiscence about the arrival of Yankee soldiers, who took away the male slaves as contraband, Elizabeth Sparks' husband did not want to leave because "he didn't know whuh [where] they'*s takin'* 'em nor what they'*s gonna* do. … They *was gonna* take 'im anyhow …" (Perdue et al. 1976: 275). In fact, he was ultimately allowed to stay with his wife rather than being taken away. In both cases of *be going to* in this part of the narrative, as well as in the progressive *they's takin'*, a-prefixing is not used, indicating that these were not so definitive or highly charged for the speaker as her other uses with a-prefixing. In other words, the availability of a-prefixing provided a means by which the speaker could express her subjective purpose.

7. Conclusion

This chapter has examined the use of a-prefixing in the ex-slave narratives collected in the 1930s. These narratives demonstrate clearly that a-prefixing was a feature used in African-American speech at the time and not just limited to use by whites. As used in these narratives, a-prefixing generally follows the phonological and morpho-syntactic patterns and constraints set out in earlier form-based studies, although as Wolfram (1976: 51) concedes, some of the constraints should be understood as violable rather than as categorical. As in other studies of a-prefixing, the types of words to which a- is prefixed here are almost all verbs ending in *-ing*, most of which can be broadly categorized as action verbs, movement verbs, or verbs of speech or sound. In terms of phonology, a-prefixing here also is found only on words with initial stress, but there are a few examples of a-prefixing on words beginning with a vowel. Again similar to other studies of morpho-syntactic distribution, the investigation here found a-prefixing on verbs in the progressive, as complements after certain types of verbs (movement, perception, starting, and continuing), in adverbial constructions, and in reduced relative clauses. Unlike some previous predictions, the ex-slave narratives provide some noun-like examples where a-prefixing is used.

The main contribution that this study offers, though, is the meaning-based analysis, which confirms the semantic and discourse functions of a-prefixing. A-prefixing does not fill a unique niche where it is obligatory, but it offers the speaker a way to highlight an extraordinary situation, as shown by the ex-slaves' use of a-prefixing with intensive meaning and in emotionally-charged narratives, such as those having to do with various emotionally-charged aspects of the life of enslavement, war-related stories, religious or superstitious beliefs, and life passages. The speaker's choice of a-prefixing can be related to the development of the progressive from its subjective purpose. For speakers who have a-prefixing in their grammar, it offers the option of expressing intense emotional content grammatically.

References

Allen, William Francis, Charles Pickard Ware & Lucy McKim Garrison. 1867. *Slave songs of the United States*. New York: A. Simpson. Available from https://library.si.edu/digital-library/book/slavesongsofunit00alle. https://doi.org/10.5479/sil.78680.39088002074508

Antieau, Lamont. 2001. "I'm a-going to see what's going on here": A-prefixing in *The adventures of Huckleberry Finn*. *Language and Literature* 10(2). 145–157. https://doi.org/10.1177/0963-9470-20011002-06

Antieau, Lamont. 2003. Plains English in Colorado. *American Speech* 78(4). 385–403. https://doi.org/10.1215/00031283-78-4-385

Atwood, E. Bagby. 1953. *A survey of verb forms in the eastern United States*. Ann Arbor, MI: University of Michigan Press. https://doi.org/10.3998/mpub.9690449

Bailey, Guy, Natalie Maynor & Patricia Cukor-Avila (eds.). 1991. *The emergence of Black English*. Amsterdam: John Benjamins. https://doi.org/10.1075/cll.8

Berlin, Ira, Marc Favreau & Steven Miller. 1998. *Remembering slavery: African Americans talk about their personal experiences of slavery*. New York: The New Press.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman grammar of spoken and written English*. Harlow: Pearson Education.

Bybee, Joan L. & Östen Dahl. 1989. The creation of tense and aspect systems in the languages of the world. *Studies in Language* 13(1). 51–103. https://doi.org/10.1075/sl.13.1.03byb

Carbado, Devon W. & Donald Weise (eds.). 2012. *The long walk to freedom*. Boston: Beacon Press.

Christian, Donna, Walt Wolfram & Nanjo Dube. 1987. *Variation and change in geographically isolated communities: Appalachian English and Ozark English*. PADS 74. Tuscaloosa, AL: University of Alabama Press.

Davis, Lawrence M., Charles L. Houck & Clive Upton. 2000. "Sett out very eairly Wensdy": The spelling and grammar in the Lewis and Clark journals. *American Speech* 75(2). 137–148. https://doi.org/10.1215/00031283-75-2-137

Elsness, Johan. 1994. On the progression of the progressive in early Modern English. *ICAME Journal* 18. 5–25.

Feagin, Crawford. 1979. *Variation and change in Alabama English: A sociolinguistic study of the white community*. Washington DC: Georgetown University Press.

Federal Writers' Project. 1936/2006. *Georgia slave narratives*. Carlisle, MA: Applewood Books.

Federal Writers' Project. 1936–1938. Slave Narrative Project. Vol. 6, *Kansas*. Retrieved from www.loc.gov/item/mesn/060

Federal Writers' Project. 1936–1938. Slave Narrative Project. Vol. 7, *Kentucky*. Retrieved from www.loc.gov/item/mesn/070

Federal Writers' Project. 1936–1938. Slave Narrative Project. Vol. 8, *Maryland*. Retrieved from www.loc.gov/item/mesn/080

Federal Writers' Project. 1936–1938. Slave Narrative Project. Vol 11, Part 2, *North Carolina*. Retrieved from www.loc.gov/item/mesn/112

Federal Writers' Project. 1936–1938. Slave Narrative Project. Vol.12, *Ohio*. Retrieved from www.loc.gov/item/mesn/120

Federal Writers' Project. 1937–38. *The African-American experience in Ohio*. Retrieved from http://dbs.ohiohistory.org/africanam/html/mss/gr7999.html

Frazer, Timothy C. 1990. More on the semantics of a-prefixing. *American Speech* 65(1). 89–93. https://doi.org/10.2307/455942

Gelderen, Elly van. 2006. *A history of the English language*. Amsterdam: John Benjamins. https://doi.org/10.1075/z.135

Hurmence, Belinda. 1997. *Slavery time when I was chillun*. New York: G.P. Putnam's.

Kautzsch, Alexander. 2008. Earlier African-American English: Morphology and syntax. In Edgar W. Schneider (ed.), *Varieties of English 2: The Americas and the Caribbean*, 534–550. Berlin: Mouton de Gruyter.

Kranich, Svenja. 2013. Functional layering and the English progressive. *Linguistics*, 51(1). 1–32. https://doi.org/10.1515/ling-2013-0001

Matyiku, Sabina, Tom McCoy & Katie Martin. 2018. A-prefixing. *Yale Grammatical Diversity Project: English in North America*. Retrieved July 24, 2019 from http://ygdp.yale.edu/phenomena/a-prefixing

McQuaid, Goldie Ann. 2017. Morpho-pragmatic faithfulness interacts with phonological markedness in Appalachian a-prefixing. *University of Pennsylvania Working Papers in Linguistics*, 23(1). 151–159.

Montgomery, Michael. 1991. The linguistic value of the ex-slave recordings. In Guy Bailey, Natalie Maynor & Patricia Cukor-Avila (eds.), *The emergence of Black English*, 173–189. Amsterdam: John Benjamins. https://doi.org/10.1075/cll.8.08mon

Montgomery, Michael B. 2009. Historical and comparative perspectives on a-prefixing in the English of Appalachia. *American Speech* 84(1). 5–26. https://doi.org/10.1215/00031283-2009-002

Myhill, John. 1995. The use of features of present-day AAVE in the Ex-Slave Recordings. *American Speech* 70(2). 115–147. https://doi.org/10.2307/455812

Nuñez Pertejo, Paloma. 1996. On the origin and history of the English prepositional type *a-hunting*: A corpus-based study. *Revista Alicantina de Estudios Ingleses* 9. 105–117. https://doi.org/10.14198/raei.1996.9.07

Perdue, Charles L., Thomas E. Barden & Robert K. Phillips (eds.). 1976. *Weevils in the wheat: Interviews with Virginia ex-slaves*. Charlottesville, VA: University of Virginia Press.

Rickford, John R. 1991. Representativeness and reliability of the ex-slave materials, with special reference to Wallace Quarterman's recording transcript. In Guy Bailey, Natalie Maynor & Patricia Cukor-Avila (eds.). *The emergence of Black English*, 191–212. Amsterdam: John Benjamins. https://doi.org/10.1075/cll.8.09ric

Sairio, Anni. 2018. Incipient and intimate: The progressive aspect. In Terttu Nevalainen, Minna Palander-Collin & Tanja Saïly (eds.), *Patterns of change in 18th-century English: A sociolinguistic approach*, 179–196. Amsterdam: John Benjamins. https://doi.org/10.1075/ahs.8.11sai

Wentworth, Harold. 1944. *American dialect dictionary*. New York: Thomas Crowell.

Wolfram, Walt. 1976. Toward a description of a-prefixing in Appalachian English. *American Speech* 51(1/2). 45–56. https://doi.org/10.2307/455354

Wolfram, Walt. 1982. Language knowledge and other dialects. *American Speech* 57(1). 3–17. https://doi.org/10.2307/455176

Wolfram, Walt. 1988. Reconsidering the semantics of a-prefixing. *American Speech* 63(3). 247–253. https://doi.org/10.2307/454821

Yetman, Norman R. (ed.). 2002. *When I was a slave: Memoirs from the slave narrative collection*. Mineola, NY: Dover.

CHAPTER 21

Trajectory of children's verb formation in Hebrew as a heritage language

Dorit Kaufman
Stony Brook University

The trajectory of language shift and attrition in the Israeli immigrant community in the United States represents a complex interaction between Hebrew, a revitalized language of heritage, and English, a language that is highly pervasive in the Israeli urban landscape. The language of the children whose Hebrew is still developing at the onset of contact with English attests to morphological interaction and innovative lexical constructions that are unprecedented in monolingual language development for Hebrew or English. Analysis of longitudinal, narrative, and innovative verb formation data in children's Hebrew language oral production confirms the vulnerability of derivational morphology and verb formation in Hebrew and the increasing diffusion of English lexical formation devices as proficiency in Hebrew declines.

Keywords: language attrition, Hebrew as a heritage language, verb formation, derivational morphology, lexical innovations

1. Introduction

Diversification of sociocultural and linguistic landscapes of cities globally has enhanced interest in exploring the processes and variables affecting heritage language use, maintenance, and waning intergenerational transmission in immigrant communities. Reduction in exposure to the heritage languages has led to changing communication patterns within families and to progressive decline and gradual loss of proficiency among the younger generation. Research into language attrition was spearheaded by Lambert & Freed (1982) and has continued to be the subject of widespread investigation encompassing a wide array of contexts. These have spanned loss of international languages learned in classroom contexts, loss of languages of returnees to their homelands, and loss of heritage languages among immigrants (Haugen 1969; Ferguson & Brice Heath 1981; Berman & Olshtain, 1983; Fishman

https://doi.org/10.1075/cilt.353.21kau

1989; Kaufman & Aronoff 1989, 1991; Seliger & Vago 1991; Extra & Verhoeven 1993; Silva-Corvalan 1994; Garcia & Fishman 1997; Hansen 1999; Ammerlaan et al. 2001; Schmid 2011; Schmid & Köpke. 2017; Kopke & Genevska-Hanke 2018; Polinsky 2018). Common to these diverse language attrition scenarios is decreased exposure and progressive decline in proficiency in one language due to linguistic, sociocultural, and affective variables and preference for another language. The process, patterns, and outcomes of language maintenance or attrition vary widely among different age groups. Language attrition research among young children, speakers of languages including Garo, French, Hebrew, German, Czech, English, and Japanese, has demonstrated robust changes in the children's first or second languages shortly after departure from the country where the language is dominant and its exposure is reduced to the home domain (Burling 1978; Celce Marcia 1978; Kaufman & Aronoff 1991; Saville-Troike, Pan & Dutkova 1995; Yukawa 1997; Flores 2014). Long-term migrants, who were born in, or arrived as young children to the country to which their family had relocated, have documented widening gaps in their mother tongue, and challenges in comprehension and lexical access in their heritage languages (Hoffman 1989; Fein 1996; de Bot & Stoessel 2000; de Bot & Schrauf 2009; Hardach 2018; Lam 2019). In immigrant communities, young children and pre-adolescents are the ones driving the process of attrition in the heritage language. For the younger generation, the bond with the heritage language is weaker than that of their parents and successful integration in the new society is uppermost. As a result, intergenerational linguistic shift may occur within a single generation, as documented for immigrant groups in the United States, including Mexican, Cuban, Chinese, Filipino, Korean, and Vietnamese (McKay & Wong 2000); in Australia, for Dutch, German and Greek (Romaine 1991) and in other countries including India, Japan, Finland, the Netherlands (Broeder & Extra 1999; Fase, Koen & Kroon 1992).

