

DE GRUYTER
MOUTON

COMPLEX ADPOSITIONS IN EUROPEAN LANGUAGES

A MICRO-TYPOLOGICAL APPROACH TO COMPLEX
NOMINAL RELATORS

*Edited by Benjamin Fagard, José Pinto de Lima,
Dejan Stosic, Elena Smirnova*

EMPIRICAL APPROACHES
TO LANGUAGE TYPOLOGY

EBSCO Publishing : eBook Collection (EBSCOhost) - printed on 2/9/2023 9:38 PM

via
AN: 135174 ; Benjamin Fagard, Jos Pinto de Lima, Dejan Stosic, Elena
Smirnova ; Complex Adpositions in European Languages : A Micro-Typological
Approach to Complex Nominal Relators
Accessions335141



Complex Adpositions in European Languages

Empirical Approaches to Language Typology



Editors

Georg Bossong

Bernard Comrie

Kristine Hildebrandt

Jean-Christophe Verstraete

Volume 65

Complex Adpositions in European Languages

A Micro-Typological Approach to Complex Nominal
Relators

Edited by
Benjamin Fagard
José Pinto de Lima
Dejan Stosic
Elena Smirnova

DE GRUYTER
MOUTON

ISBN 978-3-11-067693-8

e-ISBN (PDF) 978-3-11-068664-7

e-ISBN (EPUB) 978-3-11-068679-1

Library of Congress Control Number: 2020946661

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at <http://dnb.dnb.de>.

© 2020 Walter de Gruyter GmbH, Berlin/Boston

Typesetting: Integra Software Services Pvt. Ltd.

Printing and Binding: CPI books, GmbH, Leck

www.degruyter.com

Contents

Benjamin Fagard, José Pinto de Lima, Dejan Stosic, and Elena Smirnova

Introduction: Complex adpositions and complex nominal relators — 1

Indo-European language families & isolates

Benjamin Fagard, Dejan Stosic, and José Pinto de Lima

1 Complex adpositions in Romance: Emergence and variation — 33

Anatol Stefanowitsch, Elena Smirnova, and Matthias Hüning

2 Complex adpositions in three West Germanic Languages: German, Dutch, and English — 65

Dejan Stosic

3 Complex adpositions in Slavic languages: The case of Serbian — 139

Kevin J. Rottet

4 Complex prepositions in Breton — 195

Pietro Bortone

5 On complex adpositions in Modern Greek — 233

Guillaume Bonnet and Benjamin Fagard

6 Complex prepositions in Albanian: A first assessment — 265

Non Indo-European languages

Anni Jürine

7 What makes a complex postposition: The cyclic development of postpositions in Estonian — 299

Maria Usacheva and Timofey Arkhangel'skiy

8 Complex and simple postpositions: the split in Permic — 335

Gerd Jendraschek

9 Case marking and complex adpositions in Basque — 367

VI — Contents

Gerd Jendraschek

10 Structure and origin of Turkish adpositions — 403

Thomas Stolz

11 A chapter on compound prepositions in Maltese: PREP-PREP combinations and related issues — 439

Discussion

Benjamin Fagard, Elena Smirnova, Dejan Stosic, and José Pinto de Lima

12 Complex adpositions in Europe and beyond: A synthesis — 473

Index — 497

Language Index — 501

Benjamin Fagard, José Pinto de Lima, Dejan Stosic, and Elena Smirnova

Introduction: Complex adpositions and complex nominal relators

In this introductory chapter, we address the issue of the existence and definition of complex adpositions – and complex nominal relators in general –, and provide a working definition on the basis of the existing literature on the subject. The existence of simple adpositions as a word class is now consensual, despite a long and convoluted history, which has seen them equated among other things with conjunctions, adverbs and prefixes (Aurous 1989–2000), for obvious reasons of functional proximity. They are generally considered to be monomorphemic relational expressions whose main function in language is to introduce nouns or noun phrases. Research on simple adpositions has been quite wide-ranging, from a theoretical perspective (e.g. Lindqvist 1994, Kurzon & Adler 2008, Libert 2013), and their typological variation has been given some attention already (Svorou 1994, Hagège 2010, Fagard 2010), with a wealth of studies on adpositions in individual languages such that we could not list them all (see e.g. Vaguer 2006 for a non-exhaustive list of 1,535 references on simple adpositions).

The status of complex adpositions (henceforth CAs), conversely, is not entirely consensual. For instance, Seppänen et al. conclude their analysis of CAs in English as follows:

Introduced into the grammar on the basis of an untenable analysis, the class of complex prepositions as defined by Quirk et al. is empty, and the term itself is thus not helpful in the description of English. (Seppänen et al. 1994: 25)

At any rate, CAs have received much less attention than simple adpositions. There are very few book-length studies on the subject: to our knowledge, the only counter-examples are studies on their development in Spanish (Cifuentes Honrubia 2003), English (Hoffmann 2005), Russian (Šiganova 2001) and Esto-

Benjamin Fagard, Lattice (CNRS, ENS & Paris Sorbonne Nouvelle; PSL), Montrouge, France.
benjamin.fagard@ens.psl.eu

José Pinto de Lima, Universidade de Lisboa, Faculdade de Letras, Portugal. pintodelima@outlook.pt
Dejan Stosic, University of Toulouse Jean Jaurès & Laboratory CLLE (CNRS & UT2), Toulouse, France. dejan.stosic@univ-tlse2.fr

Elena Smirnova, Institut de langue et littérature allemandes, Université de Neuchâtel, Neuchâtel. elena.smirnova@unine.ch

<https://doi.org/10.1515/9783110686647-001>

nian (Jürine 2016). Aside from that – and from mentions of complex adpositions in grammars (e.g., for English, Quirk *et al.*, 1985) and in studies of adpositions in general (e.g. Di Meola 2000, Fagard 2010) –, most studies on complex adpositions are article length studies, often bearing on a subset of complex adpositions, in a given European language: English (Schwenter & Traugott 1995), Dutch (Hüning 2014, Bree 2014), French (Borillo 1991, 1997, Adler 2007, 2008), Portuguese (Lima 2014, 2019a, 2019b), Romanian (Ciobanu 1957, Găitănaru 1965, 1999, Căpățână 2000, 2003). A few papers are devoted to the study of complex adpositions in a more general (e.g. in Romance languages, Fagard & Mardale 2007) or theoretical perspective (Seppänen *et al.* 1994, Meibauer 1995, Adler 2001); a few others compare simple and complex adpositions (e.g. Fagard & De Mulder 2007, Fagard 2012). Finally, there is a growing body of descriptive papers which provide accounts of the paradigms of complex adpositions in individual languages, e.g. in French (Borillo 1997, Stosic & Fagard 2019), Italian (Piunno & Ganfi 2017), Occitan (Sibille 2019), Romanian (Fagard & Mardale 2019).

It now seems possible to build on these (mostly descriptive) studies on complex adpositions, and provide a global account of CAs in European languages. This is what this volume aims for. In order to move on from these descriptions to a wider typological overview, however, some common grounds seem necessary. In this chapter, we propose first an overview of the main features of simple (Section 1) and complex (Section 2) adpositions, before providing working definitions for both, including a discussion of *adpositions vs relators* (Section 3). We then focus on the main research questions addressed in the chapters of the book (Section 4), before presenting briefly each chapter (Section 5).

1 Simple adpositions: Basic features and description

The many studies on simple adpositions, especially those in the 20th and early 21st century, have made it possible to provide a consensual definition of the class. Before defining them, however, let us first take a look at what, in European languages, is generally called ‘adposition’.

1.1 Pre-, post-, ambi- and circumpositions

Adpositions are a category under which prepositions (1) and postpositions (2) are subsumed, as well as less frequent formations, such as circumpositions (3).

Sometimes, an adposition may occur pre- and postnominally; in such cases, one speaks of ambipositions (4).

- (1) *Questi* *parea* *che* ***contra*** *me* (Italian)
 DEM.M.SG seem.PST.3SG PRO.REL towards PRO.1SG.ACC
venisse
 come.PST.SBJ.3SG
 ‘He seemed to be coming **towards** me’ (Dante, *Inferno*, I, 46)
- (2) *Ik* *ben* *er* ***tegen*** (Dutch)
 PRO.1SG.NOM be.1SG PRO.DEI **against**
 ‘I’m against it’ (Multatuli, *Max Havelaar*, 4)
- (3) ***um*** *Lebens* *oder* *Sterbens* ***willen***, (German)
for life.GEN.SG or death.GEN.SG **sake**
bitt’ *ich* *mir* *ein* *Paar*
 ask.1SG PRO.1SG.NOM PRO.1SG.DAT DET.IND.NOM.SG pair
Zeilen *aus*
 line.ACC.PL out
 ‘For the sake of life or death, I ask for myself a few lines’ (Goethe, *Faust*)
- (4) ***wegen*** *des* *Unwetters* / *des*
because DET.M.GEN.SG storm.GEN.SG DET.M.GEN.SG
Unwetters ***wegen***
 storm.GEN.SG **because**
 ‘because of the storm’¹

1.2 Syntactic and distributional features

Syntactically, adpositions are basically relational expressions: an adposition establishes a grammatical relationship linking two other elements, one of which is a governing word while the other is a governed expression, commonly called “complement”. This results in a tripartite schema that could be tentatively noted [A R B] (where “R” stands for relator, meaning the adposition; cf. Pottier 1962, Cervoni 1991, Cadiot 1997, Hagège 1997, De Mulder & Stosic 2009). As is well known, this relation is an asymmetrical one, since the adposition has a stronger

¹ Most examples are taken either from the literature or the internet; a few are created.

relationship with its complement than with its governor, so that [A [R B]] would actually be a better way of describing it. Saying that the adposition is the head of phrase, called “adpositional phrase”, captures this asymmetrical relation. Thus, one can distinguish “prepositional”, “postpositional” and “circumpositional” phrases depending on where the adposition is placed with respect to its complement (B).

Most often, adpositions take as complements non-propositional expressions, the prototype of which are noun phrases (including pronouns, see 1–4 above), but many other kinds of syntactic structures can be introduced by adpositions, such as adjectival, adverbial, or prepositional phrases (see 5–7):

- (5) *De* [modeste], *il est devenu très hautain.* (French)
 from modest he AUX.PRS.3SG become.PTCP.PST very
 haughty
 ‘From modest, he has become very haughty’

- (6) *Fais- moi ça pour [demain]* (French)
 do.IMP.PRS.2SG me.DAT this for tomorrow
 ‘Do this for me tomorrow’

- (7) *Tu pars de [devant la ligne jaune]* (French)
 You start.PRS.2SG from in.front.of the line yellow
 ‘You start from in front of the yellow line’

Adpositions are also found to introduce other types of complements, namely subordinate clauses. However, one should distinguish between cases where the function of adpositions is the same as in other uses (8–9) and those where their status is more problematic, as in examples (10–11):

- (8) *Ostavi to za [kad budeš imao 7 godina]* (Serbian)
 leave.IMP.2SG this for when AUX.FUT.2SG have.PST.PTCP.M
 seven year.ACC.PL
 ‘Leave this to when you’ll be seven years old’

- (9) *Je suis nostalgique de [quand ils étaient tout petits]*
 I be.1SG nostalgic of when they be.PST.3PL all small.PL
 ‘I miss the time when they were only small kids’

- (10) [**Pendant** *qu' elle lit*], *je regarde la* (French)
 during comp she read.PRS.3SG I watch.PRS.1SG the.F
télé
 TV
 'While she reads, I watch TV'

- (11) *por ver cuán poco tiempo había [para hacer* (Spanish)
 for see.INF how little time be.PST.3SG to do.INF
la experiencia] no quiso decirle otra cosa sino
 the experiment NEG want.PST.3SG tell.INF.CL other thing than
que le agradecía el haberle descubierto tan
 that him.DAT please.PST.3SG the have.INF.CL discover.PST.PTCP so
gran secreto
 great secret
 'Seeing how little time was left **to** launch the experiment he chose not to tell him anything else than that he was thankful for his having unveiled to him such a great secret' (Cervantes, *Don Quijote de la Mancha*, II, 1146)

In some theoretical frameworks (see e.g. Dubinsky & Williams 1995), it is accepted that *pendant* and *para* in (9–10) may introduce (finite and non-finite) subordinate clauses. In more traditional syntactic approaches, however, *pendant* (*que*) and *para* in the examples above are analyzed as subordinators or conjunctions.

1.3 Head & argument relations

In the traditional syntactic view, an adposition is usually attributed the morpho-syntactic property of case assignment; this property motivates the concept of adpositional phrase in the first place. As head of an adpositional phrase, adpositions assign (usually non-nominative) case to the noun phrase which serves as complement of the adposition.

Adpositional phrases may be integrated into the more global structure of the clause in different ways. Thus, Melis (2003: 26–33) speaks of the “integrating” role of prepositions (hence, adpositions in general) and captures mechanisms behind incorporating adpositional phrase [R B] into [A] by distinguishing three possible “integration modes”:

- i) **syntactic-semantic**, typical of sentence adjuncts (12), (18–21); in this case, adpositional phrases can be seen as elements which display relatively strong semantic content and are syntactically relatively independent elements of the clause,

- ii) **lexical**, characteristic of strong lexical selection (13); in this case, the adpositional phrase is semantically and syntactically dependent on some other element of the clause, most often the verb (in terms of valency relations);
- iii) **structural**, when adpositional phrases serve to express grammatical relations and are very strongly integrated into phrases (and hence, into the clause) (14–16).

- (12) *[Selon la météo], les averses violentes perturberont la circulation pendant plusieurs jours* (French)
 according.to the.F.SG weather.forecast the.PL rainfall.PL
 violent.PL disrupt.FUT.3PL the.F.SG traffic during several
 day.PL
 ‘According to the weather forecast, heavy rainfall will disrupt traffic for a few more days’ (Melis 2003: 27)
- (13) *We count on you for this project; all this depends on the outcome of our research* (English)
- (14) *A coitada [da Maria]!* (Portuguese)
 the.F.SG poor.F.SG of.the.F.SG Maria
 ‘Poor Maria!’
- (15) *la ville [de Paris] est la capitale de la France* (French)
 the.F.SG city of Paris be.PRS.3SG the capital of
 the.F.SG France
 ‘The city of Paris is the capital of France’
- (16) *J’ ai le mal [de mer]* (French)
 I have.PRS.1SG the sickness of sea
 ‘I suffer from sea sickness’

The relationships established by adpositions thus range from purely grammatical ones as in (14–16) to others that are less semantically bleached, such as the agentive relationship in (17), and to a whole range of semantic relations. Among the

more frequent are spatial (1, 7), causal (3), temporal (18), instrumental (19–20), purposive (21) relations:

- (17) *Bio je izgrđen [od svih prisutnih]* (Serbian)
 be.PST.PTCP AUX.3SG blamed by all.PL present.PL
 ‘he was blamed **by** all persons present’
- (18) [*Po katastrofie*] *czas żałoby* (Polish)
 after catastrophe.LOC time mourning.GEN.SG
 ‘After the catastrophe has come the time of mourning’ (*Gazeta Wyborcza*,
 March 25th, 2015)
- (19) *Connectez-vous [à l’aide de votre adresse électronique]* (French)
 connect.IMP.PRS.2PL you with your address
électronique]
 e-mail
 ‘Log in with your e-mail address’
- (20) *l’altre al basso avevan fatto* (Italian)
 the others in.the bottom have.PST.3SG do.PST.PTCP
 [*con reti*] *riparo*
 with net.PL trap
 ‘The others, in the valley, had set up a trap **with** nets’ (Boccaccio, *Caccia di Diana*, II, 14th c.)
- (21) *Wir kämpfen [für eine bessere Zukunft]* (German)
 we fight.PRS.1PL for a.F better.F.ACC.SG future
 ‘We are fighting for a better future’

Depending on whether the relationship established is seen as prominently grammatical/structural or lexical/semantic, some scholars distinguish between functional and lexical adpositions (see Gougenheim 1950, 1959, Spang-Hansen 1953, Lehmann 1985, Svorou 1994); it is also possible to distinguish a third class of semi-lexical adpositions (see e.g. Mardale 2011). In short, there seems to be a continuum between the two classes, with many elements in between.

The former, i.e. structural or functional adpositions, generally correspond to a very small set of short, simple, but very frequent (and often highly polysemous) forms (see 22–23), while the latter, i.e. lexical adpositions, are generally taken to include a wider range of both formally simple (24) and complex (25) expressions:

- (22) *il vient de partir / de Londres* (French)
 he come.3SG from leave.INF / from London
 ‘he just left / he comes from London’
- (23) *il donne un livre à relier / à son ami* (French)
 he give.3SG a book to bind / to his friend
 ‘He gives a book to bind / he gives his friend a book’
- (24) *he came after us* (English)
- (25) *she parked in front of the movie theater* (English)

An important subclass of adpositions is that of complex adpositions (CAs), which may be roughly defined as multi-word linguistic expressions which serve typical functions of adpositions, and which can be shown to alternate with them. Complex adpositions pertain to the subclass of lexical adpositions, with relatively rich and specific semantics.

2 Complex adpositions: A descriptive approach

Concerning complex adpositions, there is to our knowledge no equivalent of the existing typological overviews on simple adpositions (e.g. Zelinsky-Wibbelt 1993, Svorou 1994, Hagège 2010). However, there have been a number of studies on individual languages, as noted above. Some mentions in passing are found in books on simple adpositions (e.g. Melis 2003, Kurzon & Adler 2008), and a few books provide an overview of complex adpositions in a given language (Cifuentes Honrubia 2003, Hoffmann 2005, Jürine 2016). A number of papers bear on specific questions related to complex adpositions (Lehmann 1985, Leeman 2007, 2008, Gross 2006, Fagard & De Mulder 2007, Piunno & Ganfi 2017), as well as a recent special issue of the *Revue Romane* (Fagard, Lima & Stosic (eds.) 2019a).

This body of literature broadly confirms expectations, which could be summarized as follows.

- a) CAs are hard to define, because there is a continuum from syntactically free to fixed expressions (see already, on this subject, Borillo 1991, 1997, Seppänen et al. 1994, Melis 2003);
- b) there are doubts as to what exactly happens during the evolution from one to the other, and which has been described as lexicalization, grammaticalization and constructionalization (cf. Hoffmann 2004);

- c) while some languages or language strata display a wealth of constructions, others seem to have very limited paradigms;²
- d) some morpho-syntactic patterns or *construction types* seem more productive and pervasive than others, and many authors actually only look at these more salient constructions, the most notable being [P₁(Det)NP₂] (see e.g. Huddleston 1984: 341–345);
- e) the distributional equivalence between simple and complex adpositions is true only of so-called *lexical* adpositions (Fagard 2012),
- f) CAs seem to have evolved in a mostly parallel fashion across European languages, suggesting the possibility of an *areal phenomenon* (Hüning 2014).

Complex adpositions, which could be described as less grammaticalized members of the overall category of adpositions, function in the language as adpositional units, though they are recognized as evincing some internal complexity, on account of surface structure (they are multi-word expressions, see (26–29)) and internal variation (for instance, in (27), *au mépris de* could be changed to *au plus grand mépris de* ‘to the (lit.) greatest spite of’). It thus appears that complex adpositions are morphemic sequences in which some internal structure can be discerned, but whose elements dispose of reduced autonomy, with respect to both syntagmatic and paradigmatic aspects (cf. Heine 1989). In discourse, these elements often occur together as a unit. Semantically, the meaning of the whole sequence is autonomous and mostly noncompositional.

(26) *Era verdade, o lixo da cheia* (Portuguese)
 be.PST.3SG truth the rubbish of.the.F flood
em vez de ir para a direita ia para
 instead of go.INF towards the.F right go.PST.3SG towards
a esquerda
 the.F left
 ‘It was true, the debris of the flood instead of going right went left’ (Eça de Queirós, *Alves*, 20th c.)

(27) *Menant dès ce soir même, au mépris de* (French)
 take.GER from this evening self at.the spite of

² This could be linked to different choices in *definition*, whence the need for this volume. However, the different chapters seem to confirm this trend, which is also clear in diachrony (see e.g. Fagard, to appear).

ses lois, ses compagnes danser autour
 DET.POSS.PL laws DET.POSS.PL companion.F.PL dance.INF around
 de sa statue
 of DET.POSS.F.3SG statue

‘Taking that same evening, despite his laws, her companions to dance around his statue’ (Jean de La Fontaine, *Fables*, 17th c.)

(28) *the Board has responded to the government with regard to the pensions issue* (BNC corpus) (English)

(29) *Der Bau wurde aufgrund von Streitigkeiten zwischen der Stadt Hamburg und dem Bauträger Hochtief in den vergangenen Jahren immer wieder verschoben* (German)
 the.NOM.M.SG construction AUX.PASS on.the.basis of
 conflicts between the.GEN.F.SG city Hamburg and
 the.DAT.M.SG developer Hochtief in the.DAT.N.PL past
 year.DAT.PL ever again delay.PST.PTCP
 ‘The construction was repeatedly delayed in the past years **on account of** disputes between the city of Hamburg and the promoter.’ (*Süddeutsche Zeitung*, January 18th, 2015)

Which linguistic expressions qualify as complex adpositions in a particular language, and on which grounds, is still a matter of considerable debate. Moreover, the status of the category ‘complex adpositions’, applying to pre- and postpositions equally, is not entirely clear.

Although the category is commonly accepted nowadays (in the tradition of Quirk *et al.*, 1985), it is still resisted by some researchers, who basically deem such items to be equivalent to adnominal or nominal groups that have a certain degree of idiomacity (Huddleston & Pullum, 2002). In order to deal with complex adpositions in a cross-linguistic and typological perspective, as we have tried to do in this book, we felt the need of a common working definition of this sub-class of expressions.

3 Working definitions: Adpositions and nominal relators

In this section, we propose a general definition of the category of complex adpositions, including complex adnominal relators. The definition is thus opera-

tional both for languages with adpositions and for languages with other types of nominal relators, such as case-marking and relational nouns.