The present study is part of a larger project undertaken to investigate aspects of maintenance, attrition, and interpersonal communication patterns in Hebrew, the heritage language of Israeli immigrant families in the United States (Kaufman & Aronoff 1989, 1991; Kaufman 1995, 2000, 2001, 2004). The present study focuses on the significance of Hebrew as a heritage language, the impact of English prior to immigration and the process of intergenerational language shift and changes in Hebrew verb formation strategies among pre-adolescent children who were born in, or arrived in the United States prior to having acquired literacy in formal academic settings in Israeli schools. Acquisition, maintenance, and attrition of Hebrew as a language of heritage does not operate in a vacuum. It is therefore important to examine the sociolinguistic and cultural variables and experiences that have shaped Israeli immigrant families' identity prior to immigrating to the United States and the significance of the Hebrew language in their lives. Equally significant is the status of English in Israel and the role it plays in the sociolinguistic and cultural landscape of the country. These play a critical role in Israeli immigrant families' perspectives, attitudes, and adaptation to their new environment in the United States.

2. Hebrew as a heritage language

The Hebrew language, the official language of Israel since 1948, is used in all spheres of modern life by millions of Israelis and by Jews worldwide connecting the land of Israel with the diaspora. The language has provided a shared linguistic and cultural medium for immigrant communities arriving in Israel from around the globe bridging past and present, secular and religious. The strong connection that binds Israelis to Hebrew is rooted in the history of this ancient language and its renaissance in the 19th century. Hebrew has been an integral part of the Jewish experience since ancient times. However, it had been dormant for nearly 2,000 years ceasing to exist as a language of oral discourse and communication, but remaining as a written language of scripture, scholarly writing, poetry, and the language of religious ceremonies, life cycle traditions, and rituals. The revival of the ancient language of Hebrew in the 19th century and its transformation to a dynamic, living, and lexically rich language was unprecedented. The renaissance of Hebrew is attributed to the vision, unwavering determination, and commitment of Eliezer Ben-Yehuda who is regarded as the driving force and the father of modern spoken Hebrew. Against much opposition, Ben-Yehuda unleashed an era of immense linguistic vitality and creativity by forging a three-pronged language revival plan that included the exclusive use of Hebrew at home, instruction of the language at school, and the coining of new words. He focused on teaching Hebrew to the very young, using immersion as his pedagogical approach, thereby raising a new generation that spoke the language as its mother tongue (Fellman 1973; Schiff 1996; Myhill 2004). He coined hundreds of new words that had not previously existed in Hebrew and compiled words from Hebrew literature of all periods for his publication of the Hebrew dictionary. Ben-Yehuda established the *Va'ad Halashon*, the language committee to uphold and continue the cultivation, expansion, and dissemination of the Hebrew language, often by mining classical and biblical Hebrew for new coinages. In 1953, the committee became the *Academy of the Hebrew Language* that to this day has legislative language planning authority to set standards for grammar, orthography, transliteration, and punctuation and to guide the continuing and accelerated development of the Hebrew language. In 2008, on the 150th anniversary of Ben-Yehuda's birth, UNESCO honored the linguist for his life's work noting his outstanding contribution to education, science, and culture (Mishor & Ordan 2007; UNESCO 2008). Since 2010, the *Hebrew Language Day* has been celebrated annually, on Ben-Yehuda's birth date, to mark the revival of Hebrew, the ancient language, and its continuing development as the unifying language of the Jewish people.

In a quest to create Hebrew alternatives to the lexical diffusion from English and from the multiple languages introduced into the country by immigrants, the Academy of the Hebrew Language continues to coin new words using derivational and linear processes and word-based formation devices (Eldar 2010; Berdichevsky 2014; Ephratt 2014). The Academy regularly invites Israelis to participate in the process while balancing two opposing perspectives on word coinages: the first argues that to strengthen the status and independence of Hebrew, it is imperative to coin Hebrew alternatives for all foreign words; the second asserts that the absorption of words from other languages into Hebrew strengthens the connection of Hebrew speakers to the global culture.

3. The impact of the English language in Israel

The growing status of the English language as an international language, its dominance in digital frontiers and its pervasive use in commerce and travel have also led to its spread in the Israeli linguistic, social, and cultural landscape. English is integrated in Israeli schools in the early grades and proficiency is assessed in tertiary educational institutions. The increasing influx of immigrants from English-speaking countries joined by Israelis returning from residency in these countries who continue to use English in interpersonal communication within their homes and communities further enhance the vitality and impact of the English language in Israel. English-speaking immigrants maintain a sense of dual homeness similar to that of the French-speaking immigrants in Israel (Ben Raphael & Ben Raphael 2010), thereby signaling their linguistic, social, cultural, and affective identity (Poplack 1981; Norton 2013). They stay engaged with their home countries and continue using English at home and in their communities. Various bilingual patterns often described as code switching, the juxtaposition of words and phrases from the two languages within and across sentences, are used in oral communication (Olshtain & Blum-Kulka 1989; McKay & Wong 2000; Olshtain & Horenczyk 2000), translanguaging (Garcia & Wei 2014) and Hebrish, the crossbreeding and creative blending of Hebrew and English within words.

English is prominently featured in the Israeli urban linguistic landscape and is pervasively used in public spaces, on street signs, buses, commercial establishments and arcades, in product names, packaging and advertisements. These have increased dramatically in recent decades underscoring the growing anglocentricity in oral and written communication signaling Israel's globalization endeavor

(Spolsky & Shohami 1999; Burstein-Feldman et al. 2010; Shohamy, Ben Raphael & Barni 2010). Linguistic landscapes have a considerable impact on individuals and communities. To quote Kramsch (2015: 242), "Linguistic landscapes are discourse in action, multimodal discourse …. They interpolate us in different ways and force us to respond with our senses, our memories and our imagination." The dominance of English in scholarly domains, digital platforms, and the media has further expedited its spread among Israelis who, while generally communicate online in Hebrew, prefer to use English when a wider audience and a community of potential readers and followers are of importance. This climate of anglocentricity has indeed led some to view the dominance of Hebrew in Israel as challenged and even as endangered (Kor 2019).

4. Israelis in the United States: Dual allegiance to Hebrew and English

Pre-immigration experiences shape Israeli immigrants' identity and allegiance to their sociocultural and linguistic heritage as well as their propensity for acquiring proficiency in English and the prospects for their successful integration in the United States. Being an immigrant, learning a new language, and adapting to a new sociocultural and linguistic environment are not new experiences for Jews and Israelis. Throughout history, Jews, dispersed across the globe, have assimilated, acquired the local languages, while blending their Jewish identity within the local culture of the countries of their residence. In a study of over one hundred Israeli families residing in urban cities in the United States, 42% of the fathers and 32% of the mothers had emigrated to Israel at a young age from 23 different countries in Africa, Asia, Europe, North and South America and had successfully assimilated into Israeli society prior to immigrating with their own families to the United States (Kaufman 2000). The revival of the Hebrew language in Israel has played a central role in unifying its immigrants and it has remained integral to the sociolinguistic and cultural identity of Israeli adults moving to the United States.

The Israeli community is characterized by its high ethnolinguistic vitality, strong social networks and firm ties to Israel and the Hebrew language. The community in the United States is comprised of individuals and families relocating as representatives of educational institutions, businesses and other organizations for a pre-determined period and those who arrive seeking economic or educational opportunities and whose residence may be extended indefinitely. Members of both groups intermingle and form a single cohesive community. Their immigration to the United States is voluntary, they are well educated having earned at least a high school diploma or a college degree, and are relatively proficient in English. These characteristics greatly enhance their employment opportunities and successful integration in the United States. Shared sociocultural contexts with American society reduce the need for a major readjustment, thereby facilitating their integration in their new setting. Establishment of local ethnic community organizations and centers has not been a concern for the Israeli immigrant community, nor has affiliation with the local Jewish community been a priority, due to the immigrants' secular orientation and the lack of Hebrew language proficiency in the local Jewish communities (Shokeid 1989; Gold & Phillips 1996; Schiff 1997; Gold 2002; Gedzelman 2011; Berdichevsky 2014). Those among the Israeli immigrants, for whom religion and Jewish identity are important, have been more likely to send their children to Jewish schools to help maintain their cultural heritage although, in many of these schools, oral communication in Hebrew is minimal. To strengthen and support the Hebrew language among the younger generation, beyond the home, the community established supplementary schools with an academic curriculum that is aligned with that of schools in Israel. A visit to these schools, however, demonstrates that outside the classroom, the children's preferred language for communication with their Hebrew-speaking school peers is English.

Lacking demographic density in their neighborhoods, Israeli immigrants' prospects for communicating in Hebrew with peers outside the home, in school, and in their local community are minimal, but opportunities are available with new arrivals from the homeland and during visits to Israel. Digital platforms and social media further expand interpersonal connections and offer opportunities to remain connected with family and friends and to remain current on events in Israel. However, for children and adolescents in this community whose social and emotional affinity with the homeland is weakening, engagement in social media focuses on English-speaking peers, thereby further hastening their assimilation and successful integration in the United States.

Despite the challenges, the Hebrew language remains central in the life of Israeli adults in the United States. Survey data from parents in this Israeli immigrant population demonstrate that the children's sociolinguistic environment at home is saturated with multiple opportunities for exposure to and communication in Hebrew. The parents are all first-generation immigrants to the United States. They are fluent speakers of Hebrew and are strongly committed to maintaining the Hebrew language among their children. Literacy practices are widespread in the home and play a critical role in the linguistic and sociocultural development of the children (Kaufman 2000). Shared book reading is an important agent of socialization that scaffolds young children's acquisition of linguistic and literary conventions and narrative schema (Appleby 1978; Berman & Slobin 1994; Whitehurst & Lonigan 1998). Books in Hebrew are widely available, both in print and digitally, and daily interactive shared reading practices and storytelling are introduced early in Israeli immigrants' homes, thereby immersing the children in the Hebrew language, traditions, and culture. These literacy practices play a vital role in enhancing the children's receptive and productive skills in the Hebrew language and in their socialization into the Israeli culture and heritage. They further enhance the young children's familiarity and practice with the Hebrew script and lexicon and expose them to the more formal and grammatically complex written variety of the Hebrew language (Berman & Ravid 2000; Ravid & Tolchinsky 2002; Aram & Korat 2010).

5. Maintaining the heritage language: A challenging endeavor

Home is where deep attachment to the heritage is nurtured and connection with the language is cultivated. Yet developing and maintaining the heritage language among the younger generation has been highly challenging for communities worldwide (Brown 2011; Hinton 2013; Kaufman 2013; Liang 2018). Israeli families are no exception. Exposure to Hebrew is dramatically reduced, it is available predominantly in the home, and is often limited to routines and informal conversational Hebrew. In contrast, exposure to English increases significantly and is available in the media, school, and among peers. The younger generation does not share their parents' affective connection to the Hebrew language, nor do they understand the significance and the symbolism of Hebrew as the language of their heritage. Children lack the motivation to maintain the language and have made successful integration within their peer milieu a priority. Intergenerational conflict within families increases with adolescence and emergent rebellion is not uncommon leading to statements such as these uttered by an 11-year-old: "Why should I speak in Hebrew if you understand English?" or, "It's Mommy not Ima! Speak to me in English in front of my friends". Children and adolescents' use of English in domains outside the home, and increasingly in the home, confirms their weakening Hebrew proficiency and their resolve to assimilate. The challenge of lexical access in Hebrew for interpersonal communication is described by an Israeli girl who has been in the United States since the age of seven: "The Hebrew words are in my brain but the English ones come out of my mouth". Declining proficiency in the heritage language and emerging gaps in the children's lexical access increase their reluctance to communicate in Hebrew, further expediting its attrition. Simplification in Hebrew oral discourse and narration and intra-sentential code mixing of English words serve to fill lexical gaps in the heritage language and are integral to the trajectory of language shift and attrition (Kaufman 2001; Verhoevenn & Strömqvist 2001).

In the home domain, the parents' insistence on speaking the heritage language for interpersonal communication within the family often conflicts with the children's need to convey their message. The 5-year-old child, who has been living in the United States for two and half years, says: *Why can't I first tell you this in English and then say it in Hebrew?* It is indeed not unusual for parents to speak Hebrew with their children while the children communicate with them and with their siblings in English. This unidirectional, unreciprocal communication pattern (Gal 1979) has been identified in language contact situations as a key variable leading to language shift (Haugen 1969; Silva-Corvalan 1994; Zentella 1997). Unreciprocal communication practice plays a dual antithetical role in the language acquisition and attrition process. While supporting receptive maintenance of the heritage language in the home, the primary and often the only domain where Hebrew is still dominant, it expedites attrition of productive communication skills. However, unreciprocal communication also sustains the children's exposure to the heritage language and guarantees continued receptive development. Yet, it also signals parental acknowledgement and acceptance, albeit reluctantly, of the children's increased proficiency in English and the shift in their linguistic allegiance away from the heritage language.

6. Attrition of the heritage language: Focus on Hebrew verb formation

Children's productivity and rich repertoire of lexical innovations have been documented for young monolingual children in both Hebrew and English (Clark & Berman 1984, 1987; Berman 1985; Berman 2000; Berman 2003; Ravid & Schiff 2006; Chomsky 2009). From an early age, children are attuned to the typological differences and properties of the languages they are exposed to and the language-specific patterns for coining new words. When two languages, Hebrew and English, are at play, the young children deploy their knowledge of the relevant language-specific structures and devices for their word formation and lexical innovations. English word formation involves linear affixation and zero derivation. In Hebrew, a Semitic language, construction of verbs and most nouns is characterized by the combination of consonantal root stems with an array of affixal patterns (*Binyan)* templates (Berman 2003; Shimron 2003; Shatil 2016). This process has also been viewed as a word-based affixation process (Aronoff 1994). Children acquiring Hebrew begin at an early age to analyze unfamiliar words in terms of their consonantal root and pattern structure. As their repertoire grows, they show increased sensitivity to the word formation patterns peculiar to Hebrew and coin novel verbs, nouns, and adjectives using available structural devices, including linear and compound formation options (Berman & Ravid 2000; Berman 2003).

Monolingual Hebrew-speaking children as young as three years old acquire the nominal and verbal paradigms and the morphological and inflectional operations for lexical coinages and are able to coin novel verbs from morphologically related nouns. However, in the attrition context, children's Hebrew word formation strategies have been shown to be particularly vulnerable (Kaufman & Aronoff 1989, 1991; Kaufman 1995; Kaufman 2004). As attrition of Hebrew progresses in an English dominant context, the simultaneous interaction of word formation devices from the children's two languages results in innovative blended lexical coinages that combine the templatic morphology of Hebrew with the concatenative affixal morphology of English. In this context, the children's coinages provide a fertile ground for analyzing morphological processes in acquisition and attrition and a window into emerging changes in their knowledge of Hebrew as the impact and dominance of English increases. The children's ability to access word roots and derivational patterns for their verb formation provides evidence for their language development, or the extent of its attrition.

The trajectory of attrition in Hebrew verb formation among young children in the Israeli immigrant community is examined from three different perspectives: children's verb production in oral narration; children's innovative denominal verb formation; and a longitudinal two-year study of a young child. The data from these studies focus on the children's verb formation and are discussed in the sections below. Combined, these shed light on the trajectory of attrition in verb formation in the heritage language of Hebrew. All the children, in this study, even the ones who were born in the United States, had acquired, prior to entering English-dominant daycare or schools, the nominal and verbal paradigms and the morphological and inflectional operations that monolingual Hebrew speakers acquire at a young age. However, their acquisition of Hebrew in the immigration context is still incomplete but continues to develop, albeit with reduced input, primarily restricted to the home domain. In contrast, the children's acquisition of English is dramatically enhanced through exposure to rich linguistic input, opportunities for immersion and interpersonal communication with peers and increasingly with older siblings.

7. Oral narrative data

The narrative production data was collected from 30 Israeli immigrant children (15 boys and 15 girls) ages 6;2–13;11 whose residence in the United States ranged from birth and from the age of two to ten years old. In all cases, Hebrew is the language spoken by the parents at home and English is the children's preferred language for communication with parents, siblings, and peers. Narrative production is especially suited for the study of children's verb constructions as the task entails describing a chronology of events and actions. Loss of lexical specificity and difficulty in retrieval in attrition has led to abundant use of compensatory strategies that include lexical substitution, circumlocution, redundancy, and simplification. Pauses, hesitation, false starts, and repairs are also frequently evident in the children's oral narrative data. Idiosyncratic verb forms attested in the data invariably deviate from developmental verb forms in monolingual speakers (Kaufman 2001).

The narrative data demonstrates a systematic shift from the commonly used Hebrew lexically-specified verbs, where lexical information is encoded in the verb, to analytic verb+particle (bi-lexemic) constructions to encode movement, direction, and manner, a strategy that is more common in English. This shift is attributed to the challenge of lexical access, use of simplification as a compensatory strategy for lexical gaps, and the pervasive use of this device in English. The generic verb go+particle is pervasively used in the data. Mono-lexemic *radaf* "chased" and *tipes* "climbed", which are commonly used by young monolingual speakers of Hebrew, are replaced by *halax axrey* "went after" and *halax al* "went up". The least proficient narrators use the verb+particle construction exclusively as is demonstrated in Table 1.

Table 1. Shift in verb production from lexically-specified to verb+particle

Hebrew lexically specified verb 3.S.PAST	**Children's verb+particle constructions**
Barax (escaped)	*yaca me* "went out of"
	rac me "runs from"
	holex me "goes from"
Xipes (searched)	*mistaklim ba* "look (pl.) in"
	**mistakel bishvil* "looks for"
	**mexapes bishvil* "looks.for for"
Radaf (chased)	*holxim la* "go (pl.) to"
	rac axrey "runs after"
	roxvim axrey "rode (pl.) after"
	**holex axar axrey* "go behind after"
Tipes (climbed)	*ala al* "went.up on"
	Halax al "walked on"
Yaca (exited)	*holxim baxuc* "they.walk outside"
Nixnas (entered)	*holex betox* "goes inside"

8. Innovative denominal verb production

The innovative denominal verb formation elicitation task, administered to 27 Israeli immigrant children ages 7;11–11;00, was designed to gain insights into children's knowledge of Hebrew innovative verb formation strategies that are mastered by monolingual children by age three. The innovative denominal verb formation task requires knowledge of Hebrew verb formation devices and therefore preempts production of rote-learned lexical items (Kaufman 1995, 2004; Berman 2003). The study was conducted in Hebrew and the children were given an example translated here as follows: "I am putting these toys in a box (*kufsa*), what am I doing with the toys when I put them in the box?" In this case, target responses involve extracting the consonantal root *k-f-s* and inserting it into an appropriate verbal template, the most productive verbal templates being: *Piel*: *me-kafes* and *Pa'al*: *kofes*.

Analysis of the children's lexical innovations attests to gaps in the verb formation templatic morphology of Hebrew. The data demonstrates that unlike monolingual Hebrew speakers who overwhelmingly use the highly productive verbal templates *Pi'el* and *Pa'al* for their innovative denominal verbs (Berman 2003), children in this study, across age groups, do not exhibit solid knowledge of the process of Semitic verb formation in Hebrew. The required derivational root+template strategy is replaced by an overwhelming use of affixation, primarily suffixation, attesting to varying degrees of attrition in Hebrew. The children's innovative verb formation devices include an array of responses aligned with their level of proficiency. These include, templatic verbal prefix ***me***+*Noun* or *m**a**+Noun;* prefixes juxtaposed with an idiosyncratic template; zero derivation, a strategy commonly used in English, and an array of English nominal and verbal suffixes, or Hebrew nominal suffixes, including plurals, and diminutives that are affixed to the noun to create the innovative verbs.

Table 2 includes examples of innovative verbs produced by two children who are exposed to Hebrew at home. The first child, female, aged 7;0 has been living in the United States for 1 year and speaks English at home and with peers. The second child, male, aged 7;8 has been living in the United States for 6.5 years and speaks both English and Hebrew at home and with peers. For their innovative verb formation, both children use Hebrew nominal suffixation *Noun+im* (M.PL) and *Noun+ot* (F.PL); *Noun+i* (1.GEN) *Noun+a* (3.F.GEN); as their primary strategy, thereby providing not a verb but an alternate form of the noun, with the first child doing so more consistently. The second child uses the prefix *ba-* ("inside"), a bound prefix in Hebrew and various suffixes, including random suffixes *-pa* and *-ya,* which have

Table 2. Attrition trajectory: Innovative verbs forms

Nouns	Innovative Hebrew verb forms *Piel* template	Child (F) 7;0 1 year in U.S innovative verb	Child (M) 7;8 6 years in U.S innovative verb form
sakin "knife"	*me-saken*	*sakin-im*	*sakin-i*
makel "stick"	*me-makel*	*sakin-im*	*makel-a*
panas "flashlight"	*me-panes*	*panas-im*	*panas-a*
sargel "ruler"	*me-sargel*	*sargel-im*	*sargel-im*
mitriya "umbrella"	*me-tayer*	*mitriya-im*	*mitriya-im*
karton "cardboard"	*me-karten*	*karton-im*	*ba-karton-a*
aron "wardrobe"	*me-aren*	*aron-ot*	*b-aron-a*
kufsa "box"	*me-kafes*	*kufs-im*	*ba-kofsa-im*
sal "basket"	*me-sale*	*sal-im*	*sal-im-pa*
argaz "crate"	*me-argez*	*argaz-im*	*argaz-im*
safsal "bench"	*me-safsel*	*safsal-im*	*safsal-ya*
xalon "window"	*me-xalen*	*xalon-ot-im*	*xalon-i*

no structural or semantic function. Suffixation is prevalent in Hebrew inflectional morphology for person, gender, and number agreement in nouns, adjectives, and verbs, but is not used as a verb formation device.

9. Longitudinal data: Emergence of innovative blends and an idiosyncratic verb form

The intergenerational language shift and increasing vulnerability of the verbal system is best evidenced when investigated longitudinally. A longitudinal perspective provides a unique view into the language development process, the crosslinguistic interplay between the two developing languages and the trajectory of attrition of the heritage language. The young monolingual Hebrew speaker's exposure to English began at the age of 2;6 upon the child's arrival in the United States with her parents and two older siblings. At this age, the child's Hebrew language acquisition is still incomplete and continues to develop. In the first two years of the longitudinal study, four distinct stages are discernible in the child's linguistic development and the trajectory of the acquisition and attrition processes in her speech. These are summarized in Table 3.