3.1 Simple nominal relators

It is commonly accepted that a construction such as (Fr) *à cause de* ‘because of (lit. *at cause of*)’ is a complex adposition. But what does that mean exactly? Are complex adpositions very different from simple ones? How does one identify a complex adposition? Is it a homogeneous subclass of adpositions, a different word class altogether, or neither? In order to look for answers to these questions, we believe it is fundamental to **start out with a working definition of complex adpositions** (and their distributional equivalents in languages with no adpositions), which itself requires, first, a working definition of *simple adpositions* and more generally of simple nominal relation markers or *relators*.

Simple relators: In many languages, basic *nominal relation markers* (i.e. relation markers which prototypically introduce nouns or noun phrases), which we call here relators, are CASE MARKERS and ADPOSITIONS;

- 1) these are well-known categories, rather well identified, with wide typological variation (in number of items per language, semantics, formal complexity, etc.)
- 2) they are SIMPLE relators, i.e. they consist of one morpheme or lexeme: (Fr) *à* ‘at’, *devant* ‘in front of’, *sur* ‘on’, (Eng) *on*, *in*, (Sp) *ante* ‘in front of’, *tras* ‘behind’...; this is also true of case markers, which generally consist of one or two morphemes at the most.
- 3) these simple relators have three distinct functions:
 - (at the syntactic level)
 - i. they introduce an *a priori* nominal constituent
 - ii. they establish a relation between this (lower) constituent and another (higher) constituent
 - (at the semantic level)
 - iii. they help determine the semantic nature of this relationship.³

³ In the examples below, the adpositions (i) introduce constituents (respectively *la maison*, *neuf heures*, *dinner*, *the judge*, *ścieżka*, *las*), (ii) establish a relation between these constituents and a higher constituent (respectively (i) *rentre*, (he) *came over*, (he) *stood*, *wróciliśmy*), and (iii) determine the semantic nature of this relationship – in these examples, with various spatial or temporal meanings, which are in part determined by the context itself.

- (a) (Fr) *il rentre à la maison* ‘he goes home (lit. *he comes.back to the house*)’
- (b) (Fr) *il rentre à neuf heures* ‘he comes back at nine o’clock’

- 4) SIMPLE RELATORS sometimes grammaticalize further, the result being their *decategorialization* (Hopper & Traugott 1993, Heine 2003). Simple adpositions can come to be used as determiners, complementizers, agent markers in passive constructions, and finally case makers, as noted by Lehmann (1985):
- i. determiner: (Fr) *de l'eau* ‘some water (lit. *of the water*)’
 - ii. complementizer: (Eng) *I would like **to** visit London*
 - iii. linking element: (Gm) *Zeitungsleser* ‘newspaper reader (lit. *newspaper. LINKING_s.reader*)’, (Eng) *yesterday’s paper*

These four features can be considered as definitional criteria for identifying simple nominal relators.

3.2 Complex nominal relators

Many languages have another type of nominal relators, which are not formally simple but COMPLEX;

- 1) a COMPLEX RELATOR is functionally equivalent to a simple one, but is MORPHOLOGICALLY complex;⁴
- 2) COMPLEX RELATORS have the same three distinct functions as simple relators; however, these three functions can (*but need not*) be performed by different components of the complex relator; typically:
 - i. a simple adposition (or case marker) marks the link with the lower constituent
 - ii. another simple adposition (or case marker) establishes the relation with the higher constituent
 - iii. a lexical nucleus (typically *nominal*) determines the semantic nature of the relation

(c) (Eng) *he came over before dinner*

(d) (Eng) *he stood before the judge*

(e) (Pol) *wróciliśmy ścieżką / przez las*
 come.back.PST.1PL.M path.INSTR through wood.ACC
 ‘we came back following the path / through the woods’

The semantic relations expressed by adpositions are wide-ranging, from spatial (e.g. topological and functional relations, see Vandeloise 1986, Svorou 1994) to temporal (e.g. antecedence, etc., see Haspelmath 1997) and abstract relations (e.g. cause, concession, purpose, agentivity, etc.).

4 This complexity is variable, in part because there is a diachronic tendency to routinization and fusion and/or shedding of the various elements through lexicalization and eventually grammaticalization.

In examples such as (Gm) *in Richtung auf* ‘toward (lit. *in the direction of*)’ or (Sb) *u pravcu* ‘in the direction of’ (Table 1), the simple adpositions (*auf* ‘on, to’) and the genitive case perform functions (i), and (ii) (*in* ‘in’, *u* ‘in’), while the lexical nucleus (*Richtung* ‘direction’, *pravac* ‘direction’) determines the semantic nature of the relation.

Table 1: Internal structure of complex adpositions.

Language	simple adposition / case marker	lexical nucleus	simple adposition/ case marker
German	<i>in</i>	<i>Richtung</i>	<i>auf</i>
Serbian	<i>u</i>	<i>pravcu</i>	GEN

- 3) a COMPLEX RELATOR is semantically more specific. In comparison, SIMPLE RELATORS tend to be semantically bleached and therefore seem much more polysemous.

There appears to be some sort of semantic equivalence between simple and complex relators, but complex relators typically appear in more restricted contexts, due to their higher lexical specification as in (30–32).

(30) *For / with respect to your book, he said it was okay.*

(31) *For (/ *With respect to) twenty years, I've lived here*

(32) *He came for (/ *with respect to) your book.*

- 4) COMPLEX RELATORS, unlike SIMPLE RELATORS, do not seem to have other uses than adpositional, except maybe for uses as complementizers, e.g. (Pt) *a fim de* ‘in order to’, *apesar de* ‘despite’.

The distinction between simple and complex relators is not clear-cut, however. This is because, diachronically, simple relators typically result from the grammaticalization of complex relators. There are different scenarios of simplification, with e.g. univerbation and agglutination (cf. Lehmann 2015: 160–162), namely:

- i. shedding of elements, e.g. (Gm) *in Richtung von* > *Richtung von* > *Richtung* ‘towards (lit. *in the direction of*)’, (Fr) *du côté de* > *du côté* > *côté* ‘on the side of > concerning’, *en face de* > *en face* ‘in front of’,⁵ (Pt) *em frente a* > *frente a* ‘in front of’.

⁵ The adpositional use of *en face* is non standard, but attested (i), at least since the 19th century (ii).

(i) bar de la Renaissance (*en face la poste*) ‘bar la Renaissance (across from the post office)’ (FrTenTen)

- ii. morphological and/or phonological fusion of grammatical and lexical elements, e.g. (Eng) *in stead of* > *instead of*, (Gm) *an Hand von* > *anhand von* ‘by means of (lit. *at the hands of*)’.⁶
- iii. a mixture of the two, e.g. (Gm) *an Statt* > *anstatt* > *statt* ‘instead of’.

4 Research questions

Although there has been, in the last years, an increased interest in complex adpositions, the field is in our view still vastly understudied. One reason is the lack of encompassing studies, bringing together results of this increasing body of descriptive research on the topic in various languages. Accordingly, many questions remain unanswered. Among the variety of topics open to exploration, this book aims to bring some answers to a few questions, which we describe in some detail below.

4.1 Definition and categorial status of CAs

The first question is naturally that of the definition of the class ‘complex adpositions’. Complex adpositions are a polemic category, inasmuch as not even their definition is consensual. Sequences such as (Eng.) *instead of* or *in order to*, or (Germ.) *aufgrund von* (+ GEN) are usually considered as complex prepositions, mainly due to their complex internal structure and to their relational function. However, the very definition of a complex adposition is a matter of debate, as shown e.g. by the different positions of two reference grammars of English. While Quirk *et al.* (1985) consider that there is justification for such a category, Huddleston & Pullum (2002) dismiss the concept on the basis of the syntactic properties of these sequences, although acknowledging their semantic autonomy (see also Seppänen *et al.* 1994). German reference grammars, on the other hand, admit several subcategories of adpositions, i.e. simple, complex, and ‘secondary’⁷ ones (e.g., *trotz*, *kraft*, *laut*, etc.), which are distinguished on

(ii) un appartement splendide *en face le parc Monceau* ‘a splendid flat across from the *parc Monceau*’ (Goncourt, *Journal*, II, p. 105, 19th c.; GGHF corpus).

⁶ This fusion is not always reflected in spelling, as in (Fr) *à travers*, which should be analyzed as a simple adposition [*à_travers*]: in examples such as *elle l’a jeté à travers la fenêtre* ‘she through it/him through the window’, nothing whatsoever can be inserted between *à* and *travers*.

⁷ This term has also widely been used by Slavic grammarians and linguists since the beginning of the 20th century, as opposed to “primitive” adpositions (see, among others, Hill 2012).

the basis of some prototypical features, including the transparency as to their lexical sources (Diewald 1997, 65–73).

What criteria are more relevant for the class of CAs, and to which extent are they dependent on the internal structure of CAs? For CPs, some have been proposed in the literature (Quirk *et al.* 1985, Seppänen *et al.* 1994, Borillo 1997, 2001, Adler 2001, Cifuentes Honrubia 2003, Fagard 2006, Leeman 2007), but no cluster of criteria seems consensual. Also, as research widens from concern with prepositions to concern with adpositions, criteria may be in need of revision. Comparing the system of adpositions in a series of more or less genetically and/or areally related languages is a plausible way of improving the existing criteria and definition.

4.2 Grammaticalization, lexicalization and constructionalization

Definitional uncertainty, however, has not prevented a certain consensus about the kind of grammaticalization and/or lexicalization processes that complex adpositions undergo. This concerns at least those CAs with the most common source structures, such as $[P_1(+\text{Det})+N+P_2]$ (e.g. (26–27) above; see Fagard & De Mulder, 2007). These processes can be dealt with within the framework of Lehmann's (2015) grammaticalization parameters: e.g., the change of P_1 from paradigmatic variability to invariability, the impossibility of including linguistic material between $P_1(+\text{Det})$ and N , and the eventual coalescence of $P_1(+\text{Det})+N$. However, many other grammaticalization (or lexicalization) paths need to be explored. Related to this is the issue of whether a regular order can be discerned in the processes of grammaticalization leading to the emergence of a member of one of these classes from a member of one of the others. For instance, the relation between CAs and closely related categories such as adverbs and conjunctions should be explored: while it seems reasonable to assume that grammaticalization processes involving adverbs, simple adpositions and conjunctions are unidirectional (Adverb > Adposition > Conjunction), is it possible to say the same about complex adpositions? And how do these processes fare under the unidirectionality hypothesis (Hopper & Traugott 2003)?

For CAs, how do frequency and degree of grammaticalization correlate? It is generally assumed in grammaticalization studies, as well as in usage-based theories (Bybee 2010), that higher type frequency correlates with a more advanced stage of grammaticalization. Many studies on CAs provide evidence for this hypothesis (e.g. Hoffmann 2005, Lima 2014, Stosic 2012), but more empirical studies are necessary.

4.3 Morphological and syntactic patterns of CAs

The determination of subcategories of CAs – which may reflect different stages in grammaticalization – should be discussed (see Adler’s 2008 distinction of compound prepositions and prepositional locutions).

Compiling inventories of structural types of CAs in European languages is central for a deeper understanding of their adpositional systems. Complex adpositions may have their origin in a great variety of syntactically as well as morphologically complex structures: besides $[P_1(\text{Det})NP_2]$, we can find $[P_1\text{Adv}P_2]$ (Fr. *en dehors de* ‘outside, except’), $[P_1P_2]$ (Sp. *para con* ‘towards, for’), $[\text{PTCP}P]$ (Pt. *devido a* ‘owing to, on account of’), $[\text{Adj}P]$ (Eng. *due to*), $[\text{NP}]$ (Pt. *face a* ‘in front of, given, in the face of’), $[\text{P}+\text{N}]$ (Ger. *mithilfe* ‘by means of, by dint of, thanks to’), $[\text{VP}]$ (Sp. *pese a* ‘despite’), etc. Do European languages share some of these schemas of CAs, and if so to which extent?

Related to this is the issue of productivity of CAs’ source structures. Some structures – such as $[P_1(\text{DET})NP_2]$ – seem to be highly productive and regular, while others seem to emerge *ad hoc* and have just a few members. Is the emergence of CAs always casual or are there patterns of productivity? What differences can be observed cross-linguistically regarding the productivity of structural patterns?

4.4 Emergence and evolution of CAs in European languages

In studies on grammaticalization, the need is felt for a deeper understanding of the processes that lead to the emergence of complex adpositions, as well as to their evolution and possible change into simple adpositions and case markers. The main aim of the present volume is to look for new insights into these problems in the domain of European languages. Thus, we bring together contributions on a varied array of European languages, so that a good picture of the different realities involving the emergence and the evolution of CAs may ensue. The background for these studies has been laid out in interesting work relating to complex prepositions and/or their grammaticalization, on the one hand (Lindqvist 1994, Schwenter and Traugott 1995, Di Meola 2000, Cifuentes Honrubia 2003, Hoffmann 2005, among others), and to adpositions on the other (as, e.g., Kurzon & Adler 2008, Hagège 2010, Libert 2013). We aim at widening the object under analysis from the traditional one of complex prepositions to the more embracing one of complex adpositions, thus gaining the advantage of contemplating studies on complex postpositions (and circumpositions), which have been so far neglected, at least as far as their grammaticalization and/or lexicalization are concerned.

5 Layout of the book

5.1 Complex adpositions in Romance languages

In the chapter “Complex adpositions in Romance Languages”, Benjamin Fagard, Dejan Stosic and José Pinto de Lima describe the systems of complex adpositions in modern Romance languages. One obvious question with respect to complex adpositions is that of their status: are they simply a morphosyntactically complex variant of simple adpositions? How should they be defined? While there have been a few studies on complex adpositions in individual Romance languages (for instance Spanish, Cifuentes Honrubia 2003, or Portuguese, Lima 2014, 2019a, 2019b), and a few papers on complex adpositions in Romance languages in general (for instance Fagard & Mardale 2007), there is still much to say. With a corpus-based analysis focusing on three Romance languages, French, Portuguese and Romanian, the authors show that some features of complex adpositions are common to most if not all varieties, while some seem to be specificities of one or more Romance language(s). One case in point is the way such constructions can be tested (see Adler 2001, 2008), for instance with possessives – a test which does not seem to affect French and Portuguese (Catalan, etc.) CAs in the same way. The clear contrast which appears in French between non-lexicalized constructions and CAs on account of their intolerance of the possessive (*à la place de Jean* ‘in John’s seat (or place)’ → *à sa place* ‘instead of John’; *au lieu de Jean* → **à son lieu* ‘in John’s stead’) is not found in Portuguese. In Portuguese, CAs do not exhibit this particularity, or at least to a lesser extent: *por causa do João* and *por sua causa* ‘because of John/of him’ are both possible, as is *em seu lugar* ‘in his stead’, etc.). The authors go on to formulate a hypothesis as to the development of CAs in various Romance languages, suggesting it may not have been parallel.

5.2 Complex adpositions in West-Germanic languages

In the chapter “Complex Adpositions in West-Germanic Languages”, Matthias Hüning, Elena Smirnova and Anatol Stefanowitsch provide an overview of simple and complex adpositions in West-Germanic languages, focusing particularly on Dutch, German, and English. They show that the class of prepositions is not a closed class in these languages, and in Germanic languages in general, and that there are several possibilities to extend this class.

The chapter provides a comprehensive description of the inventories of simple and complex adpositions in the three languages, and attempts at their systemati-

zation in terms of morphosyntactic features, semantics, and patterns of formation. With respect to the diachrony of complex adpositions in West-Germanic languages, the chapter proposes a unified account of their emergence and development in terms of grammaticalization and constructionalization. The focus of this chapter is on structural and diachronic aspects of so called PNP-constructions (i.e. the structural pattern ‘preposition – noun – preposition’) like NL. *met betrekking tot* – G. *in Bezug auf* – Eng. *with regard to*, constructions that are very productive today.

5.3 Complex adpositions in Slavic languages: The case of Serbian

In the chapter “Complex adpositions in Slavic languages: the case of Serbian”, Dejan Stosic describes the main morphological, syntactic and semantic features of the prepositional system of Serbian, a Slavic language, mostly in a synchronic perspective. He provides a morphosyntactic survey of inventories of simple and complex prepositions in Serbian – roughly 100 simple prepositions (see Mrazović 2009) and more than 400 complex ones. The most frequent morphosyntactic pattern of formation of complex adpositions is similar to the one found in other European languages, involving a noun preceded by one relator and followed by another, i.e. [Prep Noun + GEN], but other oblique cases can also appear as second relator. In many cases, the second relator corresponds to a preposition. Unlike simple adpositions, which generally select more than one case, complex adpositions display a very strict selection, limited to only one case. As for the evolution of Serbian complex adpositions, preliminary results of the corpus study suggest that this pattern of grammaticalization becomes especially widespread starting from the 15th century.

5.4 Prepositions in Brythonic Celtic: Stability and contact-induced change

In the chapter “Prepositions in Brythonic Celtic: Stability and Contact-Induced Change”, Kevin Rottet reports on the use of simple and complex adpositions in Brythonic Celtic. Prepositions represent a significant part of what is distinctive about the Celtic languages. Many prepositions trigger initial consonant mutations (*Breizh* ‘Brittany’ but *da Vreizh* ‘to Brittany’); simple prepositions are ‘conjugated’ for pronominal complements (*evit ma breur* ‘for my brother’ but *evitañ* ‘for him’). There are several patterns of complex prepositions including one in which pronominal complements take the form of possessive infixes (*a-raok* ‘before’ but *en ho raok* ‘before you’, cf. *ho* ‘your’). Additionally, language contact, with English

(in the case of Welsh) and French (where Breton is concerned) has played a role in recent times. For instance, Breton expressions for ‘up’ and ‘down’ (*d’al laez* ‘to the top’, *d’an traoñ* ‘to the base’) are only weakly grammaticalized whereas in Welsh, *i fyny* ‘to a mountain’ and *i lawr* ‘to a floor’ are highly grammaticalized and have undergone nearly total interlingual identification with the English words *up* and *down*, becoming full-fledged prepositions (and not just adverbs), and even competing with the traditional pattern of phrases of bodily posture, e.g. the contact-induced *eistedd i lawr* ‘to sit down’ versus the traditional *mynd ar ei eistedd* lit. ‘to go on his sitting’.

5.5 Greek complex adpositions as a key to Greek adpositional development

In the chapter “Greek complex adpositions as a key to Greek adpositional development”, Pietro Bortone describes the adpositional system of Modern Greek. In Greek, a whole system of adpositions – appearing, initially, both before and after nouns – can be observed coming into existence (in combination with an older case system), then fading and being renewed. The use of multiple adposition+case combinations ([P1+case1], [P1+case2], etc.) was slowly abandoned, but new combinations of adverbs with old adpositions ([ADV1+P1], [ADV1+P2]) enabled the system to express lost semantic nuances. The strikingly rich Ancient Greek adpositional inventory was thus gradually replaced and diversified largely thanks to the use of compounds – which have become the norm in Modern Greek. Furthermore, in Modern Greek, a number of phrases distributionally comparable to prototypical adpositions are also in use. These can express meanings that Greek previously expressed with simple adpositions. One may therefore ask whether such Greek phrases, which are an open class, could be considered fully-fledged adpositions and, ultimately, where the boundaries of the ‘adposition’ category should be drawn. Those Greek phrases, after all, match constructions that are found in other modern European languages, and which some linguists want to classify as adpositions (e.g. ‘according to’, ‘at the time of’, ‘at the presence of’, ‘on the theme of’, ‘for the benefit of’, ‘under the power of’). This fact highlights also the elusive possibility that their use in Greek might be, in part, a contact-induced change.

5.6 Complex Prepositions in Albanian: A first assessment

In the chapter “Complex Prepositions in Albanian: a first assessment”, Guillaume Bonnet and Benjamin Fagard investigate the marking of nominal relations

in Albanian. They describe the use and importance of case marking and simple prepositions, and go on to investigate the existence and use of complex prepositions, on the basis of the (limited) literature and of a corpus study. Relying on two corpora, the Albanian National Corpus (ANC) and the parallel corpus OPUS, totaling 20 and 46 million words respectively, the authors provide frequency lists and distributional features of all simple prepositions as well as a list of sequences which could be analyzed as complex prepositions. This makes it possible for the authors to account for the formal and functional differences between different types of simple and complex adpositions. The authors also address the question of the emergence of CAs in Albanian, showing a possible effect of language contact.

5.7 Recycling grammar – development of (complex) adpositions in Estonian

In the chapter “Recycling grammar – development of (complex) adpositions in Estonian”, Anni Jürine provides an overview of complex adpositions (CAs) in Estonian, with a special focus on complex *postpositions* (CPs). The development of CPs is here viewed as an instance of recycling grammar – a phenomenon whereby new complex grammatical items develop as lexical and grammatical components merge (see Jürine & Habicht 2013, 2017).

Most Estonian postpositions are simple postpositions, and the author shows that the development of CPs in Estonian is a process which is still in its early stages and has so far received very little attention. Indeed, until quite recently (Veismann & Erelt 2017) the (sub)category was not found in grammars.

CPs in Estonian include two components – a noun and a simple postposition. According to Jürine (2016), who has investigated body part-related complex postpositions, CPs in Estonian have the following characteristics: unit interpretation, desemantization, decategorialization, and extension to new contexts. These features, which are listed also in many other accounts of CAs (e.g. Hoffmann 2005, Lehmann 2002, Villada Moirón & Bouma 2003, Adler 2008), may be used as criteria to distinguish Estonian CPs from their source form. Most CPs have both readings in contemporary Estonian. As such, the development of CPs is part of a developmental cycle of function words in Estonian where new grammatical items are formed by simple grams merging with lexical items:

NOUN + LOC. CASE SUFFIX	>	SIMPLE POSTPOSITION
NOUN + SIMPLE POSTPOSITION	>	COMPLEX POSTPOSITION

As there are many parallels in the development of complex and simple postpositions, the author observes the development of CPs against the backdrop of functional and formal properties of simple postpositions in Estonian. The investigation is carried out synchronically as well as diachronically.

5.8 Complex adpositions in Permic languages

In the chapter “Complex adpositions in Permic languages”, Timofey Arkhangel'skiy and Maria Usacheva provide an overview of nominal relations in Permic languages. Just as other Uralic languages, Permic languages have postpositions and no prepositions. Complex postpositions are very rare in Permic languages. The probable reason for this is that usually the postpositions are quickly agglutinated into complex case markers instead of existing as postpositional clusters. Besides, Permic spatial systems are quite complex: they include spatial cases (up to 8 in Udmurt, up to 9 in Komi, depending on the variety), postpositions with adverbial-like properties, and relational nouns (inflected postpositions) with noun-like properties. Yazva Komi also has a very developed system of deictic units which are often used instead of spatial PPs. However, some complex postpositions do exist. They consist of elements denoting ‘front’, ‘back’, ‘bottom’, ‘top’, ‘side’ in Udmurt; ‘front’, ‘top’, ‘edge’, ‘upper part’ in Komi. Most complex postpositions arise from PPs having a head with the meaning ‘side’ (Udmurt) or ‘edge’ (Komi). In Yazva Komi there are relational nouns derived from complex postpositions which consisted of two stems: *vardâl* ‘upper part’ (< *val* ‘top’ + *dar* ‘edge’) and *urdól* ‘lower part’ (< *ul* ‘bottom’ + *dar* ‘edge’). Permic complex postpositions are similar to simple spatial relational nouns in many respects. In Udmurt, they attach nominal dependents in the nominative or in the genitive, like both full-fledged and relational nouns. Unlike full-fledged nouns and like relational nouns, they allow for nominative dependents with definite referential status. However, they only accept pronominal dependents in the genitive, which puts them one step lower on the grammaticalization path than ‘true’ relational nouns. In Komi, complex postpositions attach dependents only in the nominative.

5.9 Case marking and complex adpositions in Basque

In the chapter “Case marking and complex adpositions in Basque”, Gerd Jendraschek shows that Basque, a language isolate, widely differs from its neighbors

with respect to nominal relators, having an elaborate inventory of case marking suffixes. The combination of a postposition with a bare nominal complement is rare, as postpositions typically govern a case, typically the genitive, although the dative, locative, ablative, and instrumental are also found. Some postpositions are invariable, for example *arabera* ‘according to’ (also used without a complement, i.e. as an adverb). Many postpositional phrases are of the structure [N GEN Nrel case], where Nrel stands for a relational noun such as *aurre* ‘front’ or *barru* ‘inside’, followed by one of the local cases. Postpositional complements with inanimate referents such as *etxe* ‘house’ can form a compound-like complex nominal with the relational noun, as in *etxe aurre* ‘house front’. Adverbial phrases are routinely turned into adnominal modifiers by adding *-ko*, as in *etxe aurreko parkea* ‘the park in front of the house’ (without genitive marking of the complement *etxe*) or *zure aldameneko mutila* ‘the boy next to you’ (with the genitive marker *-re* on the possessive pronoun *zure* ‘your’). Just as there are intermediate cases between case suffix and postposition, there are genitive constructions that are not typically listed among the postpositions, even though they fulfil a similar function, such as *ozeanoaren hondoan* ‘at the bottom of the ocean’. One criterion to distinguish complex postpositions from mere possessive constructions is the modifiability of the head, as in *ozeanoaren hondo sakonean* ‘at the deep bottom of the ocean’.

5.10 Structure and origin of Turkish adpositions

In the chapter “Structure and origin of Turkish adpositions”, Gerd Jendraschek reports on the system of nominal relations in Turkish. Turkish is the main representative of the Turkic language family, whose geographic distribution reaches from East Asia to the Balkans. The Turkic languages are possibly related to other languages in North-East Asia such as Mongolian, Korean, and Japanese, with which they are often grouped together as an Altaic or Trans-Eurasian macrofamily. While such a genetic connection remains controversial, these languages share many typological features. Turkish morphosyntax is overwhelmingly head-final and agglutinating, and as such it features a moderately elaborate system of five cases.

These include three local cases, which are a conflated allative-dative, a locative, and an ablative case. A possible addition to the case paradigm is the comitative-instrumental postposition *ile* ‘with’, which in spoken language cliticizes to the noun. A particularity of Turkish is the presence of simple postpositions governing an unmarked nominative with lexical nouns, but a genitive with

pronouns. While a few postpositions are monomorphemic, Turkish has a large number of derived postpositions, such as *geline* ‘as for’, a converb form of the stem *gel* ‘come’, or the ablative-governing postpositions *önce* ‘before’ and *sonra* ‘after’, likely derived from the nouns *ön* ‘front’ and *son* ‘end’, respectively. Certain compound-like noun phrases marked by a third-person possessive suffix on the head have been reinterpreted as adverbial adjuncts, turning the head noun into a postposition, as in *anlaşma sonucu* ‘(as a) result of the agreement’ (from *sonuç* ‘result’) or *kanun gereği* ‘according to the law’ (from *gerek* ‘necessity’). Local relations involving a spatial region have the pattern [N-GEN Nrel-POSScase], where Nrel is a relational noun such as *yan* ‘side’ or *üst* ‘top’, followed by a local case indicating direction, position, or origin, e.g. *masanın üstünden* ‘from the table top/surface’ or perhaps ‘from on the table’, where *masanın üstü* ‘the surface of the table’ is followed by the ablative case suffix. As for diachrony, it is important to note that a considerable part of present-day Turkish vocabulary is the result of the language reform of the 20th century, whose primary goal was the replacement of Arabic and Persian loanwords with deliberately coined neologisms. This also affected postpositions, where e.g. *rağmen* ‘despite’ of Arabic origin now coexists with *karşın* (based on *karşı* ‘against’).

5.11 Compound prepositions in Maltese: PREP-PREP combinations and related issues

In the chapter “Compound prepositions in Maltese: PREP-PREP combinations and related issues”, Thomas Stolz provides an overview of complex prepositions in Maltese. He focuses on the formal and functional properties of certain subtypes of compound prepositions (COMPREFs) as attested in contemporary Maltese. The presentation of the Maltese evidence of COMPREFs shows a high degree of heterogeneity, to the point that it seems hardly possible to claim that there is only one candidate for the status of COMPREF in Maltese. On the contrary, several distinct types of COMPREFs emerge, their members sharing different subsets of properties with the canonical COMPREF. The diachronic developments are suggestive of gradual transitions from one class to the other. These transitions, in turn, presuppose that the domain of COMPREFs is organized like a continuum but not exclusively along the lines of the polar distinction transparent vs opaque. Univerbations result from processes which affect two different categories, namely PREP-PREP combinations and PPs. It is these two sources which occupy the opposing extremes of the continuum with univerbations being situated in the continuum’s centre. This results in the creation of new complex prepositional lexemes, which

ultimately may be reanalyzed as monomorphemic (i.e. unanalyzable) units and thus enrich the inventory of simple PREPs.

5.12 Discussion: A (micro)typological approach to Language change – the case of nominal relators

In this last chapter, the editors of the volume discuss the notion of complex nominal relator on the basis of the data presented in the previous chapters. They show that there is indeed considerable evidence pointing towards a specificity of complex nominal relators (CNRs), independently from language type, at least in European languages: whatever their basic ‘relational’ components – prepositions, postpositions, case markers, relational nouns – the definition put forward in the introduction seems to be valid for all languages taken in consideration. The importance of a diachronic approach is obvious for the study of CNRs, which show (by definition) functional affinities with simple nominal relators (SNRs) and seem to function as a reservoir, from which SNRs are ‘pulled out’ by further grammaticalization. The authors also discuss a few examples which seem to contradict this *reservoir* hypothesis, and propose a few lines of research for future studies.

Abbreviations

ACC	accusative	LOC	locative
AUX	auxiliary	M	masculine
CL	clitic	N	neuter
DAT	dative	NEG	negation
DEI	deictic	NOM	nominative
DEM	demonstrative	PASS	passive
DET	determiner	PL	plural
F	feminine	POSS	possessive
FUT	future	PRO	pronoun
GEN	genitive	PRS	present
GER	gerund	PST	past
IMP	imperative	PTCP	participle
IND	indicative	REL	relative
INF	infinitive	SG	singular
		SBJ	subjunctive

References

- Adler, Silvia. 2001. Les locutions prépositives : questions de méthodologie et de définition. *Travaux de linguistique* 42–43. 157–170.
- Adler, Silvia. 2007. Locutions prépositives temporelles et modes d’anaphorisation. In David Trotter (ed.), *Actes du XXIVe Congrès International de linguistique et Philologies Romanes*, 495–508. Tübingen: Niemeyer.
- Adler, Silvia. 2008. French compound prepositions, prepositional locutions and prepositional phrases in the scope of the absolute use. In Dennis Kurzon & Silvia Adler (eds.), *Adpositions: Pragmatic, semantic and syntactic perspectives*, 17–35. Amsterdam & Philadelphia: John Benjamins.
- Anschcombe, Jean-Claude. 2007. Les indicateurs aspectuels de déroulement processif : *en cours de*, *en passe de*, *en train de*, *en voie de*. *Cahiers de lexicologie* 90. 41–74.
- Auroux, Sylvain. 1989–2000. *Histoire des idées linguistiques*. Liège: Mardaga.
- Boissin, Henri. 1975. *Grammaire de l’albanais moderne*. Paris: Chez l’auteur.
- Borillo, Andrée. 1991. Le lexique de l’espace : prépositions et locutions prépositionnelles de lieu en français. In Liliane Tasmowski & Anne Zribi-Hertz (eds.), *Hommage à Nicolas Ruwet*, 176–190. Gand: Communication & Cognition.
- Borillo, Andrée. 1997. Aide à l’identification des prépositions complexes de temps et de lieu. *Faits de langue* 9, 173–184.
- Borillo, Andrée. 2000. Degrés de grammaticalisation : des noms de partie aux prépositions spatiales. *Travaux linguistiques du CERLICO* 13. 257–274.
- Borillo, Andrée. 2001. Il y a prépositions et prépositions. *Travaux de linguistique* 42–43. 141–155.
- Bortone, Pietro. 2010. *Greek prepositions – from antiquity to the present*. Oxford: Oxford University Press.
- Bree, Cor van. 2014. Adpositionele configuraties in het Nederlands. *Acta Universitatis Wratislaviensis. Neerlandica Wratislaviensia* 24. 11–25.
- Brøndal, Viggo R. 1940. *Præpositionernes Theori – indledning til en rationel Betydningslære* [Theory of prepositions – Introduction to a rational study of semantics]. Copenhagen: B. Lunos.
- Brøndal, Viggo R. 1950. *Théorie des prépositions. Introduction à une sémantique rationnelle*. Copenhagen: E. Munksgaard.
- Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Căpățână, Cecilia. 2000. Locuțiunile prepoziționale [Complex prepositions]. In Doina Negomireanu (ed.), *Ion Coteanu. In memoriam*, 84–89. Craiova: Editura Universitară.
- Căpățână, Cecilia. 2003. Locuțiunile prepoziționale [Complex prepositions]. *Limba Română* [Romanian language] XIII(1). 48–53.
- Casadei, Federica. 2001. Le locuzioni preposizionali. Struttura lessicale e gradi di lessicalizzazione. *Lingua e Stile* 36(1). 21–36.
- Cifuentes Honrubia, José Luis. 2003. *Locuciones prepositivas – sobre la gramaticalización preposicional en español*. Alicante: Universidad de Alicante.
- Ciobanu, Fulvia. 1957. Observații asupra prepozițiilor și locuțiunilor prepoziționale din limba română [Observations on prepositions and complex prepositions in Romanian]. *Studii de gramatică* [Grammar studies] 2. 357–388.

- Comrie, Bernard, & Greville G. Corbett (eds.). 1993. *The Slavonic languages*. London & New York: Routledge.
- Di Meola, Claudio. 2000. *Die Grammatikalisierung deutscher Präpositionen* (Studien Zur Deutschen Grammatik 62). Tübingen: Stauffenburg.
- Diewald, Gabriele. 1997. *Grammatikalisierung. Eine Einführung in Sein und Werden grammatischer Formen*. Tübingen: Niemeyer.
- Dubinsky, Stanley W., & Kemp Williams. 1995. Recategorization of Prepositions as Complementizers: The Case of Temporal Prepositions in English. *Linguistic Inquiry* 26(1), 125–137.
- Fagard, Benjamin, & Alexandru Mardale. 2007. Systèmes prépositionnels des langues romanes: la notion de partie du discours en diachronie. In David A. Trotter (ed.), *Actes du XXIVe Congrès International de Linguistique et de Philologie Romanes*, 91–104. Tübingen: Niemeyer.
- Fagard, Benjamin, & Alexandru Mardale. 2019. Les prépositions complexes du roumain – approche sur corpus. In Benjamin Fagard, José P. Lima, & Dejan Stosic (eds.), *Les prépositions complexes dans les langues romanes*, 176–203. Special issue of *Revue Romane* 54(1), Amsterdam & Philadelphia: John Benjamins.
- Fagard, Benjamin, & Walter De Mulder. 2007. La formation des prépositions complexes: grammaticalisation ou lexicalisation ? *Langue française* 156, 9–29.
- Fagard, Benjamin. (To appear). Prépositions et locutions prépositives. In Christiane Marchello-Nizia, Bernard Combettes, Sophie Prévost, & Tobias Scheer (eds.), *Grande grammaire historique du français*. Berlin & New York: Mouton de Gruyter.
- Fagard, Benjamin. 2006. Evolution sémantique des prépositions dans les langues romanes : illustrations ou contre-exemples de la primauté du spatial ? University Paris 7 & University Rome 3 Dissertation.
- Fagard, Benjamin. 2010. *Espace et grammaticalisation – L'évolution sémantique des prépositions dans les langues romanes*. Sarrebruck: Editions Universitaires Européennes.
- Fagard, Benjamin. 2012. Prépositions et locutions prépositionnelles: la question du renouvellement grammatical. *Travaux de linguistique* 64(1), 161–189.
- Fagard, Benjamin, José P. Lima, & Dejan Stosic (eds.). 2019a. *Les prépositions complexes dans les langues romanes*. Special issue of *Revue Romane* 54(1). Amsterdam & Philadelphia: John Benjamins.
- Fagard, Benjamin, José P. Lima, & Dejan Stosic. 2019b. Les prépositions complexes dans les langues romanes. In Benjamin Fagard, José P. Lima, & Dejan Stosic (eds.), *Les prépositions complexes dans les langues romanes*. Special issue of *Revue Romane* 54(1), Amsterdam & Philadelphia: John Benjamins. 1–7.
- Găitănaru, Ștefan. 1965. Descrierea locuțiunilor prepoziționale cu genitivul [Description of complex prepositions with the genitive]. *Limba Română* [Romanian language], 3–4, 87–92.
- Găitănaru, Ștefan. 1999. Criteriile deosebirii locuțiunilor prepoziționale cu genitivul de grupările libere, analizabile [Criteria for distinguishing complex prepositions with the genitive from free, analyzable sequences]. *Limba și literatura română* [Romanian language and literature] XXVIII (2). 12–14.
- Gougenheim, Georges. 1950. *Valeur fonctionnelle et valeur intrinsèque de la préposition 'en' en français*. Paris : Grammaire et Psychologie.
- Gougenheim, Georges. 1959. Y a-t-il des prépositions incolores en français ? *Le français Moderne* 27(1), 125.

- Gross, Gaston. 2006. Sur le statut des locutions prépositives. *Modèles linguistiques* 53. 33–50.
- Hagège, Claude. 2010. *Adpositions. Function-marking in human languages* (Oxford Studies in Typology and Linguistic Theory). Oxford & New York: Oxford University Press.
- Hamp, Carl. 1888. Die zusammengesetzten Präpositionen im Lateinischen. *Archiv für lateinische Lexikographie und Grammatik* 5. 321–368.
- Haspelmath, Martin. 1997. *From space to time – Temporal adverbials in the World's Languages*. München & Newcastle: Lincom Europa.
- Heine, Bernd. 1989. Adpositions in African languages. *Linguistique africaine* 2. 77–127.
- Heine, Bernd. 2003. Grammaticalization. In Brian D. Joseph & Richard D. Janda (eds.), *Handbook of historical linguistics*, 575–601. Oxford: Blackwell.
- Hill, Steven. 2012. *The N-Factor and Russian Prepositions. Their Development in 11th–20th Century Texts*. Berlin: De Gruyter.
- Hoffmann, Sebastian. 2004. Are Low-Frequency Complex Prepositions Grammaticalized? On the Limits of Corpus Data – and the Importance of Intuition. In Hans Lindquist & Christian Mair (eds.), *Corpus Approaches to Grammaticalization in English*, 171–210. Amsterdam & Philadelphia: John Benjamins.
- Hoffmann, Sebastian. 2005. *Grammaticalization and English Complex Prepositions. A corpus-based analysis*. London & New York: Routledge.
- Hopper, Paul, & Elizabeth C. Traugott. [1993] 2003. *Grammaticalization*. Cambridge: Cambridge University Press.
- Huddleston, Rodney. 1984. *Introduction to the Grammar of English*. Cambridge: Cambridge University Press.
- Huddleston, Rodney, & Geoffrey Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Hüning, Matthias. 2014. Over complexe preposities en convergentie. In Freek Van de Velde, Hans Smessaert, Frank Van Eynde, & Sara Verbrugge (eds.), *Patroon en argument. Een dubbelfeestbundel bij het emeritaat van William Van Belle en Joop van der Horst*, 433–445. Leuven: Universitaire Pers Leuven.
- Jürine, Anni, & Külli Habicht. 2013. Kaassõnade tsükliiline areng. [The cycle of postpositions] *Keel ja Kirjandus* [Language and Literature] 12. 736–749.
- Jürine, Anni, & Külli Habicht. 2017. Grammaticalization of complex items: Estonian *nii* et ‘so that’, ‘so’. *Journal of Estonian and Finno-Ugric Linguistics* 8(2).
- Jürine, Anni. 2016. *The development of complex postpositions in Estonian: a case of grammaticalization via lexicalization*. Dissertationes Philologiae Estonicae Universitatis Tartuensis vol. 38.
- Kailuweit, Rolf. 2001. Lexeme, Kasusmarker, Relatoren? Überlegungen zu den spanischen Präpositionen unter dem Aspekt der Grammatikalisierung. In Barbara Schäfer-Prieß, Hildegard Klöden, & Rolf Kailuweit (eds.), *Grammatikalisierung in den iberoromanischen Sprachen*, 33–62. Wilhelmsfeld: G. Egert..
- Klégr, A leš. 2002. *English Complex Prepositions of the Type in Spite of and Analogous Sequences. A Study & Dictionary*. Prague: The Karolinum Press, Charles University.
- Kurzon, Dennis, & Silvia Adler (eds.). 2008. *Adpositions: Pragmatic, semantic and syntactic perspectives*. Amsterdam & Philadelphia: John Benjamins.
- Leeman, Danielle. 2007. De la préposition à la locution prépositionnelle. *La préposition en français. Modèles linguistique* 55. 7–15.
- Lehmann, Christian, & Christel Stolz. 1992. *Bildung von Adpositionen im Deutschen*. Erfurt: Seminar für Sprachwissenschaft der Universität Erfurt (ASSidUE, 6).

- Lehmann, Christian. 1985. Latin case relations in typological perspective. In Christian Touratier (ed.), *Syntaxe et latin. Actes du Ilme Congrès International de Linguistique Latine, Aix-en-Provence, 28–31 mars 1983*, 81–104. Aix-en-Provence: Université de Provence.
- Lehmann, Christian. 1992. Yukatekische lokale Relatoren in typologischer Perspektive. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 45. 626–641.
- Lehmann, Christian. 1995. Synsemantika. In Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld, & Theo Vennemann (eds.), *Syntax: Ein internationales Handbuch zeitgenössischer Forschung*, 1251–1266. (Handbücher der Sprach- und Kommunikationswissenschaft, 9/2.). Berlin: Walter de Gruyter.
- Lehmann, Christian. 2002. New Reflections on Grammaticalization and Lexicalization. In Ilse Wischer, Gabriele Diewald (eds.) *New Reflections on Grammaticalization*, 1–18. Typological Studies in Language 49. Amsterdam & Philadelphia: John Benjamins.
- Lehmann, Christian. 2015 (3rd edition). *Thoughts on grammaticalization*. Munich: Lincom Europa.
- Lesz-Duk, Maria. 2007. Wtórne przyimki przyczynowe we współczesnym języku polskim. [Secondary causal prepositions in modern Polish]. *Prace naukowe Akademii im. Jana Długosza w Częstochowie* VI. 67–73.
- Libert, Alan R. 2013. *Adpositions and Other Parts of Speech*. Frankfurt am Main: Peter Lang.
- Lima, José P. 2014. Grammaticalization of complex prepositions in European Portuguese. In José P. Lima, *Studies on Grammaticalization and Lexicalization*, 163–183. Munich: Lincom Europa.
- Lima, José P. 2019a. On grammaticalized complex prepositions in Portuguese: Deployment, shift, redundancy, complementation. In Benjamin Fagard, José P. Lima, & Dejan Stosic (eds.), *Les prépositions complexes dans les langues romanes*. Special issue of *Revue Romane* 54(1), Amsterdam & Philadelphia: John Benjamins. 126–140.
- Lima, José P. 2019b. Preposições complexas em Português [Complex prepositions in Portuguese]. In Ernestina Carrilho, Ana M. Martins, Sandra Pereira, & João P. Silvestre (orgs.), *Estudos linguísticos e filológicos oferecidos a Ivo Castro* [Linguistic and philosophical studies offered to Ivo Castro], 1181–1211. Lisboa: Centro de Linguística da Universidade de Lisboa. [<https://repositorio.ul.pt/handle/10451/39619>]
- Lindqvist, Christer. 1994. *Zur Entstehung der Präpositionen im Deutschen und Schwedischen*. Tübingen: Max Niemeyer Verlag.
- Mardale, Alexandru. 2009. Sur la distinction entre prépositions lexicales et prépositions fonctionnelles. In Rodica Zafiu, Blanca Croitor, & Ana-Maria Mihail (eds.), *Studii de gramatică. Omagiu Doamnei Profesoarei Valeria Guțu-Romalo* [Grammar studies. Tribute to Professor Valeria Guțu-Romalo], 115–125. Bucharest: Editura Universității din București.
- Mardale, Alexandru. 2011. Prepositions as a semilexical category. *Bucharest Working Papers in Linguistics* XIII(2), 35–50. Bucharest: Editura Universității din București.
- Meibauer, Jörg. 1995. Komplexe Präpositionen – Grammatikalisierung, Metapher, Implikatur und *division of pragmatic labour*. In Frank Liedtke (ed.), *Implikaturen* (Linguistische Arbeiten 343), 47–74. Tübingen: Niemeyer.
- Melis, Ludo. 2003. *La préposition en français*. Paris: Ophrys.
- Mrazović, Pavica. 2009. *Gramatika srpskog jezika za strance* [Serbian grammar for foreigners]. Novi Sad: Izdavačka knjižarnica Zorana Stojanovića.
- Nowak, Tomasz. 2008. *Przyimki lokatywno-inkluzyjne we współczesnym języku polskim*. [Locative-inclusive prepositions in modern Polish]. Katowice: Wydawnictwo Uniwersytetu Śląskiego.