Table 3. Stages in the attrition process of the verbal system of Hebrew: A longitudinal perspective

Monolingual to bilingual development and attrition	Age	Number of months in the United States
Monolingual Stage	2;6	Arrival in the U.S.
Monolingual developmental Forms; Hebrew-dominant speech and code mixed English words	2;9–3;1	3–7
Hebrew/English Bilingual Stage	3;1–3;2	7–8
Emergence of English verb blending in Hebrew-dominant speech	3;2–3;5	8–11
Idiosyncratic Verbal Forms	3;6–4;6	12–24

Within seven months after arrival in the United States, the child exhibits bilingual proficiency and developmental lexical innovations in both Hebrew and English aligned with those documented for monolingual language development for speakers of Hebrew and English respectively at this age. As background, unreciprocal communication patterns emerge at age 3;2 when the child's use of English in the home increases while the parents and older siblings continue to use Hebrew exclusively for communication. The fourth stage (3;5–4;6) involves the reconstruction process, whereby the variety of Hebrew verbal patterns give way to the development of a single idiosyncratic verbal template to which English suffixes are increasingly affixed and which is code-mixed in English dominant speech. It is noteworthy that the child does not select one of the existing verbal templates in Hebrew, as might be predicted by theories of regression and simplification, but creates an idiosyncratic verbal template that, while resembling existing derivational Hebrew verbal forms, it is unprecedented in monolingual Hebrew language acquisition data. This template attests to the child's own idiosyncratic formulation of derivational verb formation which she combines with English affixation – a synthesis of the two languages in the child's environment. Instances of innovative lexical and morphological intra-word code blending are also evident in the child's speech data. These do not exist in the input and are the child's own creative forms. Typological differences between Hebrew and English trigger a myriad of innovations providing insight into the processes of acquisition and attrition respectively. The interlingual innovative lexical formation strategies that are highly pervasive at age 3;6–4;6 gradually disappear as the child gradually ceases to speak Hebrew. At the end of the fourth stage, the child's increasing unwillingness to speak Hebrew, further expedites the language shift. Examples of the child's speech at each of the stages of linguistic development are illustrated in Table 4.

Table 4. Stages of Hebrew verb formation strategies in an English dominant environment

Stages in Hebrew verb formation	**The child's verb formation and speech sample**
Monolingual Stage prior to exposure to English. Hebrew is used exclusively for interpersonal communication Age 2;6	*Hi lo yexol-a* ***li-tfos*** *et ha-kadur shel-a* She no can.PRES-FS to-catch OM the-ball of-her "She cannot catch her ball" *lo, kodem* ***ne-saxek*** *axar-kax* ***na-shir****, tov?* no, first 1PL.FUT-play then 1PL.FUT-sing, good/OK? "no, first let's play and then let's sing, O.K.?" *at* ***te-sapr-i*** *…..axshav* ***ta-dlik-i*** you 2.FUT-tell.story-FS …..now 2.FUT-switch.on-FS "you tell the story … now switch on"
Emergence of English verb blending in Hebrew-dominant speech Age: 3;2–3;5	*Im ze it-laxlex ani* ***a*****-kliyn** *ot-am* If this 3S.FUT-get.dirty I 1S.FUT-clean OM-they "If this gets dirty, I'll clean them" *ani* ***bala-ti*** *et ze xazak* I blow.PAST-1S OM this hard (English verb *blow* inflected as a Hebrew verb) "I blew this out (candle) hard" *lo yexol-im le-***step** *al ha-sheleg* "no can-MPL to-step on the-snow"
Verb blending and emergence of the Idiosyncratic Verbal Form in English dominant speech Age: 3;6 – 4;6	I didn't *i-zuz* "I didn't i-move" I'm ***me-nagev*****-ing** myself I want to ***i-nagev*** myself " I'm PRES-dry-ing myself I want to i-dry myself" Can you ***itlabesh*** me? "Can you dress (reflexive) me?" Where is the coat you ***ixabes-ed?*** "Where is the coat you (wash-ed)?"
The Idiosyncratic Verbal Form Example: Root ***k-p-l*** (fold) Age: 3;6 – 4;6	I want to ***ikapel*** one Look I ***ikapel***-ed it nice, right? No, first ***ikapel*** it for me I was just ***ikapel***-ing them I'm ***ikapel***-ing my shirt. That's how I *ikapel* my shirt
The Idiosyncratic Verb is used for all Hebrew verbs code blended in English dominant speech	***Ikarer*** (cool); ***isader*** (arrange); ***isarek*** (comb); ***inagen*** (play); ***ikalef*** (peel); ***isaben*** (soap); ***ishaber*** (break); ***ixamem*** (warm); ***ixatex*** (cut) ***ixasot*** (cover); ***ixabes*** (wash) ***inake*** (clean; ***icalcel*** (ring); ***ixaded*** (sharpen); ***ixalek*** (divide) ***imaxek*** (erase); ***itate*** (sweep); ***ikalkel*** (spoil); ***ikamet*** (crease); ***imale*** (fill)

10. Conclusion

The trajectory of heritage language attrition is characterized in families across immigrant communities by increasing challenges in the intergenerational transmission of the language. For Israeli immigrants, this issue is especially sensitive in light of the historically significant revival of Hebrew, the parents' strong linguistic and affective ties to the heritage language and their unwavering efforts for nurturing and maintaining the Hebrew language and Israeli culture within their families. The pervasive impact of the English language prior to immigration and its dominance in the new context, combined with the sociocultural and linguistic forces that drive the children to assimilate and successfully integrate in the United States, all play a critical role in the children's attrition of Hebrew and the intergenerational language shift.

From a linguistic perspective, as the children's developing Hebrew interacts with the increasingly dominant English, we gain important insights into the morphological interaction between the two typologically different languages. Children are highly resourceful in drawing upon linguistic resources for their lexical innovations. Their robust creativity and hybrid lexical innovations in their developing languages highlight the interplay between the templatic morphology and affixation of Hebrew and the concatenative morphology of English. These typological differences lend transparency to the reciprocal influence of the languages and their respective permeable domains.

The verb formation data demonstrate that manipulation of available linguistic resources is highly systematic across children and the verb formation strategies and devices they select are closely aligned with their decreasing level of proficiency in Hebrew. The vulnerability of Hebrew verb formation, as attrition progresses, is attested in three contexts: first, children's oral narrative production data demonstrate pervasive use of lexical simplification devices to replace action verbs in their narration. Second, the innovative denominal verb production data are replete with an array of nouns and nominal affixes that are produced by the children to represent innovative denominal verbs. Third, longitudinal data demonstrate developmental verbal forms leading to the emergence of a single idiosyncratic verbal template that is used by the child for all the Hebrew verbs in oral communication. Further research and additional longitudinal studies will confirm whether the developmental stages and the verb formation strategies identified are idiosyncratic or whether they may be systematic in children's bilingual development of Hebrew in similar contexts. Such studies will contribute to further illuminating the trajectory of verb formation in the attrition of Hebrew as a heritage language.

References

Ammerlaan, Tom, Madelein Hulsen, Heleen Strating, & Kutlay Yagmur (eds.). 2001. *Sociolinguistic and psycholinguistic perspectives on language maintenance and loss of minority languages*. New York: Waxmann.

Appleby, Arthur. 1978. *The child's concept of story*. Chicago: The University of Chicago Press.

Aram, Dorit & Ofra Korat (eds.). 2010. *Literacy development and enhancement across orthographies and cultures*. New York: Springer. https://doi.org/10.1007/978-1-4419-0834-6

Aronoff, Mark. 1994. Morphology by itself: Stems and inflectional classes. *Linguistic Inquiry* Monograph Series, 22. Cambridge, MA: MIT Press.

Ben Raphael, Eliezer & Miriam Ben Raphael. 2010. Diaspora and returning diaspora: French-Hebrew and vice versa. In Elana Shohamy, Eliezer Ben Raphael & Monica Barni (eds.), *Linguistic landscape in the city*. Bristol: Multilingual Matters. https://doi.org/10.21832/9781847692993-020

Berdichevsky, Norman. 2014. *Modern Hebrew: The past and future of a revitalized language*. Jefferson, NC: McFarland & Company.

Berman, Ruth A. 1985. The acquisition of Hebrew. In Dan I. Slobin (ed.), *The crosslinguistic study of language acquisition*, 255–371. Hillsdale, NJ: Lawrence Erlbaum Associates.

Berman, Ruth A. 2000. Children's innovative verbs versus nouns: Structured elicitations and spontaneous coinages. In Liss Menn & Nan Bernstein-Ratner (eds.), *Methods for studying language production*, 69–93. Mahwah, NJ: Erlbaum.

Berman, Ruth A. 2003. Children's lexical innovations: Developmental perspectives on Hebrew verb structure. In Joseph Shimron (ed.), *Language processing and acquisition in languages of Semitic, root-based morphology*, 243–291. Amsterdam: John Benjamins. https://doi.org/10.1075/lald.28.13ber

Berman, Ruth A. & Elite Olshtain. 1983. Features of first language transfer in second language attrition. *Applied Linguistics*, 4 (3). 222–234. https://doi.org/10.1093/applin/4.3.222

Berman, Ruth A. & Dorit Ravid. 2000. Acquisition of Israeli Hebrew and Palestinian Arabic: A review of current research. *Hebrew Studies* 41. 83–98. www.jstor.org/stable/27913476 (June 6, 2020).

Berman, Ruth A. & Dan Slobin. 1994. *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Brown, Clara L. 2011. Maintaining heritage language: Perspective of Korean parents. *Multicultural Education* 19(1). 31–37. https://files.eric.ed.gov/fulltext/EJ986889.pdf (June 6, 2020).

Broeder, Peter & Guus Extra. 1999. *Language, ethnicity and education: Case studies of immigrant minority groups and immigrant minority languages*. Clevedon: Multilingual Matters.

Burling, Robbins. 1978. Language development of a Garo and English speaking child. In Evelyn Hatch (ed.). *Second language acquisition*, 54–75. Rowley, MA: Newbury House.

Burstein-Feldman, Zhanna, Alek D. Epstein, Nina Kheimets, Shulamith Kopeliovich, Dafna Yitzhaki & Joel Walters. 2010. Israeli sociolinguistics: From Hebrew hegemony to plurilingualism. In Martin J. Ball (ed.), *The Routledge handbook of sociolinguistics around the world*, 226–237. New York: Routledge. https://doi.org/10.13140/2.1.2757.9205

Celce-Murcia, Marianne. 1978. The simultaneous acquisition of English and French in a two-year-old Child. In Evelyn M. Hatch (ed.), *Second language acquisition*, 38–53. Rowley, MA: Newbury House.

Chomsky, Carol. 2009. Creativity and innovation in child language. *Journal of Education* 189(3). 37–42. https://doi.org/10.1177/002205740918900306

Clark, Eve & Ruth A. Berman. 1984. Structure and use in the acquisition of word formation. *Language* 60(3). 542–590. https://doi.org/10.2307/413991

Clark, Eve & Ruth A. Berman. 1987. Types of linguistic knowledge: Interpreting and producing compound nouns. *Journal of Child Language* 14 (3). 547–567. https://doi.org/10.1017/S030500090001028X

De Bot, Kees and Robert Schrauf (eds.). 2009. *Language development over the lifespan*. New York: Routledge.

De Bot, Kees & Saskia Stoessel. 2000. In search of yesterday's words: Reactivating a long forgotten language. *Applied Linguistics* 21 (3). 364–384. https://doi.org/10.1093/applin/21.3.333

Eldar, Ilan. 2010. *Language Planning in Israel*. Jerusalem: The Academy of the Hebrew Language.

Ephratt, Michal. 2014. A model is born: Presenting the derivation in modern Hebrew. *Hebrew Studies*, 55. 121–170. National Association of Professors of Hebrew. http://www.jstor.com/stable/43151470 (June 18, 2020).

Extra, Guus & Ludo Verhoeven (eds.). 1993. *Immigrant languages in Europe*. Clevedon: Multilingual Matters.

Fase, Willem, Jaspaert Koen & Sjaak Kroon (eds.). 1992. *Maintenance and loss of minority languages*. Philadelphia: John Benjamins. https://doi.org/10.1075/sibil.1

Fein, Richard J. 1996. What can Yiddish mean to an American poet? *Shofar*, 14 (4). 79–84. http://www.jstor.org/stable/42942509 (June 3, 2020).

Fellman, Jack. 1973. *The revival of a classical tongue. Eliezer Ben Yehoda and the Modern Hebrew language*. Reprint 2011. The Hague: Mouton. https://doi.org/10.1515/9783110879100

Ferguson, Charles A. & Shirley Brice Heath (eds). 1981. *Language in the USA*. Cambridge: Cambridge University Press.

Fishman, Joshua. 1989. *Language and ethnicity in minority sociolinguistic perspective*. Clevedon: Multilingual Matters.

Flores, Cristina. 2014. Losing a language in childhood: A longitudinal case study on language attrition. *Journal of Child Language* 42(3). 562–590. https://doi.org/10.1017/S0305000914000233

Gal, Susan. 1979. *Language shift: Social determinants of linguistic change in bilingual Austria*. New York: Academic Press.

Garcia, Ofelia & Joshua A. Fishman (eds.). 1997. *The multilingual apple: Languages in New York city*. Amsterdam: Mouton de Gruyter.

Garcia, Ofeilia & Li Wei. 2014. *Translanguaging: Language, bilingualism and education*. New York: Palgrave Macmillan. https://doi.org/10.1057/9781137385765

Gedzelman, David. 2011. The possibilities and potential of Hebrew in America. *Contact* 13 (2). 3–5. New York: Steinhardt Foundation for Jewish Life.

Gold, Steven J. 2002. *The Israeli diaspora*. Seattle: University of Washington Press.

Gold, Steven J. & Bruce A. Phillips. 1996. *Israelis in the U.S. American Jewish Yearbook*. New York: American Jewish Committee.

Hansen, Lynne (ed.). 1999. *Second language attrition in Japanese contexts*. Oxford: Oxford University Press.

Hardach, Sophia. 2018. *Can you lose your native language?* https://www.bbc.com/future/article/20180606-can-you-lose-your-native-language (June 18, 2020).

Haugen, Einar E. 1969. *The Norwegian language in America*. Bloomington: Indiana University Press.

Hinton, Leanne (ed.). 2013. *Bringing our languages home: Language revitalization for families*. Berkeley, CA: Heyday Books.

Hoffman, Eva. 1989. *Lost in translation: Life in a new language*. New York: Dutton.

Kaufman, Dorit. 1995. Where have all the verbs gone: Autonomy and interaction in attrition. *Southwest Journal of Linguistics*, Special Issue Language Loss and Public Policy, 14 (1–2). 43–66.

Kaufman, Dorit. 2000. Attrition of Hebrew in the United States: Sociolinguistic perspectives. In Elite Olshtain & Gabriel Horenczyk (eds.), *Immigration, identity and language*, 173–196. Jerusalem: The Hebrew University Magnes Press.

Kaufman, Dorit. 2001. Tales of L1 attrition: Evidence from pre-puberty children. In Tom Ammerlaan, Madelein Hulsen, Helen Strating & Kutlay Yagmur (eds.), *Sociolinguistic and psycholinguistic perspectives on maintenance and loss of minority languages*, 185–202. Münster, Germany: Waxmann.

Kaufman, Dorit. 2004. Acquisition, attrition, and revitalization of Hebrew in immigrant children. In Dorit Diskin Ravid & Hava Bat-Zeev Shyldkrot (eds.), *Perspectives on language and language development*, 407–418. New York: Springer.

Kaufman, Dorit. 2013. Modern Hebrew in the diaspora: The language of Israeli immigrants in the United States. In Geoffrey Kahn (ed.), *Encyclopedia of Hebrew language and linguistics (EHLL)*. Leiden, The Netherlands: Brill.

Kaufman, Dorit & Mark Aronoff. 1989. Morphological interaction between L1 and L2 in language attrition. In Susan Gass, Carolyn Madden, Dennis Preston & Larry Selinker (eds.), *Variations in second language acquisition: Psycholinguistic issues*, 202–215. Avon, England: Multilingual Matters.

Kaufman, Dorit & Mark Aronoff. 1991. Morphological disintegration and reconstruction in first language attrition. In Herbert Seliger & Robert Vago (eds.), 175–188. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511620720.012

Köpke, Barbara & Dobrinka Genevska-Hanke. 2018. First language attrition or dominance: Same or different? https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01963/full (June 9, 2020)

Kor, Avshalom. 2019. The magic of Hebrew. Lecture delivered at the *Dov Lautman conference on Living together: Language shapes reality*, December 17, 2019. Raanana, Israel.

Kramsch, Claire. 2015. A researcher's auto-socio-analysis: Making space for the personal. In Bernard Spolsky, Ofra Inbar-Lourie & Michal Tannenbaum (eds.), *Challenges for language education and policy: Making space for people*, 235–244. New York: Routledge.

Lam, Amy. 2019. *English language glossary for a daughter of refugees: How to find the right words in another tongue*. https://gay.medium.com/english-language-glossary-for-a-daughter-of-refugees-1600177ecdcc (June 4, 2020).

Lambert, Richard D. & Barbara F. Freed (eds.). 1982. *The loss of language skills*. Rowley, MA: Newbury House Publishers.

Liang, Feng. 2018. Parental perceptions toward and practices of heritage language maintenance: Focusing on the United States and Canada. *International Journal of Language Studies* 12(2). 65–86. https://files.eric.ed.gov/fulltext/ED581470.pdf (June 8, 2020)

McKay, Sandra Lee & Sau-ling Wong. 2000. *New immigrants in the United States*. New York: Newbury House.

Mishor, Mordechay & Dena D. Ordan. 2007. *Eliezer Ben-Yehuda (1858–1922) Reviver of spoken Hebrew*. Jerusalem: Academy of the Hebrew Language. https://mevakrim.hebrew-academy.org.il/wp-content/uploads/sites/2/2014/08/BY-UNESCO.pdf (June 8, 2020).

Myhill, John. 2004. *Language in Jewish society*. Clevedon: Multilingual Matters. https://doi.org/10.21832/9781853597626

Norton, Bonny. 2013. *Identity and language learning: Extending the conversation*. 2nd edn. Bristol: Multilingual Matters. https://doi.org/10.21832/9781783090563

Olshtain, Elite & Shoshana Blum-Kulka. 1989. Happy Hebrish: Mixing and switching in American-Israeli family interaction. In Susan Gass, Carolyn Madden & Larry Selinker (eds.), Variation in second language acquisition: Discourse and pragmatics, 59–83. Clevedon, England: Multilingual Matters.

Olshtain, Elite & Gabriel Horenczyk (eds.). 2000. *Language, identity and immigration*. Jerusalem, Israel: Magnes Publications.

Polinsky, Maria. 2018. *Heritage languages and their speakers*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781107252349

Poplack, Shana. 1981. Sometimes I'll start a sentence in Spanish y termino en Español: Toward a typology of code-switching. *Linguistics* 18. 581–618.

Ravid, Dorit & Liliana Tolchinsky. 2002. Developing linguistic literacy: A comprehensive model. *Journal of Child Language* 29. 417–447. https://doi.org/10.1017/S0305000902005111

Ravid, Dorit & Rachel Schiff. 2006. Roots and patterns in Hebrew language development: Evidence from written morphological analogies. *Reading Writing* 19. 789–818. https://doi.org/10.1007/s11145-006-9004-3 (June 8, 2020)

Romaine, Suzanne (ed.). 1991. *Language in Australia*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511620881

Saville-Troike, Muriel, Junlin Pan & Ludmila Dutkova. 1995. Differential Effects of L2 on children's L1 Development/Attrition. *Southwest Journal of Linguistics*. Special Issue Language Loss and Public Policy, 14, (1–2), 125–149.

Schiff, Alvin. 1997. Hebrew in New York. In Ofelia Garcia & Joshua Fishman (eds.), *The multilingual apple: Languages in New York City*, 203–227. Amsterdam: Mouton de Gruyter.

Schiff, Alvin. 1996. *The mystic of Hebrew: An ancient language in the new land*. New York: Sheingold Publications.

Schmid, Monika S. 2011. *Language attrition*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511852046

Schmid, Monika S. & Barbara Köpke. 2017. The relevance of first language attrition to theories of bilingual development. *Linguistic Approaches to Bilingualism*. 7(6). 637–667. Amsterdam: John Benjamins. https://doi.org/10.1075/lab.17058.sch

Seliger, Herbert W. & Robert M. Vago (eds.). 1991. *First language attrition*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511620720

Shatil, Nimrod. 2016. Hebrew nominal patterns in the speaker's mind. *Hebrew Studies* 67. 239–255. https://doi.org/10.1353/hbr.2016.0012

Shimron, Joseph (ed.). 2003. *Language processing and acquisition in languages of Semitic, root-based morphology*. Amsterdam: John Benjamins. https://doi.org/10.1075/lald.28

Shohamy, Elana, Eliezer Ben Raphael & Monica Barni (ed.). 2010. *Linguistic landscape in the city*. Bristol: Multilingual Matters. https://doi.org/10.21832/9781847692993

Shokeid, Moshe. 1989. *Children of circumstances: Israeli emigrants in New York*. Ithaca, New York: Cornell University Press.

Silva-Corvalan, Carmen. 1994. *Language contact and change: Spanish in Los Angeles*. Oxford: Clarendon Press.

Spolsky, Bernard & Elana Shohamy. 1999. *The languages of Israel: Policy, ideology and practice*. Clevedon: Multilingual Matters.

UNESCO 2008. UNESCO Heritage Project. *150th anniversary of the birth of Eliezer Ben-Yehuda, linguist* (1858–1922). http://portal.unesco.org/en/ev.php-URL_ID=42392&URL_DO=DO_TOPIC&URL_SECTION=201.html (June 8, 2020).

Verhoeven, Ludo & Sven Strömqvist (eds). 2001. *Narrative development in a bilingual context*. Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.23

Whitehurst, Grover J. & Christopher J. Lonigan. 1998. Child development and emergent literacy. *Child Development* 69. 848–872. https://doi.org/10.1111/j.1467-8624.1998.tb06247.x

Yukawa, Emiko. 1997. *L1 Japanese attrition and regaining: Three case studies of early bilingual children*. Stockholm: Center for research on bilingualism, Stockholm University.

Zentella, Ana Celia. 1997. *Growing up bilingual*. Malden, MA: Blackwell.

CHAPTER 22

A primer for linguists on the reading wars

Edwin Battistella
Southern Oregon University

As linguists and teachers, we may be asked our opinions on contentious issues of reading instruction such as phonics versus whole language. This chapter reviews debates about reading instruction in the United States, introducing the relevant concepts, issues, research, and personalities to linguists. The chapter begins with a look at pre-nineteenth century views on reading and nineteenth and early twentieth century practices. The mid-twentieth century publication of *Why Johnny Can't Read* led to renewed interest in phonics and in linguistics-based reading materials. At the same time, another linguistically-inspired approach, whole language, was being proposed. Today, we have a politically fraught landscape of what are often described as 'reading wars.' The chapter ends with a set of recommendations for navigating this landscape.