- Petit, Daniel. 2015. Les prépositions nominatives de l'albanais. *Bulletin de la Société de Linguistique CX(1)*. 359–443.
- Pinkster, Harm. 1990. The development of cases and adpositions in Latin. In Harm Pinkster & Inge Genee (eds.), *Unity in diversity. Papers presented to Simon C. Dik on his 50th birthday, 195–209*. Dordrecht: Foris.
- Piunno, Valentina, & Vittorio Ganfi. 2017. Preposizioni complesse in italiano antico e contemporaneo. Grammaticalizzazione, schematismo e produttività. *Archivio Glottologico Italiano, CII(2)*. 184–204.
- Pottier, Bernard. 1962. *Systématique des éléments de relation*. Paris: Klincksieck.
- Schwenter, Scott, & Elizabeth C. Traugott. 1995. The semantic and pragmatic development of substitutive complex prepositions in English. In Andreas Jucker (ed.), *Historical pragmatics*, 243–273. Amsterdam & Philadelphia: John Benjamins.
- Seppänen, Aimo, Rhonwen Bowen, & Joe Trotta. 1994. On the so-called complex prepositions. *Studia Anglica Posnaniensia* 24. 3–29.
- Sigurd, Bengt. 1993. The problems of multi-word prepositions and subjunctives. *Working Papers in Linguistics* 40. Lund University, Dept. of Linguistics and Phonetics. 197–212.
- Spang-Hanssen, Ebbe. 1963. *Les prépositions incolores du français moderne*. Copenhagen: Gads Forlag.
- Stathi, Katerina. Forthc. *Die Entstehung der komplexen Präposition in Richtung im Deutschen: Zur Emergenz und Verfestigung von Konstruktionen*. In Alexander Lasch & Alexander Ziem (eds.), *Konstruktionsgrammatik VI. Historische Konstruktionsgrammatik: Konvergenzen und Divergenzen im Sprach- und Konstruktionswandel*. Tübingen: Stauffenburg.
- Stefanowitsch, Anatol. Forthc. Complex prepositions of contrast and opposition in English. In Marcin Grygiel & Barbara Lewandowska-Tomaszczyk (eds.), *Contrast and analogy in language: Perspectives from cognitive linguistics*. Amsterdam & Philadelphia: John Benjamins.
- Stosic, Dejan. 2012. *En passant par* : une expression en voie de grammaticalisation? *CORELA HS-12*, Special issue *Langue, espace, cognition*, <http://journals.openedition.org/corela/2844>. <https://doi.org/10.4000/corela.2844>.
- Stosic, Dejan, & Benjamin Fagard. 2019. Les prépositions complexes en français. Pour une méthode d'identification multicritère. In Benjamin Fagard, José P. Lima, & Dejan Stosic (eds.), *Les prépositions complexes dans les langues romanes*. Special issue of *Revue Romane* 54(1). Amsterdam & Philadelphia: John Benjamins. 8–38.
- Svorou, Soteria. 1994. *The Grammar of Space*. Amsterdam & Philadelphia: John Benjamins.
- Šiganova, Galina Aleksandrovna. 2001. *Sistema leksičeskix i frazeologičeskix predgov v sovremennom russkom jazyke* [The system of lexical and phrasal prepositions in Contemporary Russian]. Unpublished PhD dissertation, University of Tobol'sk.
- Vaguer, Céline. 2006. Bibliographie générale : recueils et articles consacrés aux prépositions du français. *Modèles linguistiques* 54 (XXVII–2). 171–203.
- Vandeloise, Claude. 1986. *L'espace en français : sémantique des prépositions spatiales*. Paris: Seuil.
- Veismann, Ann, & Mati Erelt. 2017. Kaassõnafraas [The Adpositional Phrase]. In Mati Erelt & Helle Metslang (eds.), *Eesti keele süntaks* [Estonian Syntax], 446–462. Eesti Keele Varamu [Estonian language library] vol. 3. Tartu: Tartu University Press.
- Veland, Reidar. 2011. Le locuzioni preposizionali *a(d) opera di, per opera di, per mano di* in italiano contemporaneo: funzionamento sintattico e proprietà semantiche. *Zeitschrift für romanische Philologie* 128(3). 537–552.

- Villada Moirón, María Begoña, & Gosse Bouma. 2003. A corpus-based approach to the acquisition of collocational prepositional phrases. In *Proceedings of the Tenth EURALEX International Congress*, EURALEX 2002, 153–158. Copenhagen, Denmark, August 13–17, 2002..
- Waters, Cathleen. 2009. The preposition cycle in English. In Elly van Gelderen (ed.), *Cyclical change. Linguistics today* 146. 285–301.
- Zelinsky-Wibbelt, Cornelia. 1993. *The Semantics of Prepositions – From Mental Processing to Natural Language Processing*. Berlin & New York: De Gruyter Mouton.



Indo-European language families & isolates

Benjamin Fagard, Dejan Stosic, and José Pinto de Lima

1 Complex adpositions in Romance: Emergence and variation

Abstract: In this paper, we describe the systems of complex adpositions in modern Romance languages, and discuss their status and definition. We do so with a corpus-based analysis focusing on three Romance languages, French, Portuguese and Romanian, and show that some features of complex adpositions are common to most if not all varieties, while some seem to be specificities of one or more of these languages. One case in point is the way such constructions can be tested, for instance with possessives – a test which does not seem to affect French and Portuguese (Catalan, etc.) CAs in the same way: the clear contrast found in French between non-lexicalized constructions and CAs on account of their intolerance of the possessive is not to be found in Portuguese. Finally, we formulate a hypothesis as to the development of CAs in various Romance languages, suggesting it may not have been parallel.

Keywords: Romance languages, corpus studies, contrastive linguistics, grammaticalization, lexicalization

1 Introduction

This article aims at describing formal and functional properties of complex adpositions in Romance languages. The paper is organized as follows: we provide an overview of Romance languages (sections 2.1 and 2.2) as well as of the methodology employed in our study (Section 2.3). We propose functional and formal definitions of simple and complex adpositions (Section 3.1); we briefly sketch an overview of complex adpositions in Romance languages (Section 3.2), and describe other possible uses of these constructions (3.3). We go on to describe the various patterns of complex adpositions, focusing first on the main patterns (4.1), then on existing subtypes (4.2), before addressing issues of frequency and productivity (4.3). We

Benjamin Fagard, CNRS, ENS & Paris Sorbonne Nouvelle; PSL, Montrouge, France. benjamin.fagard@ens.psl.eu

Dejan Stosic, University of Toulouse Jean Jaurès & Laboratory CLLE (CNRS & UT2J), Toulouse, France. dejan.stosic@univ-tlse2.fr

José Pinto de Lima, Universidade de Lisboa, Faculdade de Letras, LISBOA, Portugal. pintodelima@outlook.pt

<https://doi.org/10.1515/9783110686647-002>

then provide a diachronic account of the emergence of complex adpositions in French (5.1–5.3). We go on to examine variation in Romance languages, showing both commonalities (6.1) and divergence (6.2).

2 Language description

2.1 Romance languages – sociolinguistic and geographical features

Romance languages are a subgroup of the vast family of Indo-European languages. They originate from Latin, but have been for centuries under strong influence from other languages, both Indo-European and non-Indo-European, mainly Germanic, Arabic, Slavic and Celtic. They are mostly spoken in Europe (where all Romance languages are attested), in the Americas (mainly French, Spanish and Portuguese), and in Africa (mainly French and Portuguese), but there are also smaller, isolated communities elsewhere. The total number of mother-tongue speakers is around 700 million, among which around 200 million in Europe (by *Ethnologue*¹ counts, which we compiled). There are also Creoles based on Romance languages, with a total of around 11 million speakers in various parts of the world, from the Caribbean to the Indian Ocean; most speakers use a French-based Creole (around 10 million), followed by Portuguese-based (700,000 speakers) and Spanish-based Creoles (300,000 speakers). However, from a typological point of view, Creoles in general, and even Romance-based ones, are rather homogeneous, and very different from European Romance, and it does not seem justified to address them jointly: we believe Romance-based Creoles deserve a separate, dedicated study. For questions of simplicity, we shall henceforth use the term “Romance” to refer exclusively to European Romance languages, to the exclusion of Romance-based Creoles, and focus on European varieties.

¹ In its web version (www.ethnologue.com, accessed November 19th, 2018). Note that some numbers seem low; for French, for instance, according to the ‘Organisation internationale de la francophonie’, the actual number of native speakers in the world is 115 million, with 274 million speakers in total (Wolff (ed), 2014). The numbers in Table 1 should therefore be taken with caution; however, they do provide a rough idea of the number of speakers and the relative (numerical) importance of each Romance language.

Table 1: Number of speakers for the main Romance languages (in millions).

Variety	Number of speakers in the country of origin: all users (L1 only)	Total number of speakers: all users (L1 only)
Romansch	0.04	0.04
Occitan	0.5 ^a	2
Galician	2.3	2.3
Catalan	8.8 (3.7)	9.2 (4.1)
Romanian	18.7	23.4
Italian	59	67.8 (64.8)
French	63.2 (60.2)	285 (76.8)
Portuguese	9.9	236.5 (222.7)
Spanish	46.4 (42.7)	513 (442.4)

^aAccording to Sibille (2010), the number of speakers could be estimated at 700,000 based on data gathered in 1999. The actual number of speakers of Occitan is actually very difficult to evaluate, on account of the complex socio-linguistic situation – Occitan is a mosaic of dialects with no or little standard form, spoken mainly in France, Italy and Spain. This estimate goes well beyond Bernissan’s (2012) estimate, which seems very low compared to previous estimates (generally around 2 million speakers, see e.g. the Euromosaic report in 1996).

2.2 Romance languages – major typological characteristics

Romance languages are by and large inflectional as far as verbs are concerned, with a heavy Tense-Aspect-Modality system, but not so much for nouns and adjectives, which display variable gender and number marking and no case marking, with the exception of Romanian. Romance languages generally have preposed definite and indefinite articles. They have a large set of prepositions and conjunctions, which sometimes partly overlap, as in ‘before (prepositional)’ vs ‘before (conjunction)’: French *avant / avant que*, Italian *prima di / prima che*, Portuguese *antes de / antes que*, etc. (see Section 3.3). French differs from “Central Romance” on a number of points besides word order and pronouns; it displays for instance fully grammaticalized analytic future and past tenses, as well as two distinct paradigms for demonstratives (demonstrative pronouns vs. demonstrative determiners). Romanian, on the other hand, presents a set of distinctive Balkan features. For the nominal sphere, the most notable features are the postposed definite article (e.g. *lup.ul* ‘wolf.DET.NOM/ACC’) and a residual set of nominal case markers with genitive-dative fusion (e.g. *lup.ul.e* ‘wolf.DET.VOC’, *lup.ul.ui* ‘wolf.DET.GEN/DAT’). This specificity of French on the one hand (see e.g., Koch 2002) and Romanian on the other can also

be observed in the case of complementizers: while in ‘Central’ Romance there is a two-way distinction (*if* vs *that* complementizers), with a clear opposition between indicative and subjunctive complementation, Romanian is set apart by its three-way distinction and French by the higher degree of grammaticalization of the subjunctive (Fagard et al. 2016).

2.3 Methodology

In the following sections, we investigate formal and functional aspects of Romance complex adpositions. In order to investigate their uses, we combined different methodologies.

In a first phase, we based our hypotheses on the existing literature (for instance Borillo 2000, 2002, Cifuentes Honrubia 2003, Hoffmann 2005, Leeman 2007, Fagard and De Mulder 2007, Kurzon and Adler 2008, Lima 2014, 2019a, 2019b, Stosic to appear), and on traditional philological instruments, i.e. dictionaries and grammars (see the Reference section)². In a second phase, we relied on corpora to check the validity of our results. The use of corpora for determining how speakers make use of language is now quite well established, as one can see from the flourishing subfield of corpus-based grammars (e.g., for English: Biber et al. 1999; for French: Abeillé, Godard and Delaveau in press, and in diachrony Marchello-Nizia et al. in prep. etc.). It is based on the idea that linguistic systems are best viewed as language(s) in use, and that grammar is shaped by frequency effects (see e.g., Bybee and Hopper 2001, Bybee 2010) – in other words, it seems to us that it is more interesting to look at language use and find out from there what rules the speakers are following than to rely exclusively on introspection. The corpora we used are the following (links to the corresponding databases are given in the Reference section):

- French: Valibel, PFC, Clapi, Frantext, frTenTen;
- Italian: Badip, itTenTen;³
- Portuguese: Corpus do Português (Davies and Ferreira 2006), ptTenTen11, CRPC (*Corpus de Referência do Português Contemporâneo*), CETEM-Público;

² Grammars do not systematically provide lists of complex adpositions, and even when they do, they are typically quite limited. However, there is a tendency in the last decade for grammars to provide better accounts of Complex Adpositions. For instance, for Portuguese, the grammar of Raposo et al. lists 50 items or so (in a 3-page description, Raposo et al. 2013, 1503–1506); the *Nueva gramática de la lengua española* (2009) is another good example, with a fair section on complex adpositions (2276–2283).

³ For Italian, see also Piunno and Ganfi (2019).

- (5) Il vient **au-devant** **de** nous (French)
 he come.PRS.IND.3SG at.DET.M.SG-front to us
 ‘He’s coming toward us’
- (6) May nu intenționează să demisioneze, (Romanian)⁵
 May NEG intend.PRS.IND.3SG COMP resign.PRS.SBJ.3SG,
în pofida rezultatului conservatorilor în
 in spite result.DET.DEF.M.GEN.SG conservative.GEN.PL in
 alegerile pe care le-a convocat
 election.GEN.PL for which it-AUX.PRS.IND.3SG summon.PST.PTCP
 ‘May does not intend to resign despite the results of the conservatives in the
 elections she summoned’
- (7) **În pofida a** ceea ce cred (Romanian)⁶
 in spite at that COMP.REL believe.PRS.IND.3PL
 susținătorii unui Grexit, devalorizarea
 supporter.NOM.PL one.GEN.SG Grexit, devaluation.DET.DEF.F.NOM.SG
 monetară este doar o iluzie
 monetary.F.SG be.PRS.IND.3SG only DET.INDF.F.NOM.SG illusion.NOM.F.SG
 ‘In spite of what the supporters of Grexit believe, monetary devaluation is
 only an illusion’

In (1), the genitive case marks the relation between the proper noun *Ana* “Ann” and the NP *tatăl* “the father”. The same relation is marked in (2) by the simple adposition *de* “of”, realized as *d’*. In these two examples, the genitive case and the adposition *de* can be said to have the same function. The same can be said of English *of*, German *von*, with an alternation in both languages between case marking and prepositional marking of the genitive.

Similarly, whereas e.g. Modern French and Portuguese use the preposition *de* to introduce the complement of various types of complex adpositions (3–5), Romanian may use *de* but it alternates with case marking and the use of *a* (6–7). This complex alternation is also found in Old French, with three constructions

⁴ <https://www.news.ro/externe/may-nu-intentioneaza-sa-demisioneze-in-pofida-rezultatului-conservatorilor-inalegerile-pe-care-le-a-convocat1922403609002017061017026254>, accessed June 6th, 2017.

⁵ <http://politeia.org.ro/stiri-revista-presei/revista-presei-19-iunie-mitomanul-se-blindeaza-cu-ajutorul-lui-nitu-sial-lui-oprea/38748/>, accessed June 6th, 2017.

used to express the genitive: case marking and the adpositions *à* and *de* (e.g. *la mort le roi* ‘the king’s death (lit. *the death the king*, with an oblique case marking on *king*)’ vs *la mort du/au roi* ‘the king’s death (lit. *the death of.the/at.the king*)’).

In (8), there are two distinct adpositions, each marking a different semantic (and syntactic) relation. The adposition *de* introduces the argument of the verb *vient* ‘comes’ and marks it as one type of argument, as shown by example (9).

(8) Il [vient **de** [**chez** sa mère]] (French)
 he come.PRS.IND.3SG from at his.F.SG mother
 ‘He comes from his mother’s place’

(9) Il vient **à** Paris (French)
 he come.PRS.IND.3SG at Paris
 ‘He comes to Paris’

In contrast, the adposition *chez* establishes a type of spatial relation between the general motion event (*venir de* ‘comes from’) and the following NP *sa mère* ‘his mother’. The two adpositions are thus independent from each other, and can vary independently, as shown in (10–11).

(10) Il vient **chez** sa mère (French)
 he come.PRS.IND.3SG at his.F.SG mother
 ‘He comes to his mother’s place’

(11) Il vient **de** Rome (French)
 he come.PRS.IND.3SG from Rome
 ‘He comes from Rome’

Conversely, Fr. *près de*, Pt. *ao pé de* and Fr. *au-devant de* in examples (4–6) seem to be distributional equivalents of simple adpositions, the main difference being that they are morphologically complex. The fact that there is a functional and more specifically a semantic equivalence between such constructions and simple adpositions has been well documented e.g. for French (Melis 2003: 112–114) and Portuguese (Lima 2014, 2019a, 2019b). However, corpus studies show that this equivalence is not complete; for instance, the polysemy of complex adpositions seems much more limited than that of simple adpositions, and they have few or no purely grammatical uses (Fagard and De Mulder 2007, Fagard 2009, 2012).

Finally, Rom. *în pofida* (+ genitive marking of the dependent noun) in example (6) shows similar distributional and semantic properties, as is shown by the existence of the variant *în pofida a* (7) which corresponds more closely to the complex adposition in (6); the difference is the use of case marking (genitive) or adposition (*a* ‘at’) to establish the relation with the NP.

3.2 Canonical and non-canonical complex adpositions

Following the definition given in the introductory chapter, we define a complex adposition as having the same functions and distributional properties as simple adpositions (a–c), but with different formal characteristics (d–f).

The **typical** functions of simple adpositions are as follows – they:

- a. introduce a constituent
- b. establish a relation between this constituent and a higher constituent, or the sentence as a whole
- c. contribute to determining the semantic nature of this relationship

Complex adpositions are functionally equivalent to simple ones, i.e. their typical functions also correspond to features (a–c), but, additionally:

- d. they are MORPHOLOGICALLY complex
- e. canonically, functions (a–c) are performed by different components:
 - i. a simple adposition (or case marker) introduces the constituent
 - ii. another simple adposition (or case marker) establishes the relation with the higher constituent
 - iii. a lexical nucleus determines the semantic nature of the relation

Adpositions fulfill three different functions, and complex adpositions are – by definition! – complex, i.e. made up of different lexemes or morphemes. Therefore, as illustrated in Table 2, a *canonical* complex adposition would be one in which each one of the three functions of adpositions (a–c) is realized by a different component, while a non-canonical complex adposition would be one in which there is no such one-on-one mapping.

Complex adpositions may thus be non-canonical because there is no specific item dedicated to the introduction of a complement (e.g. Fr. *en ce qui concerne*, *à travers*), or no item establishing the relation to a higher constituent (e.g. Pt. *graças a*, *quanto a*). It can also be on account of the complexity of one of the slots, as in Pt. *no que respeita a*.

Table 2: Canonical vs non-canonical complex adpositions.

		Function			meaning/gloss
		(b)	(c)	(a)	
'canonical'	(Portuguese)	<i>ao</i>	<i>pé</i>	<i>de</i>	'next to (lit. <i>at.the foot of</i>)'
	(French)	<i>à</i>	<i>côté</i>	<i>de</i>	'next to (lit. <i>at side of</i>)'
	(French)	<i>en</i>	<i>dessous</i>	<i>de</i>	'beneath (lit. <i>in underneath of</i>)'
	(Portuguese)	<i>por</i>	<i>debaixo</i>	<i>de</i>	'beneath (lit. <i>by underneath of</i>)'
	(Portuguese)	<i>ao</i>	<i>lado</i>	<i>de</i>	'next to (lit. <i>at.the side of</i>)'
	(Portuguese)	<i>em</i>	<i>cima</i>	<i>de</i>	'above, on (lit. <i>on top of</i>)'
	(Portuguese)	<i>em</i>	<i>vez</i>	<i>de</i>	'instead of (lit. <i>in turn of</i>)'
	(Portuguese)	<i>em</i>	<i>relação</i>	<i>a</i>	'in relation to, concerning (lit. <i>in relation at</i>)'
	(Portuguese)	<i>por</i>	<i>meio</i>	<i>de</i>	'by means of (lit. <i>through means of</i>)'
	(Portuguese)	<i>por</i>	<i>causa</i>	<i>de</i>	'because of (lit. <i>through cause of</i>)'
'non-canonical'	(Portuguese)		<i>graças</i>	<i>a</i>	'owing to (lit. <i>thanks to</i>)'
	(French)		<i>grâce</i>	<i>à</i>	
	(Portuguese)		<i>quanto</i>	<i>a</i>	'concerning (lit. <i>how.much at</i>)'
	(French)		<i>quant</i>	<i>à</i>	
	(Portuguese)	<i>no</i>	<i>que</i>	<i>a</i>	'concerning (lit. <i>in.the what concerns at</i>)'
			<i>respeita</i>		
	(French)	<i>en</i>	<i>ce qui</i>		'concerning (lit. <i>in that which concerns</i>)'
		<i>concerne</i>			
(French)	<i>à</i>	<i>travers</i>		'through (lit. <i>at transverse</i>)'	

3.3 Other uses of complex adpositions

Simple adpositions often have other uses than adpositional ones, for instance subordination (9) (cf. Section 2.2). While this is sometimes taken to be a natural extension of their uses, to the point that Pottier (1962: 195–197) considers that e.g. French *avant* should be analyzed as one morpheme with different uses, depending on the context: adverb, preposition, conjunction. In (12–14), he would thus not posit the existence of an adposition, a conjunction and an adverb with similar form and meaning (*depuis* 'since'), but of only one morpheme *depuis* with three different uses.