Keywords: reading, phonics, whole language, whole word, linguistics and education, literacy

1. Teaching reading in the nineteenth century (and before)

The phonic method of teaching reading in English, or simply *phonics*, is often attributed to the sixteenth-century spelling reformer and educator John Hart. Hart proposed an augmented Roman alphabet for the spelling of English sounds and, in 1570, published a textbook on reading. His idea, later developed by William Bullokar, Charls [*sic*] Butler, Alexander Ellis, and Isaac Pittman (of shorthand fame), was to make the relationship between letters and sounds more transparent and transcription-like. The greater the transparency, the thinking went, the easier it would be for novice readers to decode text.

Phonic alphabets, while often quite ingenious, did not effectively address the problem of reading conventional English spelling. Actual classroom practice tended to involve spelling taught as a prerequisite to reading, and in nineteenth-century American schoolhouses, Noah Webster's *American Spelling Book* was often used,

https://doi.org/10.1075/cilt.353.22bat

even with its many idiosyncrasies. Over time, the most popular reading textbooks became the McGuffey readers, first published in 1836 by William H. McGuffey, then a professor at Miami University in Ohio. His eventual six levels of readers offered short stories, poems, prayers, speeches, and moral tales intended to be sounded out and read aloud. The approach was implicitly phonic, but as an 1879 edition notes, "'This First Reader may be used in teaching reading by any of the methods in common use; but it is especially adapted to the Phonic Method, the Word Method, or a combination of the two." (McGuffey 1879: ii).

In the nineteenth century, there was in fact quite a spirited discussion of approaches based on whole words as opposed to sounds and letters. Horace Mann, who became secretary of the newly established Massachusetts Board of Education in 1837, was unhappy with the alphabet as the basis for reading pedagogy. Under his leadership, the Boston primary schools adopted a new primer designed by Thomas Gallaudet and based on methods used for deaf education, where sounding out words was not an option. Mann believed that since the relationship between letters and sounds was so irregular, "learning the letters first is an absolute hindrance" to learning to read (Mann 1844: 91). Instead, he proposed that "Whole words should be taught before teaching the letters of which they are composed" (Mann 1841: 13). Proponents of whole-word reading believed that being able to recognize common words by sight – so-called sight words – allowed readers to focus on comprehension without having to stop and decode each word as they read.

Mann claimed that the experience of the Boston schools supported the whole-word method. However, by 1844, Mann's school board colleagues rebelled. Their report on "Modes of Teaching Children to Read" (Association of Masters 1844) characterized the whole-word method, disapprovingly, as "converting our language to Chinese" and offered a series of objections including inefficiency, confusion, defective pronunciation, and idleness and pleasure seeking (1844: 94–95). But the whole-word method retained its advocates, among them Colonel Francis Parker, who adopted whole-word pedagogy as a superintendent in Quincy, Massachusetts, Boston, and Chicago.

2. See Dick run

Education reformer John Dewey would swing the pendulum further toward the whole-word method. Dewey was greatly influenced by Colonel Parker, and the progressive education movement of which Dewey was the intellectual leader also incorporated insights from Edmund Burke Huey's 1908 *The Psychology and Pedagogy of Reading*. Huey argued that children were not developmentally prepared for print, suggesting that it was "a great waste to devote, as at present, the main part of a number of school years to the mere mechanics of reading and spelling" (Huey 1908: 301). Instead, Huey proposed that children should be acquiring experiences and interests "that will in time make reading a natural demand, and meaningful process …." (1908: 303). Other educational psychologists of the time conducting experiments in word recognition concluded that words were first recognized as wholes, and thus reading books of the time used the approach of associating words with pictures and systematically repeating the words so that learners would remember them by sight.

By the 1930s, the Dick and Jane readers would arrive to cement the sight-word method in pedagogy. These books, popular from the 1940s through the early 1960s, were associated with William S. Gray and Zerna Sharp. Gray had a Ph.D. in education from the University of Chicago, served as dean of its College of Education, and co-authored the Basic Reading Series and the Elson-Gray Readers (Elson & Gray 1930) before becoming the senior editor of the Scott, Foresman reading program. Zerna Sharp had spent a decade teaching first grade, had served as principal of an elementary school, and was hired in 1924 as reading consultant for Scott Foresman. She proposed the new series and supervised the Dick and Jane stories, which were illustrated by a series of artists over the years.

As for Dick and Jane, their adventures revolved around a small set of repeated words at each grade level. One of Dick and Jane's first appearances was in a story called "The Old, Old Doll," in which Dick pulls up in his red wagon while Jane is sitting on the steps playing with her dolls. It opens this way:

> Dick said, "Come, Jane.
> We will go for a ride.
> Will you go?"
> Jane said, "Yes, Dick.
> I want a ride.
> My dolls want a ride.
> We will go."

In the two pages that follow, Jane puts her new doll and her old doll (named Polly) in the wagon and the four of them go down the hill. The story ends with Polly falling out of the wagon.

Fun with Dick and Jane, published in 1940, included five stories – "Look Up," "Who is It?," "Something Pretty," "Where is Sally?," "Jane Helps Mother," and "A Funny Ride." The opening tale begins with a picture of Dick pointing upwards at a hat floating above the trees while Jane looks on. The next page has them running to investigate the floating hat and, on the third page, Dick and Jane see that their father is carrying young Sally on his shoulders and she is wearing his hat. The story concludes:

> "Look, look," said Dick.
> "See Sally.
> See funny Sally and Father."
> "See, see," said Sally.
> "Sally is up, up, up.
> This is fun for Sally."

It was perhaps less fun for young readers. The key to Dick and Jane, and their various imitators, was repetition of a set of prescribed sight words, introduced in the context of illustration, along with the depiction of a type of middle-class fantasy life in Anytown, USA. First graders were exposed to about 300 words, the number expanding at each grade level.

Attention and public perception began to shift in 1954, with the publication in *Life* magazine of an article titled "Why Do Students Bog Down on the First R?" Written by John Hersey, the author of *Hiroshima*, the article criticized the content of the Dick-and-Jane style readers – "pallid primers" as he called them. The following year another writer would weigh in with a book-length condemnation of the whole-word method.

3. Rudolph Flesch and Doctor Seuss: The Fish and the cat

Rudolph Flesch was an Austrian lawyer who fled the Nazis in 1938, finding his way to New York City. There he completed a Ph.D. in Library Science at Columbia University, studying the factors that made a text easy or difficult to read. His 1943 Ph.D. dissertation, *Marks of a Readable Style*, provided a mathematical formula to predict the difficulty of adult reading material. The formula included such variables as the number of names and personal pronouns, which enhanced readability, and the number of prefixes and suffixes, which did the opposite. Successors of Flesch's formula are still in use today.

In 1955, Flesch extended his study of reading in a different direction, with a sharp critique of the whole-word method. He wrote the bestselling *Why Johnny Can't Read, and What You Can Do About it*. Its eleven chapters began with "A Letter to Johnny's Mother" and ended with "A Letter to Johnny's Teacher." In between, he offered anecdotes based on visits to schools and conversations with parents, explanations of phonic and whole-word methods, and critiques and endorsements of various basic readers. Flesch quoted from education professors, whom he characterized as "complacent" (1955: 95–96). The whole-word method was "totally wrong" (1955: 2) because it required children to memorize sight words and encouraged learners to guess or skip over words they could not recognize. Flesch even threw in the complaint that the whole word method was like teaching Chinese characters (1955: 3–8), echoing Horace Mann's critics. As for the Dick and Jane books, Flesh called them "horrible, stupid, emasculated, pointless, tasteless little readers" (1955: 6).

The year after *Why Johnny Can't Read* was published, the International Reading Association was established. The first president was William S. Gray, Dick and Jane's godfather. In *The Reading Teacher* journal, Gray had responded with a short article titled "Phonics Versus Other Methods of Teaching Reading," defending the whole-word method as promoting "the development of a thoughtful reading attitude" as opposed to mere word-calling (Gray 1955). But he conceded a role for phonics as well, proposing that "Effective progress results from parallel emphasis upon both meaning and word recognition." Gray's article was followed immediately in the journal by another article called "An Analysis of Propaganda Techniques Used in *Why Johnny Can't Read*" (Lamkin 1955).

While Flesch did not succeed in winning over educators with his polemical approach, he did catch the attention of William Spaulding, the Houghton Mifflin publishing executive who had commissioned Theodore Geisel to write *The Cat in the Hat*. Spaulding had given Geisel a list of just over two-hundred words and the instruction to write "a story that first-graders can't put down." Geisel succeeded. Published in 1957, the book met with rave reviews (Geisel was called "the moppets' Milton," among other accolades) and, priced at just under two dollars, *The Cat in the Hat* sold three million copies in its first three years (Morgan & Morgan, 1995: 154–156).

Geisel, a canny entrepreneur, wrote to a friend before *The Cat in the Hat* came out that he expected the book to "make a tremendous noise in the discussion of Why Johnny Can't Read" (Nel 2007: 9). And in a 1981 interview in *Arizona* magazine, he implied that the word list Spaulding had provided was actually a sight-word list:

> That damned Cat in the Hat took nine months until I was satisfied. I did it for a textbook house and they sent me a word list. That was due to the Dewey revolt in the Twenties, in which they threw out phonic reading and went to word recognition, as if you're reading a Chinese pictograph instead of blending sounds of different letters. I think killing phonics was one of the greatest causes of illiteracy in the country. Anyway, they had it all worked out that a healthy child at the age of four can learn so many words in a week and that's all. So there were two hundred and twenty-three words to use in this book. I read the list three times and I almost went out of my head. I said, "I'll read it once more and if I can find two words that rhyme that'll be the title of my book." (That's genius at work.) I found 'cat' and 'hat' and I said, 'The title will be The Cat in the Hat.' (Carlinsky, 1981: 51)

As for Flesch, who played the fish to Geisel's cat, his work on readability found renewed interest in the wake of the 1970s Plain Language Movement, which gave him a platform to revisit his critique in a 1981 book called *Why Johnny Still Can't Read: A New Look at the Scandal of Our Schools.* Flesch replied to his critics and even claimed that education publishers and professors were engaged in a "great cover up" (Flesch 1981: 40) that concealed the findings of research and offered only token phonics. Academic reviews of *Why Johnny Still Can't Read* largely dismissed Flesch as a single-minded polemicist. Nevertheless, Flesch's two books spurred others to take phonics seriously. And the titular Johnny would go on to spawn dozens of later book titles by other authors.

4. From "Let's Read" to Project Literacy

The interest in phonics in the 1960s brought renewed attention to Leonard Bloomfield's work on reading. Bloomfield had died in 1949 but *Let's Read: A Linguistic Approach*, co-authored with lexicographer Clarence L. Barnhart, appeared in 1961. The book excerpted and expanded Bloomfield's 1942 article on "Linguistics and Reading," which had critiqued earlier phonics and the whole-word approaches. According to Bloomfield, phonics approaches often mistakenly assumed that letters represented sounds uniformly rather than contextually. As for the whole-word method, Bloomfield saw it as crucially ignoring the alphabet principle – the idea that letters represent sounds:

> The word method proceeds as though our writing were word writing. Every word has to be learned as an arbitrary unit … In order to read a new word, the child must learn the new word character; he can best do this by memorizing the letters which make up this new word character, but these letters are arbitrarily presented and have nothing to do with the sound of the word. (1961: 28–9)

Bloomfield's pedagogical suggestions involved teaching phonics explicitly by distinguishing between regular and irregular spellings and by introducing each letter with just one phonetic value at first. Irregular spellings were classified according to the types of deviation from the alphabetic principle, and silent letters, double letters, and digraphs were to be avoided along with the problematic letters *x* and *q*. The approach was unabashedly behavioristic and stressed the need for constant repetition toward the goal of distinguishing phonemes. It also abstracted away from concerns about meaning and even included nonsense words alongside actual words in the lessons.

Bloomfield was not the only structural linguist developing reading pedagogy. Michigan's Charles Carpenter Fries published *Linguistics and Reading* in 1963. Reading, in his view, was "process of transfer from the auditory signs for language signals, which the child has already learned, to the new visual signs for the same signals" (1963: 120). In other words, learning to read involved decoding orthography into phonemic patterns. For Fries, decoding was a matter of pattern recognition. Learners thus needed to first identify contrasting patterns of letters (such as G versus C or F versus T), then identify sequenced patterns of about 500 one-syllable words that were connected to the learners' experience grammatically and semantically. The initial phonics sequence proposed by Fries was as follows:

a. Same final consonant letter: *at, bat, cat, fat, gat, hat, mat*
b. Same initial consonant letter: *bat, bad, bag, ban*
c. Different vowel letter: *bad, bid, bed, bud*
d. Digraph in place of initial or final consonant (*than, bath*)
e. Consonant cluster of two phonemes in place of initial or final consonant (*step, best*)
f. Consonant clusters of two or more phonemes (*clamp, string, helps*)
g. Final consonant doubled (*back, gaff, pass, ball, bell*)

Only complete words were used, and the learner was to infer the pronunciation-spelling patterns from contrasts systematically introduced. Fries explained that "We avoid completely such a question as 'What does the letter C say?'" (1963: 204). The approach became the basis for the Merrill Linguistic Reading Series "based on the philosophy of Charles Fries" (Fries et al. 1966).

Bloomfield and Fries defended the idea that reading involved the skill of decoding orthography and they each proposed materials for explicit, systematic phonics instruction. During the 1960s and early 1970s, other linguists also weighed in on the reading process, some under the auspices of a program called Project Literacy. Developed in the early 1960s by Harry Levin and Joanne R. Mitchell at Cornell University and funded by the US Office of Education, the multiyear project aimed at bringing together scholars in a coordinated, interdisciplinary research initiative. A 1970 collection related to Project Literacy, edited by Levin and Williams, included contributions by Noam Chomsky (1970, on the ways in which English orthography might be optimal at an abstract morphophonemic level), Richard Venezky (on predictable and unpredictable spellings), and William Labov (on vernacular culture versus school culture).

5. Whole language

Work in Chomskyan generative linguistics was also the impetus for what has become known as the whole-language approach to reading (the term is often capitalized). Whole language views learning to read as analogous to learning to speak and best fostered by relatively unstructured – and often phonics-free – immersion in simple printed texts and in reading aloud. A key difference with phonics approaches involves dealing with unfamiliar words, which could be skipped, guessed at, or inferred from context in whole language. In phonics approaches, unfamiliar words would be sounded out, in what was derided as "grunt-and-groan" methods.

Whole language emerged from the work of Kenneth Goodman and Frank Smith. Goodman's short article "Reading: A psycholinguistic guessing game" (1967), proposed that readers used rapid, unconscious hypothesis testing to identify words. He was referring to skilled readers rather than beginners, but he extended the focus on hypothesis-formation ("word guessing" as he called it) to beginning readers as well. Goodman based his proposal in part on a deep structure-surface structure "model of sentence production" he attributed to one of Chomsky's presentations at Project Literacy (1967: 130).

The whole-language approach reflected a philosophical shift in the teaching of reading that aimed at creating an environment that encouraged children to develop their reading skills. Words are recognized by various cuing systems – orthographic, semantic, syntactic, and pragmatic, with the emphasis on cues relevant to meaning. Other components of whole-language pedagogy include a natural learning environment that values literacy, authenticity of reading materials, and empowerment of readers, especially those not exposed to much reading in their homes (see Goodman 1982). And while whole language is not a prescribed sequence of activities, it is often associated with the program called Reading Recovery, a successful intervention program developed by clinical psychologist Marie Clay.

The whole-language philosophy does not necessarily exclude the direct teaching of decoding skills, but some whole-language advocates dispute the value of phonics instruction. One is Frank Smith, whose 1971 *Understanding Reading: A Psycholinguistic Analysis of Reading and Learning to Read*, incorporated cognitive psychology and early generative linguistics. Smith devoted a chapter of his *Understanding Reading* to critiquing phonics, concluding that it was inadequate and time-consuming strategy due to the complexity of English orthography and the limitations of short-term memory. In the most recent edition of *Understanding Reading*, he characterizes phonics this way:

> Before most children come to school, well-intentioned adults say to them "That word is *John*," or "That word is *cereal*," just as on other occasions they say, "That animal is a cat," in all cases leaving it to the child to solve the more complex problem of working out exactly how to recognize the word or animal on future occasions. But when the children get to school this support is frequently taken away from them, at least as far as reading is concerned. Another well-intentioned adult is likely to say to them, "Good news and bad news today, children. The bad news is that no one is ever likely to tell you what a word is again. The good news is that we are going to give you 166 rules and 45 exceptions so that you can work it out for yourselves.
> (2004: 109)

Smith's characterization of phonics as 166 rules and 45 exceptions is a rhetorical exaggeration of actual phonics practice. More measured concerns have to do with the effect of pre-school reading readiness on skills development, with time management in a classroom, with the consistency and accuracy of phonics generalizations, with the potential for the pedagogy to become mere drill, and with the context-dependent nature of some phonics rules. Critics of whole language in turn argue that explicit recognition of the alphabetic principle and the early development of phonemic awareness are necessary skills for comprehension and that learning to read is not a process parallel to the acquisition of language.

6. Who's right?: Research and meta-studies

A sharper picture of the debate over phonics and whole language comes from the various research studies and meta-studies on reading and reading pedagogy. In *Why Johnny Can't Read*, Flesch cited about a dozen studies; Gray's reply cited a pair of research summaries: his own "Preliminary Survey of Methods of Teaching Reading and Writing" (1953) and Witty and Sizemore's "Phonics in the Reading Program: A Review and Evaluation" (1955), arguing that the results were inconclusive. Jeanne Chall's *Learning to Read: The Great Debate* (1967) was a landmark review of the research on teaching reading from 1910 to 1965 focusing on actual methods, available experimental research, and interviews. Chall concluded that "at the end of grade 1 or by grade 2, the stronger phonics programs produce better results in both word recognition and comprehension" (Chall 1989: 8).

Researchers Guy Bond and Robert Dykstra led a federally-funded study conducted in hundreds of first grade classrooms. The goal was to understand methods of beginning reading programs, but also the characteristics of communities, schools, and students. Known as the First Grade Studies, their research also supported the need for phonics, but found that emphasis on comprehension combined with phonics was most effective. As they put it, "Evidently, reading achievement is influenced by factors peculiar to school systems over and above differences in pre-reading capabilities" (Bond & Dykstra 1967: 121–122).

The studies of the 1960s were influential but failed to establish a professional consensus, particularly as whole language grew in popularity through the last two decades of the twentieth century. Another study by Marilyn Jager Adams concluded that "programs including systematic instruction on letter-to-sounds correspondences lead to higher achievement in both word recognition and spelling, at least in the early grades and especially for slower or economically disadvantaged learners" who may come from homes not rich in reading opportunities (1990: 31). A review conducted by the National Research Council eschewed specific recommendations about pedagogy but also noted that beginning reading depended crucially on decoding skills (Snow, Burns & Griffin 1998: 321).

Public and political concerns remained, and a fourteen-member National Reading Panel (NRP) was formed by the National Institute of Child Health and Human Development in 1997. Its 450-page report, *Teaching Children to Read*, was a meta-analysis of peer-reviewed research on phonemic awareness, phonics, fluency, comprehension strategies, vocabulary development, technology, and teacher preparation. The phonics section of the report concluded that

> growth in word-reading skills is strongly enhanced by systematic phonics instruction when compared to non-phonics instruction for kindergartners and 1st graders as well as for older struggling readers. Growth in reading comprehension is also boosted by systematic phonics instruction for younger students and reading disabled students. These findings should dispel any belief that teaching phonics systematically to young children interferes with their ability to read and comprehend text. Quite the opposite is the case. (National Reading Panel 2000: 2–94)

While the report was broadly accepted and influential, it was also criticized as selective, and one member issued a minority report (see Yatvin 2005). *Teaching Children to Read* became the basis for parts of the No Child Left Behind program, under which states were required to set annual targets for improving achievement.

Most recently, there is the report of the National Early Literacy Panel, titled *Developing Early Literacy*. Published in 2008, *Developing Early Literacy* was a meta-analysis of approximately 300 studies showing which early measures correlate with later literacy achievement. The discussion of what the authors called "code-focused interventions" (they avoided the term "phonics") concluded that such interventions "have a significant, substantial, and positive impact both on young children's conventional literacy skills and on early skills that predict later literacy achievement" (National Early Literacy Panel 2008: 109). The National Early Literacy Panel characterized this as the difference between scoring at the 50th and 79th percentiles.

Meta-studies, from Chall and the First Grade Studies to Adams and Snow et al. to *Teaching Children to Read* and *Developing Early Literacy*, point to the value of explicit phonics in early reading instruction. And while teachers do not generally review meta-studies before planning their lessons, they seem to value phonics as well. According to both the International Reading Association (IRA) and the National Reading Panel Report, the majority of teachers support phonics instruction. The IRA policy statement of 1997 cited a study finding that "98% of primary-grade teachers regard phonics instruction as a very important part of their reading program." The NRP report added that 89% believed that skills instruction should be combined with literature and language-rich activities. Currently many teachers (and schools districts) opt for a middle ground known as balanced literacy, which includes phonics and elements of whole-language methods.

Beyond the meta-studies and surveys of teachers, there is also a body of more recent linguistically interesting research on reading. Rayner et al. (2001) review much of the underlying learning theory connected with reading development and reading deficits along with laboratory studies of eye movement, word identification, reading comprehension, brain imaging, and computer simulations of reading. Rayner et al. (2001: 57) note that two of the basic tenets of whole language have been refuted: Goodman's idea that skilled reading is a psycholinguistic guessing game and the view that learning to read is a natural act parallel to learning to speak. Nevertheless, they endorse the whole-language principles of engaging learners with literature and of focusing on the learners rather than "the teachers as an agent of instruction" (2001: 57). As for phonics, they find convincing evidence that, while both direct unmediated access of words and phonologically-mediated access play a part in reading, decoding words as phonemes is a key aspect of learning to read.

7. Politics as usual, and conspiracies too

Reading has always been a contentious subject, connected to philosophies about learning, about the nature of language, and about the nature of learners. But since *Why Johnny Can't Read*, the debates have become political and cultural clashes. In the Cold War era, reading and education were a matter of national security. Over time, phonics has become a rallying point for right-wing critiques of learner-centered progressive education in various screeds. One example is the writing of Samuel L. Blumenfeld, whose publications include *The Whole Language/OBE Fraud* as well as guides to phonics and homeschooling. In *The Whole Language/OBE Fraud*, Blumenfeld attacks John Dewey as subverting literacy "[b]ecause high literacy produced that abominable form of independent intelligence which was basically, as Dewey believed, anti-social" (1996: 55). In Blumenfeld's view, Dewey crafted "a master plan, involving the entire progressive education community, to create a new socialist curriculum for the schools of America, a plan, based on the new psychology, that was indeed carried out and implemented" (56).

Ideological critiques of phonics echo similar dark themes: In *The Linguistics, Neurology and Politics of Phonics*, Steven L. Strauss (2005: 26–27) writes that phonics fits the agenda of corporate America, which "wants public schools to manufacture a workforce of information technology (IT) workers." Strauss adds that phonics "is ideally adaptable to the pedagogy that is required for imposing an authoritarian, top-down, externally defined 'standards' curriculum on classrooms." Brinkley & Weaver (2005) look at religious fundamentalists' preference for phonics, noting that fundamentalists of various faiths are literalists who "pay close attention to the reading of text and often insist that their children memorize, word for word, significant passages of the Bible, the Torah, or the Koran" (98). Later they add that "A steady diet of literal and uncritical reading … leaves readers of any age vulnerable to those who have an agenda for which they want to enlist followers" (100). And Matthews (2005) critiques the financial connections between phonics researchers/advocates and corporate publishers.