- (12) il est changé **depuis**
 it be.PRS.IND.3SG change.PST.PTCP since
 la guerre
 the war
 'it has changed since the war' (Goncourt, *Journal*, 1896)

- (13) il est changé **depuis que** la guerre
 it be.PRS.IND.3SG change.PST.PTCP since COMP the war
 a commencé
 AUX.PRS.IND.3SG start.PST.PTCP
 ‘it has changed since the war started’
- (14) il est changé **depuis**
 it be.PRS.IND.3SG change.PST.PTCP since
 ‘it has changed since’

One limitation of this view is that it does not account for the fact that despite the frequent polyfunctionality of such morphemes, be it within a given language or across languages (cf. the other chapters of this volume, which show that uses of adpositions as adverbs or conjunctions are well represented across Europe), it is not universal. Some items have uses as adposition and adverb but not as conjunction, for instance Modern French *devant* ‘in front (of)’; others have all three uses but with differences in form, e.g. Italian *prima* and Portuguese *antes* ‘before’, which cannot be used as adpositions without a functional *di / de* ‘of’ (*prima / prima che / prima di, antes / antes que / antes de*). What is true, however, is that there is some functional and semantic overlap between adverbs and prepositional phrases on the one hand, prepositions and conjunctions on the other. To put it simply, an adposition links a sentence with a noun phrase, a conjunction links a sentence with a subclause, and it could be said that an adverb just anaphorically incorporates the noun phrase: in (11), *depuis* can stand for *depuis la guerre* ‘since the war’ if it is contextually salient.

Though it is not systematic, there seems to be a similar feature in complex adpositions: there are frequent formal correspondences between complex adpositions and complex conjunctions, for instance *au lieu de / au lieu que* (‘instead of’ / ‘except that’) in Medieval and Modern (rather non-standard) French (15).

- (15) au lieu que dans les écoles parisiennes et dans
 in.the stead that in the.PL school.PL Parisian.PL and in
 le primaire et le secondaire c’était le hasard
 the primary and the secondary it.be.PST.3SG the chance
 qui faisait les choses
 that do.PST.3SG the thing.PL
 [some people have received this officially...]
 ‘while in Parisian schools, and in primary and secondary schools, what did
 that was chance’ (CFPP2000_16-01_MARC-JAKOBSON_H_83_16E)

Beyond these uses, which remain functionally similar, simple adpositions may also grammaticalize further, acquiring a wide range of functions, including uses

as determiner (16), complementizer (17–18) (see e.g. English *to* used as complementizer or infinitive marker, German *zu*, Dutch *te*, etc.), differential object marking (19) (DOM, Regional French, 16; see also Romanian *pe*, Spanish *a*, Regional Italian *a*, among other Romance varieties), etc.

- (16) donne-moi **de** l' eau
 give-me of DET.DEF.F.SG water
 'give me (some) water'
- (17) elle m' a dit **de** venir
 she me.ACC AUX.PRS.IND.3SG say.PST.PTCP.M.SG of come.INF
 'she told me to come'
- (18) Il vient **pour** manger
 he come.PRS.IND.3SG for eat.INF
 'he comes to eat'
- (19) [ils] veulent t'interroger **à** toi et
 (they) want.PRS.IND.3PL CL.ACC.2SG.question.INF DOM you and
 moi hein (rires)
 me huh (laughs)
 they want to question you and me both (laughs) (PFC, Midi-Pyrénées,
 Fagard and Mardale 2014)

In such contexts, these morphemes are no longer adpositions. It should be noted that complex adpositions do not seem to display this characteristic of simple adpositions, at least in Romance.

4 Complex adpositions in French

In this section, we describe in detail the system of complex adpositions found in one Romance language, namely French.

4.1 Main schema

In terms of types, the most productive schema of complex adpositions in French is made up of three to four elements, i.e. [PREP₁ (DET) N PREP₂]:

- 1) preposition₁ (typically *à*)
- 2) definite article
- 3) noun
- 4) preposition₂ (typically *de*)

This schema accounts for complex adpositions such as *au-dessus de* ‘above’, *au lieu de* ‘instead of’, *à l’instar de* ‘like’, etc. For each of these four elements, there is a wide range of variation: the first and fourth – adpositions – are most often *à* and *de*, but other adpositions can be found in their place, such as *en* ‘in’, *avec* ‘with’, etc., including sometimes in a given construction (e.g. *en-dessous de* vs *au-dessous de* ‘beneath’); they can also be absent. The same applies for the determiner, which is rarely present. Finally, the third element is not necessarily a noun; it can be a verb, an adverb, etc. The next subsection, on subtypes of complex adpositions in French, provides more detail on the extent of this variation.

4.2 Subtypes

The most productive subtypes in Modern French are the main schema, [PREP₁ (DET) N PREP₂], and variants thereof; there are also, however, a number of marginal subtypes (see Table 3).

Table 3: Typical patterns of CAs in Modern French (adapted from Stosic, to appear, on the basis of grammars, dictionaries and existing literature on CAs; the list cannot be exhaustive, but is representative. Numbers indicate the number of *constructions* found for each pattern, not their frequency).

Pattern	Total	Examples
P ₁ DET N <i>de</i>	27	<i>à l’égard de</i> ‘concerning’, <i>à la fin de</i> ‘at the end of’, <i>au bord de</i> ‘at the edge of’, <i>sous l’égide de</i> ‘under the aegis of’, <i>sur les traces de</i> ‘on the steps of’, <i>au-dessous de</i> ‘below’
P ₁ N <i>de</i>	122	<i>à base de</i> ‘with, made from’, <i>en dépit de</i> ‘in spite of’, <i>sous prétexte de</i> ‘under the pretense of’ <i>sur ordre de</i> ‘by order of’
P ₁ N P ₂ (P ₂ ≠ <i>de</i>)	20	<i>par comparaison avec</i> ‘in comparison to’, <i>par rapport à</i> ‘in relation to’
ADV P ₂	20	<i>auprès de</i> ‘beside’, <i>autour de</i> ‘around’
P ₁ ADV P ₂	16	<i>à moins de</i> ‘unless, short of’, <i>en dehors de</i> ‘outside’
P ₁ P ₂	13	<i>d’après</i> ‘according to, based on’, <i>par-devers</i> ‘on’
P ₁ ADV	11	<i>à même</i> ‘next to, directly on’, <i>par-dessus</i> ‘on top of’
N P ₂	11	<i>dos à</i> ‘back to’, <i>face à</i> ‘in front of’, <i>grâce à</i> ‘thanks to’
P ₁ DET ADJ N <i>de</i>	10	<i>au beau milieu de</i> ‘in the middle of’, <i>au fin fond de</i> ‘in the depths of’

Table 3 (continued)

Pattern	Total	Examples
Other patterns	63	<i>au plus bas de</i> ‘at the lowest of’, <i>pour ce qui regarde</i> ‘concerning’, <i>quant à</i> ‘according to, concerning’, <i>quitte à</i> ‘at the risk of’, <i>proche de</i> ‘close to’, <i>à égale distance de</i> ‘at equal distance from’, <i>pas loin de</i> ‘not far from’, <i>au vu et au su de</i> ‘known to’, <i>en passant par</i> ‘by way of, including’, <i>mis à part</i> ‘excluding’, <i>il y a</i> ‘ago’

The initial preposition is most typically *à*, which appears roughly in half of the constructions, less frequently *en*, *sous*, *par*, *de*, *sur* and *dans*, rarely other adpositions (see Table 4a); the final preposition is almost always *de* (Table 4b).

Tables 4a & 4b: Patterns of CAs in Modern French: (a) initial and (b) final adposition (adapted from Stosic, to appear).

Prep ₁	Productivity (number of CAs)	Prep ₂	Productivity (number of CAs)
<i>à</i> ‘at’	240	<i>de</i> ‘of, from’	476
<i>en</i> ‘in’	79	<i>à</i> ‘at’	28
<i>sous</i> ‘under’	50	<i>avec</i> ‘with’	10
<i>par</i> ‘by, via’	35	others	11
<i>de</i> ‘of, from’	31		
<i>sur</i> ‘on’	19		
<i>dans</i> ‘in’	18		
others	26		

4.3 Number & frequency

The frequency of simple adpositions in Modern French is far from homogeneous. On the one hand, a few adpositions are very frequent: *à*, *de*, *en* are the most typical *functional* adpositions (maybe not the best term, see Lehmann 2019, Footnote 2), which are highly grammaticalized, and generally considered as a subtype of simple adpositions. Their frequency ranges from 15‰ (i.e. per thousand words) to 70‰ in Modern Corpora (Fagard, to appear). On the other hand, some fifty to a hundred other simple adpositions, identified as *lexical* adpositions, display much lower frequency levels, generally less than 1‰ (i.e. a thousand occurrences per million words), as shown in Table 5. There is also much variation in the frequency levels of complex adpositions; however, it can be shown that their frequency is much lower than that of functional adpositions, and even lower than some lexical adpositions (Table 5).

Table 5: Relative frequency of simple and complex adpositions in Modern French (GGHF corpus, 19th–20th c., roughly 5 million words).

Frequency range (%)	Simple adpositions	Complex adpositions
above 5	<i>de</i> 'of, from', <i>à</i> 'at', <i>dans</i> 'in, into', <i>en</i> 'in', <i>pour</i> 'for', <i>par</i> 'by'	
between 1 and 5	<i>sur</i> 'on', <i>avec</i> 'with', <i>sans</i> 'without', <i>chez</i> 'at'	
between 0.5 and 1	<i>sous</i> 'underneath', <i>après</i> 'after', <i>entre</i> 'between', <i>jusque</i> 'until', <i>contre</i> 'against', <i>vers</i> 'toward', <i>depuis</i> 'since', <i>devant</i> 'in front of'	
between 0.1 and 0.5	<i>pendant</i> 'during', <i>avant</i> 'before', <i>dès</i> 'from, since', <i>malgré</i> 'despite', <i>parmi</i> 'among', <i>derrière</i> 'behind', <i>selon</i> 'according to', <i>par-dessus</i> 'above, on top of'	<i>autour de</i> 'around', <i>d'après</i> 'according to, based on', <i>au fond de</i> 'at the bottom of', <i>auprès de</i> 'beside', <i>quant à</i> 'according to, concerning', <i>au-dessous de</i> 'below'
between 0.01 and 0.1	<i>hors</i> 'outside', <i>sauf</i> 'except', <i>envers</i> 'towards', <i>suivant</i> 'according to', <i>durant</i> 'for, during', <i>environ</i> 'about', <i>excepté</i> 'except, besides', <i>moyennant</i> 'by means of, in exchange for'	<i>il y a</i> 'ago', <i>au bord de</i> 'at the edge of', <i>à la fin de</i> 'at the end of', <i>par rapport à</i> 'in relation to', <i>grâce à</i> 'thanks to', <i>en dépit de</i> 'in spite of', <i>en dehors de</i> 'outside', <i>à l'égard de</i> 'with respect to', <i>face à</i> 'in front of', <i>dos à</i> 'back to', <i>proche de</i> 'close to', <i>à même</i> 'next to, directly on', <i>en passant par</i> 'by way of, including', <i>sous prétexte de</i> 'under the pretense of'
less than 0.01	<i>outré</i> 'beyond, besides', <i>touchant</i> 'concerning', <i>vu</i> 'seen, given, owing to', <i>nonobstant</i> 'despite', <i>par-delà</i> 'beyond', <i>concernant</i> 'concerning', <i>hormis</i> 'save, except', <i>devers</i> 'on', <i>fors</i> 'except' (archaic)	<i>pas loin de</i> 'not far from', <i>quitté à</i> 'even if it means, at the risk of', <i>à moins de</i> 'unless, short of', <i>au beau milieu de</i> 'in the middle of', <i>sur les traces de</i> 'on the steps of', <i>mis à part</i> 'excluding', <i>à base de</i> 'with, made from', <i>au fin fond de</i> 'in the depths of', <i>sous l'égide de</i> 'under the aegis of', <i>au plus bas de</i> 'at the lowest of', <i>au vu et au su de</i> 'known to', <i>sur ordre de</i> 'by order of', <i>à égale distance de</i> 'at equal distance from', <i>par comparaison avec</i> 'in comparison to', <i>pour ce qui regarde</i> 'concerning', <i>pardevers</i> 'on'

For simple adpositions, we included all items with 'PRP' tags, i.e. those categorized as 'prepositions', but checked manually and excluded some marginal items such as Old French *és* 'in' (lit. in.DET.PL), which appears twice in the corpus but only in quotations of a medieval text. For complex adpositions, we included those listed in Table 3.

Table 6: Main patterns of CAs in Old French (GGHF corpus).

Pattern	Constructions
<i>à</i> ‘at’ (+ DET) + N + P ₂	16
<i>en</i> ‘in’ (+ DET) + N + P ₂	27
<i>de</i> ‘of, from’ / <i>par</i> ‘by, via’ / <i>pour</i> ‘for’ (+ DET) + N + P ₂	9
Total	52

This pattern is illustrated in (22) with *por l’amor de* ‘thanks to’ (lit. ‘for the love of’), which also provides evidence of semantic bleaching in CAs. An alternate pattern can be found, much as CA in Modern Romanian: [PREP₁ (DET) N + CASE MARKING], with the object case on the governed noun, illustrated in (23) with [*en som* [*lo mur*]_{NP}]_{PP} lit. ‘[in/on top [the wall]_{NP}]_{PP}’.

(22) Tout **por** l’ **amor** de cest afaire li
 all for the.M.SG love of this.M.SG matter the.M.NOM.SG
 varlés au cuer debonaire a
 valet at.the.M.SG heart good-natured AUX.IND.PRS.3SG
 fait le poulain jus abat
 do.PST.PTCP.M.SG the.M.SG colt down put.INF
 ‘Because of this matter, the good-hearted valet had the colt put down’ (*Eracle*, 12th c., p. 58, v. 1883)

(23) Chascuns qui l’ ot corant
 each.M.SG.NOM who.NOM PRO.SG.ACC hear.PRS.IND.3SG run.PRS.PTCP
 i vet, d’ **en som lo mur**
 there go.PRS.IND.3SG, of on top the.M.SG.ACC wall
 lo vont veoir
 PRO.M.SG.ACC go.PRS.IND.3PL see.INF
 ‘Anyone who heard him ran up to him, they went to watch him from upon the wall’ (*Eneas*, 12th c., v. 7340-1)

Given the fact that there are only two cases in Old French (subject case and object case), the distinction between this second type and a true adpositional use ([PREP [NP]]_{PP}) is not obvious, and seems to rely mainly on the analysis of the central element as retaining nominal features or having undergone reanalysis. Thus, it seems that *de lez* ‘next to (lit. ‘of side’, ‘from side’, ‘from beside)’ has undergone full reanalysis and can be considered a ‘simple’ adposition, given the fact that it can be reinforced by a prefix, yielding the surprising form *dedelez*

‘next to’ (lit. ‘of of beside’ or ‘from from beside’). The productivity of the paradigm of constructions such as *en som* is hard to evaluate, depending mostly on the analysis retained for individual constructions.

5.2 Middle and Classical French

In Middle and Classical French (roughly from mid 13th c. through 18th c.), the productivity and frequency of the main pattern of CAs steadily increase. The importance of this phenomenon in Middle French is already noted by Shears (1922: 19)⁷. There is some renewal: not all constructions which appeared in Old French survive, and many new constructions appear, including for instance *au regard de* ‘with respect to’ (lit. ‘at the eye gaze of’). However, quite a few remain and gain frequency, for instance *au long de* ‘along’ (lit. ‘at the length of’) or *au-dessus de* ‘above’ (lit. ‘at the upper-side of’). As suggested in the chapter on CAs in Germanic languages (Hüning et al., this volume), and in line with construction grammar approaches (Goldberg 1995, Traugott and Trousdale 2013), it may be considered that Complex Adpositions may undergo constructionalization to some extent, losing variability and acquiring internal coherence. For instance, in the GGHF corpus, the sequence *au bord de* ‘at the edge of’ seems to appear in the 13th century, with a very low relative frequency in Middle and Pre-Classical French (roughly 4 occ. per million words). In Classical (and Modern) French, along with other similar sequences, it becomes much more frequent (37 occ. per million words) and thus reaches a frequency which is akin to that of simple (lexical) prepositions. At that period, for this construction, the proportion of occurrences with modification drops from 14% (2 in 14 occ.) to less than 1% (3 in 355 occ.).

The result is that, by our counts, in Middle French some 20 CAs have a frequency of about or more than 10 occurrences per million words, with more than 60 CAs in total; in Classical French, more than 30 CAs have a frequency above 10 occ. per million, with more than 70 CAs in total. The cumulative (relative) frequency of CAs seems to have reached 200 to 300 occurrences per million (0.2–0.3‰) in Middle French, and more than 600 per million (0.6‰) in Classical French. The rise in frequency continues after that: the cumulative frequency of CAs listed in Table 5 (Section 4.3, Modern French – 19th and 20th centuries) is roughly 1.3‰. This is still much lower than functional or even lexical adpositions

⁷ He thus writes that the tendency to substitute prepositional phrases for simple prepositions is one of the most remarkable features of Middle French (“l’un des caractères les plus remarquables du moyen français est la tendance marquée à substituer aux prépositions simples des locutions prépositives”; our translation).

during the same period; however, it shows the gradual emergence of the category, from a few infrequent constructions to a productive, consistent and relatively frequent pattern.

5.3 Modern French

In the later centuries, this pattern becomes much more productive, yielding hundreds of new constructions, some of which not only acquire a high relative frequency, but also tend to become lexicalized (see Table 7); when that happens, they tend to deviate from the typical (and original) pattern.

Table 7: Most productive patterns of CAs in Modern French.

pattern	examples
Prep ₁ + N + Prep ₂	<i>à cause de</i> ‘because of’, <i>à côté de</i> ‘next to’, <i>à défaut de</i> ‘for want of’, <i>à fleur de</i> ‘on the surface of’, <i>à force de</i> ‘by dint of’, <i>à raison de</i> ‘at the rate of’, <i>en face de</i> ‘in front of’, <i>de peur de</i> ‘for fear of’, <i>par rapport à</i> ‘with regard to’, <i>par suite de</i> ‘as a result of’, <i>pour cause de</i> ‘on account of’, <i>sous couleur de</i> ‘under the guise of’, <i>en faveur de</i> ‘in favor of’, <i>en guise de</i> ‘by way of’, <i>en raison de</i> ‘because of’, <i>en dépit de</i> ‘despite’
Prep ₁ + Adv + Prep ₂	<i>en dessous de</i> ‘below’, <i>en-dehors de</i> ‘outside, apart from’, <i>en sus de</i> ‘above’, <i>en amont de</i> ‘uphill from’, <i>en arrière de</i> ‘behind’, <i>en aval de</i> ‘downhill from’, <i>en bas de</i> ‘below’, <i>en deçà de</i> ‘short of’, <i>en dedans de</i> ‘within’, <i>en plus de</i> ‘besides’
Prep ₁ + Det + N + Prep ₂	<i>à l’aide de</i> ‘by means of’, <i>à l’égard de</i> ‘with respect to’, <i>à l’endroit de</i> ‘concerning’, <i>à l’entour de</i> ‘around’, <i>à l’exception de</i> ‘except’, <i>à l’exclusion de</i> ‘except’, <i>à l’instar de</i> ‘like’, <i>à l’issue de</i> ‘after’, <i>à l’occasion de</i> ‘on the occasion of’, <i>à la faveur de</i> ‘thanks to’, <i>à la merci de</i> ‘at the mercy of’, <i>à l’insu de</i> ‘unbeknownst to’, <i>à l’intérieur de</i> ‘inside’, <i>au bord de</i> ‘on the edge of’, <i>au fond de</i> ‘at the bottom of’, <i>au lieu de</i> ‘instead of’, <i>au milieu de</i> ‘in the middle of’, <i>au moyen de</i> ‘thanks to’, <i>au pied de</i> ‘at the foot of’, <i>au prix de</i> ‘by means of’, <i>au travers de</i> ‘by means of’, <i>du côté de</i> ‘near’, <i>le long de</i> ‘along’, <i>de la part de</i> ‘on behalf of’
Prep ₁ + Det + Adv + Prep ₂	<i>à l’arrière de</i> ‘behind’, <i>à l’encontre de</i> ‘counter to’, <i>au dedans de</i> ‘within’, <i>au dehors de</i> ‘outside’, <i>au delà de</i> ‘beyond’, <i>au dessous de</i> ‘below’, <i>au dessus de</i> ‘above’, <i>au devant de</i> ‘in front of’, <i>auprès de</i> ‘near’

Their relative frequency is much higher than it was in Medieval French. While the most frequent CAs in Medieval French barely reached 20 occ. per million (*au lieu de*, BFM), they can be at least three times as frequent in Modern French, as illustrated in Table 8.

Some of the older CAs grammaticalize into simple adpositions. In some cases, this simplification is found even in Standard French, and the corresponding

Table 8: Relative frequency of a few CAs, in Old and Modern French.

corpus	sequence	relative frequency (occ. per million)
BFM (11th–15th c.: 3 million words)	<i>au lieu de</i> ‘instead of’	23.3
Frantext (after 1968: 54.2 million words)	<i>au moyen de</i> ‘by means of’	6.2
	<i>au lieu de</i> ‘instead of’	68.2
	<i>à travers</i> ‘through’	133.8
	<i>à l’instar de</i> ‘like’	5.3
	<i>faute de</i> ‘for lack of’	19.2

constructions have an even higher frequency, as is the case for *à travers* ‘through’, or *par (la) faute de / à faute de > faute de* (Table 8). In other cases, this simplification is limited to some contexts, i.e. mainly to informal or regional French, for instance *en face de > en face, du côté de > du côté > côté* (Fagard 2008), *par rapport à > rapport à* (24–25).

- (24) c’ est-il vrai, monsieur Joigneau, ce
 this.M.SG be.PRS.IND.3SG.IT true sir Joigneau this.M.SG
 qu’ on raconte, **rapport à** la
 which.ACC.SG one say.PRS.IND.3SG about at the.F.SG
 mère Daigne?
 mother Daigne
 ‘is it true, Mr. Joigneau, what they say ’bout Mother Daigne?’ (Roger Martin du Gard, *Vieille France*, 1933, p. 1030)

- (25) Je me souviens très bien maintenant. Mon
 I me.ACC.SG remind.PRS.IND.1SG very well now my
 taxi faillit vous écraser, il y
 cab fail.PST.IND.3SG you.ACC.SG run-over.INF it there
 a une quinzaine de jours, **en face** la
 have.PRS.IND.3SG one.F.SG fortnight of day.PL in face the.F.SG
 gare du Nord (...)
 station of.the.M.SG North
 ‘I remember perfectly now. My cab almost ran you down, two weeks ago, in front of the Gare du Nord’ (Raymond Queneau, *Le Chiendent*, 1933, p. 43)

6 Variation in the language family

In this section, we provide an overview of CAs in Romance languages, focusing first on common features (6.1), then on variation (6.2): as we will see, there are both.

6.1 Common features

There seems to be a set of core patterns of CAs which are found in all Romance languages. As illustrated by Table 9, the most wide-spread patterns besides the main pattern ([PREP₁ (DET) N PREP₂]) are [PREP₁ (DET) ADV PREP₂] and [N/ADV PREP₂]. Among these 6 patterns (3*2 possibilities: PREP₁ / zero, DET / zero, N/ADV), it could be argued that the main patterns are actually [PREP₁ N PREP₂], [PREP₁ ADV PREP₂], [ADV PREP₂] and [PREP₁ DET N PREP₂], since the [N PREP₂] pattern is less frequent, generally resulting from the loss of the initial P, while in the [PREP₁ DET ADV PREP₂] pattern the adverb is nominalized, and could or should be analyzed as a noun (the same could be said of all complex adpositions in which the lexical nucleus is not originally – or not considered to be – a noun, but takes the definite article). Finally, there are very few ‘other’ patterns. Note that most of these patterns are already found in Medieval Romance, in all languages for which medieval texts are available (Fagard 2006).

Table 9: Patterns of Complex Adpositions in Romance languages (C: Catalan, I: Italian, O: Occitan, P: Portuguese, R: Romanian, S: Spanish).

		Pattern			
		[PREP ₁ N/ADV PREP ₂]	[N/ADV PREP ₂]	[PREP ₁ DET N/ADV PREP ₂]	other
LEXICAL NUCLEUS	NOUN	all Romance languages		FCI?OPRS	FP
	ADVERB			?(FCOR)	FCPS
	VERB	FCOPR	FCPR	C	FCPS
	OTHER	CP	IP?R	C	?

The four main patterns, PNP, PAdvP, AdvP and PDetNP, also seem to be the most productive and those which yield the most frequent complex adpositions. There is also a striking lexical homogeneity: for quite a few CAs, patterns are instantiated almost identically across languages. Indeed, in many cases, the very same noun, adverb or verb form is used as nucleus of a CA *and* follows similar patterns in different Romance languages, as illustrated in Table 10.

Table 10: A few Complex Adpositions in Romance – similarities and differences.

Lexical Nucleus	Portuguese	Spanish	Catalan	Occitan	French	Italian	Romanian
'base'	<i>na base de, com base em, à base de</i>	<i>a base de</i>	<i>a base de</i>	<i>a basa de</i>	<i>à base de</i>	<i>a base di, sulla base di</i>	<i>pe baza (+ GEN)</i>
'cause'	<i>por causa de</i>	<i>por causa de</i>	<i>a causa de</i>	<i>a causa de</i>	<i>à cause de</i>	<i>per causa di</i>	<i>din cauza (+ GEN)</i>
'outside'	<i>fora de</i>	<i>fuera de</i>	<i>fora de, en/al defora de</i>	<i>en/al defòra de</i>	<i>en-dehors de</i>	<i>all'infuori di</i>	<i>în afară de</i>

These three lexical nuclei illustrate quite well on the one hand the proximity between languages, since the same lexical nucleus appears in all 7, but also the possible variations, for instance the absence or presence of Prep₁, with Ibero-Romance standing out, or the use of the genitive instead of Prep₂ in Romanian. In the next subsection, we further illustrate these sources of cross-linguistic variation.

6.2 Variation

There is indeed a great deal of variation across Romance languages. Variation can be found in the *lexical* elements which form the core of CAs, and which vary along with the rest of the lexicon; it is also present in the respective proportions of [PREP₁ (DET) N PREP₂] and [ADV P] patterns; finally, the degree of grammaticalization of CAs, as measured by the way they react to syntactic tests, seems uneven. We address these three types of variation in sequence.

6.2.1 Lexical variation

Some lexical nuclei are less wide-spread than others, on account of grammaticalizations which took place after the differentiation of Romance. For instance, Late Latin **bassius* 'lower than' acquired adverbial (and later on prepositional) uses only in parts of Romania, with Portuguese *baixo*, Spanish *bajo*, French and Occitan *bas* 'low' > 'down' in Ibero-Romance and Gallo-Romance, respectively. The same can be said of the grammaticalization of Latin **monte* and **valle*, with corresponding adpositions mainly in Gallo-Romance: French *en amont de, en*

aval de, Catalan *al damunt de*, *davall de*, Occitan *daval de*, **en amont de* (Portuguese equivalents, *a montante de* ‘upstream’ and *a jusante de* ‘downstream’, seem to be a later creation). More locally, individual languages may stand out with the grammaticalization of a new CA, for instance Portuguese *à beira de* ‘(very) close to, on the verge of’ on the noun *beira* (probably from *ribeira* < (Lat) *ripariam*, on *riparius* ‘which frequents the banks of rivers’), or Catalan *a la vora de* and French *à l’orée de* with the same meaning, but a different etymology: (Lt) *ora* ‘border, edge’, plural of *os*, *oris*, n. ‘mouth’.

6.2.2 Patterns

The main differences as far as patterns are concerned are (i) the importance of the AdvP pattern in Ibero-Romance, (ii) the dialectal variation in Italo-Romance and the use of *di* for personal pronouns: Italian *senza di me* ‘without (lit. of) me’ / *senza questo* ‘without this’, and finally (iii) the complex system of CAs in Romanian, with an alternation between Prep₂ and genitive marking.

The importance of the AdvP pattern in Ibero-Romance is quite obvious (Fagard 2006, 2010); they are described in Spanish and Portuguese as “relational adverbs” (*advérbios relacionais*, Raposo et al. 2013: 1500), on account of their distributional characteristics, i.e. mainly their capacity to appear either alone (in the form [Adv]), or with a complement ([AdvP [SN]]), as in (26–27).

(26) o comboio já está perto (Portuguese)
 DET.M.SG train already be.PRS.IND.3SG near
 da estação
 of.DET.F.SG train.station
 ‘the train is already near the train station’ (Raposo et al. 2013, *ibid.*)

(27) o comboio já está perto
 DET.M.SG train already be.PRS.IND.3SG near
 ‘the train is already nearby’

Even if the pattern is found in other languages, there is a much more systematic correspondence between adverbial and adpositional uses in Ibero-Romance than in other Romance languages. Table 11 illustrates the systematicity of the Adv/CA pattern in Portuguese, and the fact that both Italian and French differ in this respect. While there is a perfect systematicity in Portuguese, for these examples at least, it is not the case in French and Italian. Indeed, in most instances there is either no variation (e.g. for ‘after’, in both languages) or various types of

Table 11: Adv and AdvP/CA uses in Portuguese, Italian and French.

Meaning	Adverbial uses			Adpositional uses		
	Portuguese	Italian	French	Portuguese	Italian	French
'above'	<i>acima</i>	<i>sopra</i>	<i>dessus</i>	<i>Adv + de</i>	<i>Adv (+ a/di)</i>	CA
'below'	<i>debaixo</i>	<i>sotto</i>	<i>dessous</i>			(<i>au-dessus de</i> , <i>en-dessous de</i>)
'far'	<i>longe</i>	<i>lontano</i>	<i>loin</i>		<i>Adv + di</i>	<i>Adv + de</i>
'before'	<i>antes</i>	<i>prima</i>	<i>avant</i>			ADV = PREP
'near'	<i>perto</i>	<i>vicino</i>	<i>près</i>		<i>Adv (+ a)</i>	
'after'	<i>depois</i>	<i>dopo</i>	<i>après</i>		ADV = PREP	

Note that French *près (de)* 'close (to)', mainly has uses as a CP. It also has adverbial uses; these are generally found in specific constructions such as *tout près* 'very close', *assez près* 'quite close', but there are examples of adverbial *près* alone (e.g. *C'est passé près* ! 'That was close!' – lit. *that went close by*, of bullets –, Cavanna, 1979, *Les Russkoffs*, Frantext).

alternation between adverbial and prepositional uses (e.g. with an optional Prep₂ as in Italian, or a CA construction as in French).

This table also illustrates the variation in Italian, with the same form displaying uses as an adverb or an adposition, but also CA constructions, e.g. *sopra (a/di) + NP*, for which a corpus search (itTenTen and Repubblica, Sketch-Engine) shows the existence of various possibilities: *sopra di + pronoun* (28), *sopra a + Adv/NP* (29), *al di sopra di* (30), *sopra + NP* (31) (with widely different frequencies: *sopra* is above 20 occurrences per million for prepositional uses only, while all other constructions are well below 5 occurrences per million). This variation is partly constrained by dialectal, sociolinguistic or register-related considerations.⁸

- (28) **Sopra di** noi torreggiano i palazzi. (Italian)
 above of us tower.PRS.IND.3PL the.M.PL building.PL
 'The skyscrapers loom above us' (Repubblica corpus, SketchEngine, #26298006)

⁸ This is also true for other Romance languages, of course, with e.g. the alternation between French *dessous la table* and *sous/en-dessous de la table*, or *en face la gare* and *en face de la gare* (the first variant being clearly standard, in both cases). But it is quite frequent in Italian, possibly for historical reasons, the standard language having emerged only recently in comparison with other (national) Romance languages.

- (29) La casa è bassa, un solo piano
 the.F.SG house be.PRS.IND.3SG low.F.SG one.M.SG only.M.SG floor
sopra a quello terreno
 above at that.M.SG ground
 ‘The house is low, only one floor above the ground floor’ (Repubblica corpus, SketchEngine, #30558684)
- (30) Un testimone **al di sopra di** ogni sospetto
 a witness at.the.M.SG of above of all.M.SG suspicion
 ‘A witness above suspicion’ (Repubblica corpus, SketchEngine, #873531)
- (31) **Sopra** la sua voce si sono sentite
 above the.F.SG POSS.F.SG voice REFL be.PRS.IND.3PL hear.PST.PART.F.PL
 distinte quelle di
 distinct.F.PL this.F.PL of
 due signore che si raccontavano i fatti
 two lady.PL REL.PRO REFL tell.PST.IND.3PL the.M.PL fact.PL
 loro.
 POSS.PL
 ‘Above his voice one could hear distinctly those of two ladies who were telling each other their own stories’ (Repubblica corpus, SketchEngine, #21425897)

Romanian naturally stands out on account of its case-marking system, which allows for a pattern found in no other Modern Romance language: the [PREP₁ N + GEN] pattern, in which the link with the complement is marked not by an adposition, but by case marking (Găitănaru 1999, Petrescu 2005, Vlasin 2015). The [PREP₁ N PREP₂] pattern is also found (Ciobanu 1957, Căpățână 2003), and in some cases both constructions are possible for a given CA. This may entail a difference in meaning, as can be seen in examples (32–35) for *în afara* + GEN / *în afară de* ‘outside of, except’ and *în jurul* + GEN / *în jur de* ‘around’.

- (32) *inclusiv în afara Uniunii Europene* (Romanian)
 even in outside.DET.F.SG Union.GEN European.GEN
 ‘even outside the European Union’ (roTenTen16)
- (33) *Ce altceva mai sculptați, în afară de portrete?*
 what else still sculpt.IND.PRS.2PL in outside of portrait.PL
 ‘what else do you make, besides portraits?’ (roTenTen16)

- (34) *Protestul va dura în jur de două ore.*
 protest.DET will.PRS.3SG last.INF in turn of two.F.PL hour.PL
 ‘the protest will last around two hours’ (roTenTen16)
- (35) 20 aprilie 2016, *în jurul* orei 06.00
 20 april 2016 in turn.DET hour.GEN 06:00
 ‘On April 20th, 2016, at around 6 am’ (roTenTen16)

Most frequently, though, CAs follow only one pattern. The [PREP₁ N + GEN] construction is clearly the most productive one, with close to a hundred items, for instance *cu excepția* ‘except’, *din cauza* ‘because of’, *în ciuda* ‘despite’. The [PREP₁ N PREP₂] pattern is less productive, with less than thirty items, among which *în (de)curs de* ‘during’, *în jur de* ‘around’, *în urmă cu* ‘ago’ (Fagard and Mardale 2019).

6.2.3 Degree of grammaticalization

All languages, it seems, evolve in part through grammaticalization processes. These result in a progressive renewal of grammatical systems. Thus, in the evolution from Latin to Modern Romance languages, there has been a steady renewal of grammatical items, including TAM markers, demonstratives, pronouns, and even the emergence of new paradigms of grammatical items, for instance determiners. In this respect, it has been argued that, among Romance languages, French is the furthest along the grammaticalization cline (Lamiroy 1999, De Mulder 2001, Fagard 2010, 2011). It has for instance developed the most extensive paradigm of determiners, including indefinite plural and partitive uses (Carlier 2007). This could also be said of the adpositional system:

French grammaticalized further its renewed stock, creating more simple adpositions than the other languages and more rigid patterns of complex adpositions, followed by Italian, then Spanish, Portuguese and Romanian. (Fagard and Mardale 2012: 335)

Of course, the tendency of CAs to grammaticalize and thus form less complex – and eventually simple – adpositions is probably universal, and examples could be found easily in the other Romance languages, e.g. Catalan *a la vora de* > *vora de* > *vora* ‘close to’, Italian *per tramite di* ‘by way of’ > *tramite* ‘through’. However, French CAs, in comparison with other Romance CAs, seem to have both emerged more consistently as a class, and – possibly as a consequence – to have contributed more constructions available to further grammaticalization into simple adpositions. One possible indication, among others (see e.g. Stosic and Fagard 2019, Hüning et al., this volume, Stosic, this volume), of the greater degree of grammaticalization of

CAs in French – more specifically of their greater internal coherence (Lehmann 2002) – is the possible alternation with the possessive, which is almost never possible in French once the CA is grammaticalized, but often remains possible in other Romance languages, as illustrated by examples (36–39)⁹. As noted in the *Nueva gramática de la lengua española* (2009, I:55), for some scholars, the possibility of using the possessive automatically excludes a sequence from the group of CAs; however, the authors (*ibid.*) plead for a more flexible approach, considering that these sequences may display ‘semilexicalization’ (*semilexicalización*). The possibility of modifying such sequences may evolve with time: for instance, in Medieval and Classical French, insertion was still possible even with demonstrative determiners, as illustrated in (40) – which would be translated in Modern French by *à cause de cela* ‘because of that’, while *#à cette cause* can no longer alternate with an adpositional use of *à cause de* (*j’ai raté mon examen à cause de la fête* ‘I failed my exam because of the party’ → **à cette cause* ‘for this reason’).

(36) (French) **en-dessous de lui** ‘below him’ → ***en son dessous**

(37) (Portuguese) **ao seu lado / ao lado dele** ‘next to him’, **a teu respeito** ‘about you’ / ***a respeito de ti, na sua esteira / na esteira dele** ‘in his wake’

(38) (Italian) **per mia colpa** ‘because of me’ / ***per colpa di me**

(39) (Catalan) **davall de mi / davall meu** ‘beneath me’, **davall nostre** ‘below us’, **sobre meu** ‘above me’

(40) en quoy ne se peürent (Medieval French)
 in REL.PRO.ACC.SG NEG REFL can.PST.IND.3SG
 accorder et, à ceste cause,
 agree.INF and at this.F.SG cause
 partit très mal content d’ eulx de
 leave.PST.IND.3SG very bad.M.SG happy of they.3PL.ACC of
 la court
 the.F.SG court
 ‘on which question they could not find an agreement, and, for that reason [lit. *at this cause*], he left the court feeling very mad at them’ (Philippe de Comynes, *Mémoires*, 15th c., vol. 7, p.7)

⁹ In general, Romance languages display individual exceptions to this tendency; for instance, in Portuguese, the grammaticalization of *a pesar de* ‘in spite of’ > *apesar de* ‘despite’ led to the loss of the alternate construction with the possessive (initially *a seu pesar* ‘in his spite’) (Lima 2019b).

A thorough contrastive study should check whether this is indeed the case, i.e. whether there is a clear contrast between French and other Romance languages.

7 Conclusion

Systems of CAs are fairly well-established in Romance languages, with a few productive patterns providing constructions with high frequencies, which sometimes end up grammaticalizing into simple adpositions. There are CAs in all Romance languages, and they appear to be quite homogeneous, with some variation, however, in terms of lexical nuclei and patterns. There seem to be areal tendencies, with a greater importance of the AdvP pattern in Ibero-Romance, of the P(Det)NP patterns in French, of variation in Italian, and the existence of specific patterns in Romanian.

There is evidence of the emergence of CAs as a class between Medieval and Modern Romance. Indeed, the data available for French and other Romance languages point to the existence of individual constructions in Medieval Romance. It is also quite clear that CA patterns were much more limited in number, frequency and systematicity in Medieval Romance than they are now. After that, there has been a steady rise of CAs. This can be documented, for instance, from Old French to Modern French: more constructions appear, gaining in frequency and internal coherence, and progressively yielding systematic patterns which become available for the creation of new CAs (*grammaticalization by analogy*, see Meillet 1912, Hoffmann 2004). It remains to be seen whether there has been, as hypothesized by Van der Horst (2013, cf. Hüning 2014¹⁰), a global emergence of CAs in the languages of Europe in the 19th c. – one feature among others indicating a Sapirian *drift* of central average European, or more exactly western European languages, as a result of a more general *cultural drift* (ibid.: 182). If it is the case, we should try to understand whether this phenomenon is linked to cultural trends, to language contact, or to typological shifts (from synthetic Latin to analytic Romance).

¹⁰ Van der Horst actually shows that these strong relations between European languages are only a few centuries old (“Ze zijn er niet altijd geweest, die vaste verbindingen. Duizend jaar geleden, in de periode van het Oudnederlands, waren ze er nog niet of nauwelijks. Het is pas in de laatste paar eeuwen dat ze opduiken en de wind mee krijgen. Maar dan gaat het ook hard.” “These fixed connections have not always existed. A thousand years ago, during the period of Old Dutch, they did not exist, or barely. It is only in the last two centuries that they appeared and gained strength. But they took off quickly” [our translation]; Van der Horst 2013: 131, quoted in Hüning 2014: 439).

Abbreviations

ABL	ablative	M	masculine
ACC	accusative	NEG	negation
ADV	adverb	NOM	nominative
AUX	auxiliary	PL	plural
COMP	complementizer	POSS	possessive
DAT	dative	PRO	pronoun
DEF	definite	PRS	present
DET	determiner	PTCP	participle
DOM	differential object marking	REFL	reflexive
F	feminine	REL	relative
GEN	genitive	SBJ	subjunctive
IND	indicative	SG	singular
INDF	indefinite	VOC	vocative
INF	infinitive		

References

- Abeillé, Anne, Danièle Godard, & Annie Delaveau. In press. *Grande Grammaire du français*. Paris: Actes Sud.
- Adler, Silvia. 2001. Les locutions prépositives: questions de méthodologie et de définition, *Travaux de linguistique* 42–43(1). 157–170.
- Adler, Silvia. 2008. French compound prepositions, prepositional locutions and prepositional phrases in the scope of the absolute use. In Dennis Kurzon & Silvia Adler (eds.), *Adpositions. Pragmatic, semantic and syntactic perspectives*, 17–35. Amsterdam & Philadelphia: John Benjamins.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, & Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Limited.
- Borillo, Andrée. 1997. Aide à l'identification des prépositions complexes de temps et de lieu. *Faits de langue* 9. 173–184.
- Borillo, Andrée. 2000. Degrés de grammaticalisation des prépositions composées de lieu. *Actes du CERLICO, La grammaticalisation: (dé)motivation et contrainte*. Rennes: Presses universitaires de Rennes.
- Borillo, Andrée. 2002. Il y a prépositions et prépositions. *Travaux de Linguistique* 42–43. 141–155.
- Bybee, Joan, & Paul Hopper (eds.). 2001. *Frequency and the emergence of linguistic structure*. Amsterdam & Philadelphia: John Benjamins.
- Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Căpățână, Cecilia. 2003. Locuțiunile prepoziționale [Complex prepositions]. *Limba Română* [Romanian language] XIII(1). 48–53.
- Cifuentes Honrubia, José Luís. 2003. *Locuciones prepositivas – sobre la gramaticalización preposicional en español*. Alicante: Universidad de Alicante.