The ideological divide is exacerbated by harsh characterizations of proponents of phonics or whole language. Teachers using whole-language methods are dismissed as taking the easy approach of merely reading to students; colleges of education are attacked for willfully ignoring science (see Hanford 2018). Academic proponents of phonics are dismissed as intellectual imperialists (see Strauss 2005). Supporters are characterized as "phonics nuts" advocating mindless drills that promote word-calling without meaning-making and who lack practical classroom knowledge (Lemann 1997). And teacher organizations often push back against challenges to their autonomy and professionalism. The International Reading Association, for example, has adopted policy statements that express its concern with "exaggerated claims found in the press and other media" and the "growth in the number of curricular and legislative mandates that require teachers to blindly follow highly prescriptive plans for phonics instruction" (IRA 1997).

By the 1990s, phonics versus whole language has become a wedge political issue at the level of state education policy, especially in states whose size gives them disproportionate influence in the textbook industry. But as Flesch himself noted in *Why Johnny Can't Read*, the issue of best practices for teaching reading should be distinct from one's politics. In his closing chapter (1955: 127), he explained that he was "not a reactionary but a liberal" with a great admiration for John Dewey. He remarked:

> Who says a progressive, liberal minded teacher must not tell her pupils anything about sounds and letters … Why is the word method always labeled modern and phonics branded as reactionary? There is no earthly reason for pigeonholing them this way.

No reason other than politics.

8. How to read the reading wars: Beyond name calling

Phonics has been discussed since the sixteenth century and hotly debated from the nineteenth century on, competing at first with the whole-word approach and later with whole language. Among the controversies, we find issues of the nature of reading, the role of the alphabetic principle, the complexity of English orthography, the development of the learner, and much more. For the linguist who is not a psychologist or reading specialist, this chapter is intended as guide to the history, issues, pitfalls, subtleties and not-so-subtleties in the phonics-whole language debates.

When entering into discussions of reading in the classroom or the community, it is worth remembering that term 'reading' includes activities as diverse as decoding unfamiliar words in beginning books and constructing meaning in literary and disciplinary contexts. The terms 'phonics' and 'whole language' are likewise used in a variety of ways by researchers, practitioners, and the public and are prone to uncharitable misconstrual. 'Phonics' may be misunderstood as simply drilling the alphabet and the pronunciation of letters and syllables, where students sound out words without understanding them. 'Whole language' may be misconstrued as simply reading to students and hoping they learn on their own. In discussions of phonics and whole language with colleagues and students, it may be useful to begin with the question "What do you mean by that?"

With respect to early reading and reading development, the crucial question for the whole-language philosophy is whether reading is acquired naturalistically like spoken language or learned by developing skills that recode letters as phonemes. Experimental evidence and classroom studies support the view that reading is learned, not acquired by immersion, and that success in reading requires phonemic awareness and decoding skills. These skills are supported by early, explicit phonics instruction.

Objections that phonics does not work because the English spelling is too arbitrary miss the point. While all phonics programs are not the same, the complexity of English orthography can be simplified and scaffolded in accessible and useful ways which build the phonemic awareness that underlies decoding and that support meaning-making and comprehension in reading.

Finally, we should recognize that reading instruction can and should use insights and materials that build a broad, critical literacy and a love of reading and that support all groups of learners equally. In actual teaching practice, phonics versus whole language becomes a theoretical dichotomy, useful for polemicists but not educators. Good teachers blend elements of each while keeping in mind that the research consensus strongly supports the early use of explicit phonics instruction in reading pedagogy.

Acknowledgements

The author is indebted to Margaret Perrow, Marcia Haag, and Janie Rees-Miller for helpful comments and suggestions.

References

Adams, Marilyn Jager. 1990. *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Association of Masters of the Boston Public Schools. 1844. *Remarks on the seventh annual report of the Hon. Horace Mann, Secretary of the Massachusetts Board of Education. Association of Masters of the Boston Public Schools*. Boston: C. C. Little and J. Brown.

Bloomfield, Leonard. 1942. Linguistics and reading. *The Elementary English Review* 19(4). 125–130 & 19(5). 183–186.

Bloomfield, Leonard & Clarence L. Barnhart. 1961. *Let's read: A linguistic approach*. Detroit, MI: Wayne State University Press.

Blumenfeld, Samuel L. 1996. *The whole language/OBE fraud: The shocking story of how America is being dumbed down by its own education system*. Boise, ID: Paradigm Publishing.

Bond, Guy & Robert Dykstra. 1967. The cooperative research program in first-grade reading instruction. *Reading Research Quarterly* 2(4). 5–142. https://doi.org/10.2307/746948

Brinkley, Ellen & Constance Weaver. 2005. Phonics, literalism and futuristic fiction: Religious fundamentalism and education policy. In Leslie Poynor & Paula Wolfe (eds.), *Marketing fear in America's public schools: The real war on literacy*, 93–110. Mahwah, NJ: Lawrence Erlbaum.

Carlinsky, Dan. 1981. A story (it's true) of a Geisel with a view. *Arizona* 14(June). 48–53.

Chall, Jeanne. 1967. *Learning to read: The great debate*. New York: McGraw-Hill.

Chall, Jeanne. 1989. *The role of phonics in the teaching of reading: A position paper prepared for the Secretary of Education*. Washington, DC: Office of Educational Research and Improvement. ERIC ED328899.

Chomsky, Noam. 1970. Phonology and reading. In Levin & Williams, 3–18.

Elson, William H. & William S. Gray. 1930. *Elson-Gray basic reader pre-primer*. The Philippine Islands, Scott Foresman & Co.

Flesch, Rudolph. 1943. *Marks of readable style; a study in adult education*. New York: Teachers College Press.

Flesch, Rudolph. 1955. *Why Johnny can't read, and what you can do about it*, New York: Harper & Row.
Flesch, Rudolph. 1981. *Why Johnny still can't read: A new look at the scandal of our schools*. New York: Harper & Row.
Fries, Charles. C. 1963. *Linguistics and reading*. New York: Holt, Rinehart & Winston.
Fries, Charles. C., Agnes C. Fries, Rosemary G. Wilson & Mildred K. Rudolph. 1966. *Merrill linguistic reading series*. Columbus, OH: Charles E. Merrill.
Geisel, Theodore. 1957. *The cat in the hat*. New York: Houghton Mifflin.
Goodman, Kenneth. 1967. Reading: A psycholinguistic guessing game. *Journal of the Reading Specialist* 6(4). 126–135. https://doi.org/10.1080/19388076709556976
Goodman, Kenneth. 1982. *Language, literacy and learning*. London: Routledge & Kegan Paul.
Gray, William. S. 1953. *Preliminary survey of methods of teaching reading and writing*. Paris: UNESCO.
Gray, William S. 1955. Phonics versus other methods of teaching reading *The Reading Teacher* 9(2). 102–106.
Gray, William. S., May Hill Arbuthnot, Eleanor Campbell & Keith Ward. 1940. *Fun with Dick and Jane*. Chicago: Scott, Foresman & Co.
Hanford, Emily. 2018. Hard words: Why aren't kids being taught to read? *APMReports*. Online: https://www.apmreports.org/story/2018/09/10/hard-words-why-american-kids-arent-being-taught-to-read. (12 December, 2019.)
Hart, John. 1570. *A methode or comfortable beginning for all vnlearned, whereby they may be taught to read English: In a very short time, vvith pleasure: So profitable as straunge, put in light*, London: I. H. Chester Heralt.
Hersey, John. 1954. Why do students bog down on first R? *Life* 36(21). 136–150.
Huey, Edmund. B. 1908. *The psychology and pedagogy of reading: with a review of the history of reading and writing and of methods, texts, and hygiene in reading*. New York: Macmillan.
International Reading Association. 1997. *The role of phonics in reading instruction*. Newark, Delaware: International Reading Association.
Labov, William. 1970. The reading of the *–ed* suffix. In Levin & Williams, 222–245.
Lamkin, F. Duane. 1955. An analysis of propaganda techniques used in "Why Johnny Can't Read". *The Reading Teacher* 9(2), 107–117.
Lemann, Nicholas. 1997. The reading wars. *Atlantic Monthly* 280(5). 128–133.
Levin, Harry & Joanna R. Williams (eds.). 1970. *Basic studies on reading*, New York: Basic Books.
Mann, Horace. 1841. *Lecture on the mode of preparing and using spelling books*. Boston: William D. Ticknor.
Mann, Horace. 1844. Seventh annual report of the Secretary of the Board. *Massachusetts Board of Education. Seventh annual report of the Board of Education*. Boston: Dutton & Wentworth.
Matthews, Sharon. 2005. Who gets to play? How and why reading researchers were left out of the No Child Left Behind Act. In Leslie Poynor & Paula Wolfe (eds.), *Marketing fear in America's public schools: The real war on literacy*, 51–61. Mahwah, NJ: Lawrence Erlbaum.
McGuffey, William. H. 1879. *First eclectic reader* (revised edition). Cincinnati: Van Antwerp, Bragg & Co.
Morgan, Judith & Neil Morgan. 1995. *Dr. Seuss & Mr. Geisel: A biography*. New York: Random House.
National Early Literacy Panel. 2008. *Developing early literacy: Report of the National Early Literacy Panel*. Washington, DC: National Institute for Literacy.

National Reading Panel. 2000. *Report of the National Reading Panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. Washington, D.C.: National Institute of Child Health and Human Development, National Institutes of Health.

Nel, Philip. 2007. *The annotated cat: Under the hats of Seuss and his cats*. New York: Random House.

Rayner, Keith, Barbara R. Foorman, Charles A. Perfetti, David Pesetsky & Mark S. Seidenberg. 2001. How psychological science informs the teaching of reading. *Psychological Science in the Public Interest* 2(2). 31–74. https://doi.org/10.1111/1529-1006.00004

Smith, Frank. 1971. *Understanding reading: A psycholinguistic analysis of reading and learning to read*. New York: Holt, Rinehart & Winston.

Smith, Frank. 2004. *Understanding reading: A psycholinguistic analysis of reading and learning to read*. New York: Routledge.

Snow, Catherine E., M. Susan Burns & Peg Griffin (eds.). 1998. *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Strauss, Steven L. 2005. *The linguistics, neurology and politics of phonics*. Mahwah, NJ: Lawrence Erlbaum. https://doi.org/10.4324/9781410612014

Venezky, Richard. 1970. Regularity in reading and spelling. In Levin & Williams, 30–42.

Witty, Paul A. & Robert A. Sizemore. 1955. Phonics in the reading program: A review and an evaluation. *Elementary English* 32(6). 355–371

Yatvin, Joanne. 2005. Making whole language disappear: How the National Reading Panel worked its magic. In Leslie Poynor & Paula Wolfe (eds.), *Marketing fear in America's public schools: The real war on literacy*, 81–91. Mahwah, NJ: Lawrence Erlbaum.

Index of languages

Index of names

Index of terms

This book provides a view of where the field of morphology has been and where it is today within a particular theoretical framework, gathering up new and representative work in morphology by both eminent and emerging scholars, and touching on a very wide range of topics, approaches, and theoretical points of view. These seemingly disparate articles have a common touchstone in their focus on a word-based, paradigmatic approach to morphology.

The chapters in this book elaborate on these basic themes, from the further exploration of paradigms, to studies involving words, stems, and affixes, to examinations of competition, inheritance, and defaults, to investigations of morphomes, to ways that morphology interacts with other parts of the language from phonology to sociolinguistics and applied linguistics.

The editors and contributors dedicate this volume to Prof. Mark Aronoff for his profound influence on the field.

ISBN 978 90 272 5963 9

9 789027 259639

JOHN BENJAMINS PUBLISHING COMPANY