- Ciobanu, Fulvia. 1957. Observații asupra prepozițiilor și locuțiunilor prepoziționale din limba română. *Studii de gramatică* [Grammar studies] 2. 357–388.
- Diewald, Gabriele, & Marijana Kresić. 2010. Ein übereinzelsprachliches kontrastives Beschreibungsmodell für Partikelbedeutungen. *Linguistik online* 44. http://www.linguistikonline.de/44_10/diewaldKresic_a.html (accessed 2 July 2013).
- Fagard, Benjamin. To appear. Prépositions et locutions prépositives. In Christiane Marchello-Nizia, Bernard Combettes, Sophie Prévost, & Tobias Scheer (eds.), *Grande grammaire historique du français*. Berlin & New York: De Gruyter.
- Fagard, Benjamin. 2009. Prépositions et locutions prépositionnelles : un sémantisme comparable ? *Langages* 173. 95–113.
- Fagard, Benjamin. 2012. Prépositions et locutions prépositionnelles : la question du renouvellement grammatical. *Travaux de linguistique* 64(1). 161–189.
- Fagard, Benjamin, & Alexandru Mardale. 2007. Systèmes prépositionnels des langues romanes : la notion de partie du discours en diachronie. In David Trotter (ed.), *Actes du XXIVe Congrès International de Linguistique et de Philologie Romanes*, 91–104. Tübingen: Niemeyer.
- Fagard, Benjamin, & Alexandru Mardale. 2012. The pace of grammaticalization and the evolution of prepositional systems: Data from Romance. *Folia Linguistica* 46(2). 303–340.
- Fagard, Benjamin, & Alexandru Mardale. 2019. Les prépositions complexes du roumain – approche sur corpus, *Revue Romane* 54(1). 176–203 (10.1075/rro.00020.fag).
- Fagard, Benjamin, Paola Pietrandrea, & Julie Glikman. 2016. Syntactic and semantic aspects of Romance Complementizers. In Kasper Boye & Petar Kehayov (eds.), *Complementizers in European Languages*, 75–130. Berlin & New York: De Gruyter.
- Fagard, Benjamin, & Walter De Mulder. 2007. La formation des prépositions complexes : grammaticalisation ou lexicalisation ? *Langue française* 156. 9–29.
- Găitănaru, Ștefan. 1999. [Criteria for distinguishing complex prepositions with the genitive from free, analyzable sequences]. *Limba și literatura română* [Romanian language and literature] XXVIII(2). 12–14.
- Hoffmann, Sebastian. 2005. *Grammaticalization and English Complex Prepositions. A corpus-based analysis*. London & New York: Routledge.
- Horst, Joop van der. 2013. *Taal op drift: lange-termijnontwikkelingen in taal en samenleving*. Amsterdam: Meulenhoff.
- Hüning, Matthias, Elena Smirnova, & Anatol Stefanowitsch. This volume. Complex adpositions in three West Germanic Languages: German, Dutch, and English.
- Hüning, Matthias. 2014. Over complexe preposities en convergentie. [About complex prepositions and convergence]. In Freek Van de Velde, Hans Smessaert, Frank Van Eynde, & Sara Verbrugge (eds.), *Patroon en argument. Een dubbelfeestbundel bij het emeritaat van William Van Belle en Joop van der Horst* [Pattern and argument. A double festschrift on the occasion of the retirements of William Van Belle and Joop van der Horst], 433–445. Leuven: Universitaire Pers Leuven.
- Koch, Peter. 2002. Das Französische in typologischer und historisch-vergleichender Sicht. In Ingo Kolboom, Thomas Kotschi, & Edward Reichel (eds.), *Handbuch Französisch*, 36–58. Berlin: Schmidt.
- Kurzon, Dennis, & Silvia Adler (eds.). 2008. *Adpositions. Pragmatic, semantic and syntactic perspectives*. Amsterdam & Philadelphia: John Benjamins.
- Leeman, Danielle. 2007. De la préposition à la locution prépositionnelle. *Modèles linguistiques* 55, 715.

- Lima, José P. 2014. Grammaticalization of complex prepositions in European Portuguese. In José P. Lima, *Studies on Grammaticalization and Lexicalization*, 163–183. Munich: Lincom Europa
- Lima, José P. 2019a. On grammaticalized complex prepositions in Portuguese: Deployment, shift, redundancy, complementation. In Benjamin Fagard, José P. Lima, & Dejan Stosic (eds.) 2019. *Les prépositions complexes dans les langues romanes*. Special issue of *Revue Romane* 54(1). Amsterdam & Philadelphia: John Benjamins, 126–140.
- Lima, José P. 2019b. Preposições complexas em Português [Complex prepositions in Portuguese]. In Ernestina Carrilho, Ana Martins, Sandra Pereira, & João P. Silvestre (eds.), *Estudos linguísticos e filológicos oferecidos a Ivo Castro* [Linguistic and philological studies offered to Ivo Castro], 1181–1211. Lisboa: Centro de Linguística da Universidade de Lisboa. [<https://repositorio.ul.pt/handle/10451/39619>]
- Marchello-Nizia, Christiane Bernard Combettes, Sophie Prévost, & Tobias Scheer (eds.). In prep. *Grande grammaire historique du français*. Berlin & New York: De Gruyter.
- Meillet, Antoine. 1912. L'évolution des formes grammaticales. *Scientia, Rivista di scienza* 12(26). (reprinted in Antoine Meillet. 1958. *Linguistique Historique et Linguistique générale*, Paris: Champion, 130–149).
- Melis, Ludo. 2003. *La préposition en français*. Paris: Ophrys.
- Petrescu, Maria. 2005. Despre locuțiunile prepoziționale cu genitivul. *Annales Universitatis Apulensis. Series Philologica*, 6.3.
- Piunno, Valentina, & Vittorio Ganfi. 2019. Usage-based account of Italian Complex Prepositions denoting the Agent. In Benjamin Fagard, José P. Lima, & Dejan Stosic (eds.). *Les prépositions complexes dans les langues romanes*. Special issue of *Revue Romane* 54(1), Amsterdam & Philadelphia: John Benjamins, 141 – 175.
- Raposo, Eduardo P., Maria F. Nascimento, Maria A. Mota, Luísa Segura, & Amália Mendes. 2013. *Gramática do Português* [Portuguese grammar]. Vols. I & II. Lisboa: Fundação Calouste Gulbenkian.
- Schoonjans, Steven. 2013. Modal particles: Problems in defining a category. In Liesbeth Degand, Bert Cornillie, & Paola Pietrandrea (eds.), *Modal Particles and Discourse Markers. Categorization and Description*, 133–162. Amsterdam & Philadelphia: John Benjamins.
- Shears, Fred. 1922. *Recherches sur les Prépositions dans la Prose du Moyen Français (XIV^e et XV^e siècles)*. Paris: Honoré Champion.
- Sibille, Jean. 2019. Les prépositions complexes en occitan languedocien: premières explorations. In Benjamin Fagard, José P. Lima, & Dejan Stosic (eds.). *Les prépositions complexes dans les langues romanes*. Special issue of *Revue Romane* 54(1), Amsterdam & Philadelphia: John Benjamins, 62–92.
- Squartini, Mario. 2013. From TAM to discourse: The role of information status in North-Western Italian *già* 'already'. In Liesbeth Degand, Bert Cornillie, & Paola Pietrandrea (eds.), *Modal Particles and Discourse Markers. Categorization and Description*, 163–190. Amsterdam & Philadelphia: John Benjamins.
- Stosic, Dejan. 2012. *En passant par* : une expression en voie de grammaticalisation?, *CORELA / Special issue / Langue, espace, cognition*. (<http://corela.edel.univ-poitiers.fr/index.php?id=2844>)
- Stosic, Dejan. This volume. Complex adpositions in Slavic languages: the case of Serbian.
- Stosic, Dejan. To appear. Les prépositions complexes du français. *Carnets de grammaire*.
- Stosic, Dejan, & Benjamin Fagard. (2019). Les prépositions complexes en français: Pour une méthode d'identification multicritère. In Benjamin Fagard, José P. Lima, & Dejan Stosic

(eds.). *Les prépositions complexes dans les langues romanes*. Special issue of *Revue Romane* 54(1), Amsterdam & Philadelphia: John Benjamins, 8–38.

- Vlasin, Veronica. 2015. Câteva considerații asupra statutului morfologic al locuțiunilor prepoziționale cu genitivul [A few considerations about the morphological status of complex prepositions with the genitive]. In Ionuț Pomian & Nicolae Mocanu (eds.), *Înspre și dinspre Cluj, contribuții lingvistice: omagiu profesorului G. G. Neamțu la 70 de ani* [To and from Cluj, linguistic contributions: A tribute to professor G. G. Neamțu on the occasion of his 70th birthday]. Editura Scriptor, Cluj-Napoca, 645–657.

Corpora

- French: Valibel, PFC, Clapi, Frantext, frTenTen; Italian: Badip, itTenTen;
Portuguese: Corpus do Português (Davies & Ferreira 2006), ptTenTen11, CRPC (Corpus de referência do português contemporâneo), CETEM-Público;
Romanian: CORV (Dascălu-Jinga 2002), Ruxândoiu, Romanian web corpus; Spanish: CREA, CORDE, Corpus del Español (Davies 2002), esTenTen.

Grammars and dictionaries (selection)

- Anglade, Joseph. 1921. *Grammaire de l'ancien provençal*. Paris: Klincksieck.
- Avram, Mioara. 1960. *Evoluția subordonării circumstanțiale cu elemente conjuncționale în limba română* [Evolution of circumstantial clauses with conjunctive elements in Romanian]. Bucharest: Editura Academiei Republicii Populare Române.
- Badia Margarit, Antoni. 1962. *Gramàtica catalana*. Madrid: Gredos.
- Blumenthal, Peter, & Achim Stein. 2002. *Tobler-Lommatzsch, Elektronische Ausgabe*. Stuttgart: Steiner.
- Buridant, Claude. 2000. *Grammaire nouvelle de l'ancien français*. Paris: Sedes.
- Coutinho, Ismael. 1977. *Gramática histórica* [Historical grammar]. Rio de Janeiro: Ao Livro Técnico.
- De Mauro, Tullio, Federico Mancini, Massimo Vedovelli, & Miriam Voghera. 1993. *Lessico di frequenza dell'italiano parlato*. Milano: Etaslibri.
- De Rochegude, Henri Pascal. 1819. *Essai d'un glossaire occitanien pour servir à l'intelligence des poésies des troubadours*. Toulouse: Benichet Cadet.
- Densusianu, Ovid. 1938. *Histoire de la langue roumaine*. Paris: Leroux.
- Diccionario Medieval Español*. 1986. Salamanca: Universidad Pontificia de Salamanca.
- Dicționarul explicativ al limbii române* [Explanatory dictionary of Romanian](DEX). 1975. Bucharest: Academia Republicii Socialiste România.
- Ernout, Alfred, & François Thomas. 1964. *Syntaxe latine*. Paris: Klincksieck.
- Etienne, Eugène. 1980. *Essai de grammaire de l'ancien français (IX^e–XIV^e siècles)*. Genève: Slatkine Reprints. [Original edition: 1895, Nancy, Paris: Berger-Levrault].
- Euromosaic – The production and reproduction of the minority language groups in the European Union. 1996. Luxembourg: Office for Official Publications of the European Communities. [<http://www.lavplu.eu/central/bibliografie/euromosaic1.pdf>]

- Fabra, Pompeu. 2008 [1933], *Gramàtica catalana* [Catalan Grammar]. Barcelona: Institut d'Estudis Catalans.
- Gougenheim, Georges. 1974. *Grammaire de la langue française du seizième siècle*. Paris: Picard.
- Gramatica Limbii Române* [Romanian Grammar]. 2005. Bucharest: Editura Academiei Române.
- Jensen, Frede. 1994. *Syntaxe de l'ancien occitan*. Tübingen: Max Niemeyer.
- Migliorini, Bruno. 1966. *Storia della lingua italiana*. Firenze: Sansoni.
- Moignet, Gérard. [1973] 1984. *Grammaire de l'ancien français*, 2nd edn. Paris: Klincksieck.
- Moll, Francesc de Borja. 1952. *Gramática històrica catalana*. Madrid: Gredos.
- Musso, Giancarlo. 2004. *Gramática astèsan-a*. Torino: Gioventura Piemontèisa.
- Penny, Ralph J. [1993] 2014. *Gramática històrica del español*. Barcelona: Ariel Lingüística.
- Raposo, Eduardo P., Maria F. Nascimento, Maria A. Mota, Luísa Segura, & Amália Mendes. 2013. *Gramática do Português*. Vols. I & II. Lisboa: Fundação Calouste Gulbenkian.
- Real Academia Española y Asociación de Academias de la Lengua Española. 2009. *Nueva gramática de la lengua española*. Madrid: Espasa.
- Rohlf, Gerhard. [1949] 1954. *Grammatica storica della lingua italiana e dei suoi dialetti*. Torino: Einaudi [*Historische Grammatik der Italienischen Sprache und ihrer Mundarten*. Bern: Francke].
- Tekavčić, Pavao. 1972. *Grammatica storica dell'italiano. II: Morfosintassi*. Bologna: Il Mulino.
- Väänänen, Veikko. 1981 [1963]. *Introduction au latin vulgaire*. Paris: Klincksieck.
- Vocabolario degli accademici della Crusca*. 1866. Accademia della Crusca.
- Wolff, Alexandre (ed.). 2014. *La langue française dans le monde*. Paris: Nathan.

Anatol Stefanowitsch, Elena Smirnova, and Matthias Hüning

2 Complex adpositions in three West Germanic Languages: German, Dutch, and English

Abstract: In this chapter, we provide an overview of simple and complex adpositions in West Germanic languages, focusing in particular on German, Dutch, and English. We show that the class of prepositions is not a closed class in these languages, but that there are several ways in which it is productively extended. The focus of this chapter is on structural and diachronic aspects of the highly-productive [P (Det) N P]-constructions (i.e. the structural pattern ‘preposition – (determiner) – noun – preposition’), for example, Nl. *met betrekking tot*, G. *in Bezug auf*, and Eng. *with regard to*.

Keywords: German, Dutch, English, complex prepositions, morphological constructions

1 Introduction

In traditional descriptions of the major West Germanic languages, adpositions have more or less tacitly been assumed to constitute a closed (or nearly closed) class, in line with their characterization as serving a primarily grammatical function. Modern approaches, both descriptive and theoretical, are more varied. In English and Dutch linguistics, the closed-class nature of adpositions is still widely taken for granted. For example, for Baker (2003: 304), it is an “obvious fact”, not just synchronically but also diachronically: “languages do not have many adpositions and are reluctant to acquire more”. Major descriptive grammars of English (cf. Quirk et al. 1985: 71ff., Huddleston and Pullum 2002: 603) and Dutch (Broekhuis 2013: 5) tend to agree.

Anatol Stefanowitsch, Freie Universität Berlin, FB Philosophie und Geisteswissenschaften, Institut für Englische Philologie, Berlin. anatol.stefanowitsch@fu-berlin.de

Elena Smirnova, Institut de langue et littérature allemandes, Université de Neuchâtel, Neuchâtel. elena.smirnova@unine.ch

Matthias Hüning, Freie Universität Berlin, FB Philosophie und Geisteswissenschaften, Institut für Deutsche und Niederländische Philologie, Berlin. matthias.huening@fu-berlin.de

<https://doi.org/10.1515/9783110686647-003>

Typologists and historical linguists, in contrast, have long pointed out the dynamic nature of the class of adpositions. For example, König and Kortmann (1991: 109), while conceding that “the inventory of prepositions in English and other languages is only a very limited one”, stress that “the class of prepositions is by no means a closed class, but admits of new additions and in fact constantly changes its extension in the course of the history of a language.” Descriptive grammars of German tend to agree, describing the class of adpositions as ‘not fully closed’ (Zifonun et al. 1997: 2075) or even straightforwardly ‘open’ (Duden 2009: 600).

It is precisely this openness that makes the class of adpositions worthy of continued study, both from a diachronic and from a synchronic perspective. In this paper, we will combine both of these with a comparative perspective within the West Germanic languages (in particular, German, Dutch, and English). After a general overview of adpositions and their properties in the three languages (Section 2), we will focus on the subclass that is often called ‘complex adposition’, ‘adpositional locution’, or ‘adposition-like collocation’. We will discuss the synchronic properties that warrant their inclusion in the class of adpositions (Section 3) and then discuss their diachronic role as a starting point for the development of new adpositions.

Complex adpositions have been quite extensively (but by no means exhaustively) studied from both a synchronic and a diachronic perspective, but we feel that a comparative approach to the parallels and differences that these expressions display even in closely related languages can add insights. Like the other contributions in this volume, the present article is intended to serve as an inventory of the relevant aspects and problems, based both on descriptions in the existing literature (including dictionaries and descriptive grammars) and on complementary corpus data.

As there are no strictly comparable corpora for the three languages, we work with what is available. For German, this is the *DWDS-Kernkorpus*, a 120-million-word representative collection of German written texts from the 20th century (www.dwds.de) and, for historical aspects, the *Deutsches Textarchiv* (DTA, www.deutsches-textarchiv.de, 215 million words), which is also accessible via the DWDS-interface. For Dutch, it is *OpenSoNaR*, a 500-million-word reference corpus of contemporary written Dutch (opensonar.inl.nl), and the *Corpus Hedendaags Nederlands* (CHN), another 440-million-word corpus of contemporary Dutch (chn.inl.nl). For the historical perspective, we turn to the historical dictionaries of Dutch available at the ‘Instituut voor de Nederlandse taal’ (gtb.inl.nl). For English, our data mainly comes from the 100 million word *British National Corpus* (BNC, www.natcorp.ox.ac.uk), which contains written and spoken language covering a wide range of text types and situations, complemented by the

Corpus of Contemporary American English (COCA) and the *Corpus of Historical American English* (COHA), both accessible online (www.english-corpora.org) and the citation base of the *Oxford English Dictionary* (OED, www.oed.com).

2 Adpositions in the West Germanic languages: Characteristics and classification

2.1 Definition

The word class ADPOSITION is typically characterized as a (relatively) closed class, consisting of (i) invariable (non-inflecting) elements that (ii) combine with nominal complements to (iii) form a phrase type (called adpositional phrase) which occurs as a modifier in noun phrases, as an adverbial or, less typically, as a complement, where the adposition (iv) establishes a semantic relation between the referent of its nominal complement and some other entity or proposition (cf. e.g. Huddleston 1988: 123–124, Quirk et al. 1985: 657, Zifonun 1997: 2077–2098, Kurzon and Adler 2008: 2). In the West Germanic languages under discussion here, they are typically prepositions, i.e., they (v) precede their nominal complement; in German, they also (vi) assign (oblique) case (cf. Zifonun 1997: 2077f).

The most frequent prepositions in these languages also tend to be (vii) derivationally simple and diachronically old, as the data in (1) show:

- (1) Top 15 adpositions, ordered by decreasing frequency (CELEX, Baayen et al 1995)
- a. German¹:
- in* ‘in’, *zu* ‘to’, *von* ‘of’, *mit* ‘with’, *auf* ‘on’, *für* ‘for’, *bei* ‘by’, *an* ‘at’, *aus* ‘out’, *nach* ‘after’, *vor* ‘before’, *durch* ‘through’, *um* ‘(a)round’, *zwischen* ‘between’, *über* ‘over’

¹ As is well-known, prepositions are not translatable outside of specific usage contexts. We have tried to give the closest equivalent for the most literal use of each preposition throughout the paper. However, we do not include such translations in Tables 1 and 3, which are complex enough as is, and trust that readers who are interested enough to read a chapter on prepositions in German, Dutch and English will also be motivated enough to consult a dictionary in the few instances where they are unable to recognize what lexical material they are dealing with.

b. Dutch:

van ‘of’, *in* ‘in’, *te* ‘to’, *met* ‘with’, *op* ‘on’, *voor* ‘for’, *aan* ‘at’, *naar* ‘to’, *om* ‘(a) round’, *bij* ‘by’, *door* ‘through’, *over* ‘over’, *tot* ‘to’, *uit* ‘out’, *tegen* ‘against’

c. English:

of, *in*, *for*, *to*, *with*, *on*, *at*, *by*, *from*, *about*, *than*, *over*, *through*, *after*, *between*

A systematic exception to the property of invariability is found in German, where the determiner of the nominal complement may cliticize (or fuse) with the preposition, as in (2):

(2)	DAT	<i>in + dem = im</i>	<i>auf + dem = aufm</i>	<i>an + dem = am</i>
		‘in the’	‘on the’	‘at the’
	ACC	<i>in + das = ins</i>	<i>auf + das = aufs</i>	<i>an + das = ans</i>
		‘in(to) the’	‘on(to) the’	‘at the’

The distribution of the cliticized and non-cliticized forms is quite complex, as are the conditions under which cliticization can happen at all (see e.g. Duden 2009: 615–619; Nübling 1998; Schiering 2005; Augustin 2018). In the other two languages discussed here, cliticization does not occur, although Dutch has historical remnants showing that cliticization used to be possible (e.g. *te + der = ter*, *te + den = ten*). As Dutch has lost its case system, these cliticized forms survive only in fixed expressions (e.g. *ter plaatse* ‘on the spot’, *ten eerste* ‘in the first place’).

While the prepositions in (1) meet all criteria of the traditional definition, there is a large class of expressions that are, by now mostly uncontroversially, regarded as members of the class of adpositions and that call into question their status as a closed class: so-called complex adpositions. Table 1 provides an overview of these expressions based on the formal criterion of synchronically transparent morphosyntactic complexity. What is synchronically transparent is, of course, a matter of degree – initially complex adpositions of all types can become simple (or simpler) adpositions through various processes of language change discussed further in Section 4.

Aside from their morphological or syntactic complexity, these expressions conform to the definition provided above. They can thus be seen as less central, but nevertheless full members of the class of adpositions – less central for two reasons: first, their age and their role in diachrony, and second, their discourse frequency.

With respect to diachrony, simple, highly frequent adpositions are the oldest members of the category. With respect to discourse frequency, simple adpositions are much more frequent than the various types of complex adpositions shown in Table 1 above: In the FLOB corpus of (written) British English, which

Table 1: First overview of words usually considered adpositions in German, Dutch, and English.

	GERMAN	DUTCH	ENGLISH
Simple	<i>in, für, auf, an, zwischen</i>	<i>in, voor, op, aan, tussen</i>	<i>in, for, on, at, among</i>
Morphologically complex			
conversion (real or apparent)			
from nouns	<i>kraft, trotz, dank, zeit</i>	<i>richting, hartje</i>	<i>inside, outside</i>
from adjectives and adverbs	<i>nah, fern, unweit</i>	<i>binnen, buiten, rond</i>	<i>near, round, opposite, absent, effective, [worth]</i>
from present participles	<i>während, entsprechend, betreffend</i>	<i>aangaande, gedurende, hangende, lopende, staande</i>	<i>including, following, concerning, regarding, considering, pending, barring, excepting, (during)</i>
from past participles	<i>ungeachtet, unbeschadet, ausgenommen</i>	<i>gezien, ongeacht, gegeven</i>	<i>given, granted, [aged]</i>
affixation			
on nouns	<i>angesichts, zwecks, bezüglich, nördlich, südlich, hinsichtlich</i>	<i>krachtens, middels, ondanks, wegens; overeenkomstig</i>	– (but cf. <i>thanks to</i>)
on adjectives	<i>unterhalb, jenseits</i>	–	–
on verbs	<i>einschließlich, ausschließlich</i>	<i>volgens</i>	–
on adpositions	–	–	<i>toward(s), amongst</i>
on adverbs	–	–	<i>against</i>
compounding			
of adverbs and/or adpositions	<i>gegenüber, mitsamt, bis zu/um/in/auf</i>	<i>achterin, binnenin, bovenop, nabij, tegenover, voorbij</i>	<i>into, onto, out of, without, within, throughout, underneath (upon, unto)</i>
Syntactically complex (including cases showing signs of univerbation)			
Pres. Part. + P	<i>aufbauend auf, basierend auf</i>	–	<i>according to, owing to, depending on, relating to, consisting of, referring to, belonging to, arising from</i>
Past Part. + P	<i>abgesehen von</i>	<i>gebaseerd op</i>	<i>based on, associated with, caused by, provided by</i>

Table 1 (continued)

	GERMAN	DUTCH	ENGLISH
Adj + P	–	–	<i>due to, prior to, short of, close to, far from, dependent on, contrary to, consistent with</i>
Adv + P	<i>weg von, fernab von</i>	<i>ver van</i>	<i>away from, instead of, apart from, apart from, all over, forward to, regardless of</i>
N + P	–	–	<i>subject to, courtesy of, thanks to</i>
P + (Det)	<i>in Folge GEN/von, im</i>	<i>aan de hand van, in</i>	<i>in spite of, in front of, in terms of,</i>
N + NP _{gen} /	<i>Fall(e) GEN/von, im</i>	<i>de loop van, in plaats</i>	<i>in (the) case of, in the course of,</i>
PP _{of}	<i>Verlauf(e) GEN/von,</i> <i>(anhand GEN/von,</i> <i>aufgrund GEN/von,</i> <i>zugunsten GEN/von)</i>	<i>van, op basis van, op</i> <i>grond van, met behulp</i> <i>van</i>	<i>on top of, (instead of, because of)</i>
P + N + P	<i>in Bezug auf, im</i> <i>Hinblick auf, in</i> <i>Zusammenhang mit</i>	<i>met betrekking tot,</i> <i>in verband met, in</i> <i>tegenstelling tot, (met</i> <i>(het) oog op, inzake)</i>	<i>in regard to, with respect to, in</i> <i>relation to, in accordance with</i>

tags complex prepositions as prepositions rather than sequences of other word classes and thus allows a joint search for all types of adpositions, the most frequent synchronically complex prepositions, *out of* and *within*, rank in 18th and 19th position, the most frequent preposition derived from a participle, *including*, ranks 25th, and the most frequent preposition of the type [P + N + NP_{gen}/PP_{of}], *in terms of*, ranks 45th.

However, complex adpositions are by no means a marginal phenomenon, as the discussion in the second part of this paper will show. First, however, let us expand on the definition of adpositions in general.

2.2 Delineation

The definition states that adpositions combine with nominal complements to form adpositional phrases, but this is a simplification. While noun phrases are the most typical and most frequent type of complement in adpositional phrases, other types of complements may also occur, most importantly adjectives and adverbs, see Table 2:

Table 2: Adjectives and adverbs as complements of adpositions.

GERMAN	DUTCH	ENGLISH
<i>für richtig halten</i> 'consider (lit. 'hold for') correct'	<i>sinds kort</i> 'since recently (lit. short)'	<i>in particular, in general, in short, to have sth in common, at best, in vain, at large, for good/sure/certain, in private</i>
<i>in bar zahlen</i> 'pay in cash'		
<i>nach oben</i> 'to the top (lit. 'to up)'	<i>naar buiten</i> 'to the outside (lit. 'to out)'	
<i>bis gestern</i> 'until yesterday' <i>für immer</i> 'for ever'	<i>tot gauw</i> 'until soon' <i>voor altijd</i> 'for ever'	<i>at once, for ever, by far, until recently</i>

There is little disagreement that these cases should nevertheless be classified as adpositions, presumably because they are rare and typically occur in fixed or semi-fixed and often idiomatic expressions, so that it makes little difference to the grammatical description of a language how they are categorized.

Of course, allowing complements other than noun phrases in adpositional phrases raises the question how, if at all, adpositions can be distinguished from items like the following:

1. adverbs in sentences like *The children played **inside**, John ran **away**, I have not seen him **since**.*
2. words that govern adpositional phrases and that would usually be regarded as adjectives (***Due to** the rain, the children played inside*), adverbs (*John ran **away from** home*), adpositions (*John climbed **out of** the window*), etc.;
3. subordinating conjunctions in sentences like *I have not seen him **since he left**.*

Phrases headed by members of these word classes are distributed like adpositional phrases (i.e., they modify nouns or function as adverbials and complements), and the elements are often formally identical to clear cases of adpositions:

- (3) a. adverb:
 - I have not seen him **since**.*
 - Ich habe ihn **seitdem** nicht mehr gesehen.*
 - Ik heb hem **sindsdien** niet meer gezien.*
- b. adposition:
 - I have not seen him **since** his departure.*
 - Ich habe ihn **seit** seiner Abreise nicht gesehen.*
 - Ik heb hem **sinds** zijn vertrek niet meer gezien.*

c. conjunction:

*I have not seen him **since** he left.*

*Ich habe ihn nicht gesehen, **seit** er abgereist ist.*

*Ik heb hem niet meer gezien, **sinds** hij vertrokken is.*

Cases of formal identity across all three categories are rare and may, within the West-Germanic languages, be limited to English – note that the German and Dutch equivalents of the examples in (3) are identical only in their uses as adpositions and conjunctions. In German and Dutch, there is clear evidence that the adverbial use is derived from the prepositional use by univerbation of an adposition and a determiner (*seitdem* < *seit* + *dem*, *sindsdien* < *sinds* + *dien*).

However, such cases of formal identity across all three categories are not necessary to make the argument that adverbs, adpositions, and subordinating conjunctions are quite naturally treated as subcategories of a single word class, differing only in their valency requirements: adverbs like those mentioned above would be intransitive adpositions, complex adpositions of the type *away from* or *out of* would be simple adpositions that govern a PP, traditional adpositions could be analyzed as transitive adpositions, and (some) subordinating conjunctions could be analyzed as adpositions with a sentential complement. This view is argued forcefully, for example, by Huddleston and Pullum (2002) for English and Broekhuis (2013) for Dutch, but it is not (yet) universally accepted. We will briefly comment on the boundary between adpositions and subordinators here, but generally, we will remain agnostic with respect to the exact delineation of adpositions and conjunctions.

For German, two major cases of functional overlap between adpositions and subordinators may be distinguished (cf. Crössmann 1973). On the one hand, the same lexical item may function as an adposition and as a subordinator introducing a subordinated finite clause. This is the case for a very small number of German adpositions: a few simple ones (*bis* ‘till’, *seit* ‘since’, *außer* ‘except’) and one more complex one (*während* ‘while’). The adpositional uses are given in (4 a,b), and the subordinating uses are given in (5 a,b):

- (4) a. *Und der Bursch steht da und tut, als könnt’ er nicht **bis** fünfzählen.* [Christ, *Madam Bäurin*, 2001 (1920)]
 ‘And the boy is standing there, pretending he can’t count to five.’
- b. *In Tunesien nahm Humphrey laut AP erstmalig **während** seiner Reise ausführlich zum Vietnam-Problem Stellung.* [Archiv der Gegenwart, 2001 (1968)]
 ‘According to AP, Humphrey commented extensively on the Vietnam problem for the first time during his trip in Tunisia.’

- (5) a. *Lisa vergaß, warum sie hergekommen war, und wartete, **bis** die Kabinentür sich öffnete.* [Jentzsch, *Seit die Götter ratlos sind*, 1999]
 ‘Lisa forgot why she had come here and waited until the cabin door opened.’
- b. *In der Hauptsache war es allerdings der Wunsch meines Vaters, **während** meine Mutter dem Gedanken, daß ich studieren sollte, mehr und mehr zuneigte.* [Wien, *Ein Rückblick*, 2004 (1927)]
 ‘However, it was mainly my father’s wish, while my mother was more and more inclined towards the thought that I should study.’

For these usages, it has been proposed that the direction of change proceeded from adposition to subordinator. It has been assumed that before the reanalysis from adposition to subordinator there was an intermediate stage where the adposition combined with the default complementizer *dass* ‘that’ forming a complex subordinating expression such as [*bis* ‘till’ + *dass*], [*während* ‘while’ + *dass*]. It was presumably only in the later stages of the development that the complementizer *dass* was lost (cf. Paul 1920, Behaghel 1924, di Meola 2000).

On the other hand, some adpositions (e.g. *ohne* ‘without’, (*an*)*statt* ‘instead’, *außer* ‘except’) may function as subordinators and introduce non-finite clauses with the infinitive with *zu* ‘to’, as exemplified in (6):

- (6) a. *es geht auch **ohne** König.* [Schwanitz, *Bildung*, 1999]
 ‘it is possible without a king as well’
- b. *es sind Stoffe, die chemische Vorgänge beschleunigen, **ohne** sich selbst dabei **zu verändern**.* [Kölling, *Fachbuch für Kellner*, 1962]
 ‘they are substances that accelerate chemical processes without changing themselves.’

For these usages, there is no unified account so far, and they still need to be investigated from a diachronic perspective.

In Dutch, we find an overlap of adpositions (7a, b) and subordinators (8a, b) as well, as in the case of *sinds* ‘since’ briefly mentioned above:

- (7) a. *Dat zei de actrice [Angelina Jolie], die vorige maand beviel van dochter Shiloh, dit weekend in het eerste interview **sinds** haar bevalling, op de Amerikaanse nieuwszender CNN.* [OpenSoNaR]
 ‘This was said by the actress [Angelina Jolie], who last month gave birth to daughter Shiloh, this weekend in the first interview since her delivery, on the American news channel CNN.’

- b. *De oude vrouw bleef **tot** haar dood begin jaren negentig een verstokte communiste.* [OpenSoNaR]
 ‘The old woman remained an inveterate communist until her death in the early 1990s.’
- (8) a. *Het is het eerste interview van Angelina [Jolie] **sinds** ze eind mei in Namibië beviel van een dochtertje.* [OpenSoNaR]
 ‘It is the first interview of Angelina [Jolie] since she gave birth to a little daughter in Namibia at the end of May.’
- b. *En zijn moeder was ziek geweest, geestelijk labiel, en hij had niet per se voor haar hoeven zorgen **tot** ze stierf.* [OpenSoNaR]
 ‘And his mother had been sick, mentally unstable, and he hadn’t necessarily had to take care of her until she died.’
- c. *Colbert, die samen met zijn broers en zussen voor zijn moeder zorgde **totdat** ze stierf.* [mantelzorgelijk.nl]
 ‘Colbert, who took care of his mother together with his brothers and sisters until she died.’

Note that the subordinator in *tot ze stierf* in (8b) can be replaced by *totdat*, without a change in meaning (8c). This univerbation of [adposition + complementizer] can also be observed in other cases, such as *voor* and *voordat*. In these cases, it is clear that, as in German, the subordinators developed from the prepositions, with the lexicalized compounds consisting of adposition and complementizer paving the way for the use of the bare preposition as a complementizer.

In English, there is little reason in synchrony to distinguish adpositions from subordinators at the level of parts-of-speech. Most (if not all) adpositions that can take a nominal complement can also take a gerund clause (cf. 9 a, b). This is due to the fact that the latter can generally occur in positions where NPs also occur in English, which already blurs the distinction between adposition and subordinator when compared to German or Dutch. Some words that function like adpositions in other places can also occur as ‘traditional’ subordinators, which simply means that they take a subordinate clause as a complement (cf. 9 c):

- (9) a. *Members of the Salvation Army have remained silent **since their announcement** that their money had disappeared.* [BNC K5]
- b. *Knighon has hardly been out of the news **since announcing** his intention to buy United on 18 August, the day before the Football League season began.* [BNC A2S]
- c. *Lord Mackay said he believed the bill was a full response to the protests he had received **since he announced** the proposed changes.* [BNC A96]

Not all adpositions/subordinators can occur with all three types of complements – *while*, for example, can only take a gerundial clause or a subordinate clause, but no NP, *under* can only take an NP or a gerundial clause. However, as Huddleston and Pullum (2002) point out, the same types of differences exist for verbs, which are nevertheless seen as subcategories of a single category. Thus, they suggest that subordinating conjunctions simply be treated as adpositions with sentential complements (a categorization also found elsewhere, e.g. in the Penn Tagset, cf. Santorini 1990). Diachronically, this is also plausible, as elements have extended from either category to the other. For example, *after* already had both functions in Old English, *since* was used as a conjunction before it was extended to adpositional uses, and *before* was used as an adposition before it was extended to sentential complements.

2.3 Semantic properties

The definition states that adpositions establish a semantic relation between the referent of their complement and some other entity or proposition. Typically, adpositions (especially the old, simple ones) are highly polysemous, with a spatial core meaning (e.g. Huddleston 1988, Lindstromberg 2000) from which other relations are derived more or less transparently. English *on*, for example, is, as far as is possible to say, originally a spatial adposition (*the book is on the table*). It may be used to express temporal relations (*it was on 10 March*), causal relations (*she read the book on my suggestion*), or relations of manner (*she spoke on a rising tone*), amongst others (Rauh 1991: 170). We find comparable sets of extensions for many prepositions in all three languages:

- (10) a. LOCAL: **in** der Küche / **in** de keuken / **in** the kitchen
 b. TEMPORAL: **in** kurzer Zeit / **in** een korte tijd / **in** a short time
 c. MODAL: **in** großen Schwierigkeiten / **in** grote moeilijkheden / **in** great trouble
- (11) a. LOCAL: **unter** dem Tisch / **onder** de tafel / **under** the table
 b. TEMPORAL: **unter** der Woche / **onder** (more typically: **tijdens**) de week / **during** the week
 c. MODAL: **unter** Protest / **onder** protest / **under** protest
- (12) a. LOCAL: **an** der Bushaltestelle / **bij/aan** de bushalte / **at** the bus stop
 b. TEMPORAL: **um** sechs Uhr / **om** zes uur / **at** six o'clock
 c. MODAL: **auf** meine Bitte / **op** mijn verzoek / **at** my request

Note that there is no one-to-one correspondence between etymologically related adpositions in the three languages with respect to the semantic extensions. Despite the fact that such extensions have been plausibly argued to be conceptually motivated (e.g. Lakoff 1987, Brugman 1988, Cuyckens 1993, 1995, Meex 2002), the set of uses of any given adposition is highly language specific and must be treated as a problem of lexicography (cf. e.g., Lindstromberg's (2000) semantic description of simple English prepositions). German *unter*, for example, has a temporal extension that is possible but rare for Dutch and not possible for English; *at* has temporal and modal uses that are not found in German or Dutch, and Netherlands Dutch even uses *bij* instead of the cognate *aan* in some local contexts (Belgian Dutch uses *aan*).

One of the relevant properties of simple adpositions, particularly when compared with complex adpositions, is that in addition to the more lexical uses just described, they have semantically neutral (or bleached) structural variants that are part of verbal/nominal valency (see e.g. Eichinger 1995) and functionally resemble cases (e.g. *helfen* 'help' DAT, *ähnlich* 'similar' DAT etc. in German). Simple adpositions are found in combination with verbs (13a, b), verbal nouns (13c, d), as well as adjectives (13e, f):

- (13) a. *auf* etwas achten / letten **op** iets / pay attention **to** sth.
 b. *sich nach* etwas sehnen / naar **iets** verlangen / to long **for** sth.
 c. *Hoffnung auf* etwas / hoop **op** iets / hope **for** sth.
 d. *Nachfrage nach* etwas / eis **naar** iets / demand **for** sth.
 e. *stolz auf* etwas / trots **op** iets / proud **of** sth.
 f. *verrückt nach* etwas / gek **op** iets / crazy **about** sth.

As an extreme case of this semantic bleaching, prepositions may take on purely grammatical functions that may be motivated diachronically by their original meaning, but that are no longer transparently related to it – for example, English *of*, which functions almost exclusively as a genitive-like structural marker, and *von* and *van* in German and Dutch, which have this grammatical use in addition to their more lexical uses.

2.4 Position

As mentioned above, adpositions overwhelmingly occur as prepositions in the West Germanic languages, i.e., they precede their complement. However, in all three languages there are also elements that could be (and have been) described as

postpositions (following their nominal complement), “ambipositions” (alternating between prepositional and postpositional use) and circumpositions (with elements occurring on both sides of the complement), see Table 3.

Table 3: German, Dutch, and English adpositions and their position.

	GERMAN	DUTCH	ENGLISH
PREPOSITION	<i>in, an, auf</i>	<i>aan, achter, met, naar, van</i>	<i>in, on, at</i>
POSTPOSITION	<i>zufolge, halber, zuliebe</i>	<i>af, geleden, terug</i>	<i>ago, on, hence</i>
AMBIPOSITION	<i>entlang, nach, wegen</i>	<i>door, in, om, op, over, uit, voorbij</i>	<i>notwithstanding, aside, apart</i>
CIRCUMPOSITION	<i>um ... willen, von ... an</i>	<i>tussen ... door, naar ... toe, achter ... aan, tegen ... in</i>	<i>from ... on, within ... of</i>

The existence of postpositions in these three languages is contentious in the literature. If they have postpositions at all, these are extremely rare. Potential examples from German are *halber* or *zuliebe* (cf. 14a, b), in Dutch we find *af* ‘down’, *geleden* ‘ago’ and *terug* ‘ago’ (15a, b), and English has *ago*, *on*, *hence* and the semi-fixed expression [[*all N*] *long*] (16a–d):

- (14) a. *Was er Juliane erzählte, ließ er – der Einfachheit **halber** – an einem Ort namens Buenos Aires spielen, aber er erzählte wahre Dinge.* [Kopetzky, *Grand Tour*, 2003]
 ‘What he told Juliane, he transferred – for simplicity’s sake – to a place called Buenos Aires, but he told true things.’
- b. *Dem schöpferischen Leben **zuliebe** kämpft Schiller dagegen an.* [Safransky, Friedrich Schiller, 2004]
 ‘For the sake of creative life, Schiller fights against it.’
- (15) a. *Hij vloog werkelijk die berg **af!*** [OpenSoNaR]
 ‘He really flew down the mountain!’
- b. *Twee weken **geleden** was de politie ook al eens binnengevallen in de tempel* [OpenSoNaR]
 ‘Two weeks ago the police had already entered the temple.’
- c. *de legendarische BBC-dj John Peel, die twee jaar **terug** overleed* [OpenSoNaR]
 ‘legendary BBC DJ John Peel, who passed away two years ago’

- (16) a. *Two years **ago**, less than 25 per cent had a car, now almost 35 per cent.* [BNC A1]
 b. *Twenty-five years **on**, the claims for compensation for medical disability are still trickling in.* [BNC AHN]
 c. *Three weeks **hence**, The Canadian will head west for the last time.* [BNC AAF]
 d. *He had been drinking all night **long**, beer and whisky.* [BNC BNC]

As just mentioned, there is no agreement in the literature as to whether these words are, in fact, adpositions at all. In traditional grammars, they are treated as adverbs or adjectives, with the temporal NP as a pre-modifier. This analysis would certainly increase the typological consistency of the languages under discussion, in which complements generally follow their heads so that the existence of postpositions would be highly exceptional. For German, Di Meola (2003) presents a detailed synchronic and diachronic discussion supporting, to some extent, a category postposition. For English, too, Huddleston and Pullum (2002: 632) have made the argument that the three English words should be treated as adpositions based on the fact that the temporal NP is obligatory and must therefore be regarded as a complement rather than a premodifier. While this is a plausible argument, Culicover (1999: 72) points out that if we analyze *ago* as an adposition, then it is a very irregular one, as unlike all other adpositions in English, it does not allow stranding (the same is true of *on* in the relevant sense, *hence* and *long*). In any case, if these elements are adpositions, they are so exceptional within their class that nothing follows from this classification.

The case for circumpositions is much stronger. All three languages have combinations of adpositions and adverbs/adpositions that are quite plausibly analyzed as circumpositions, such as *um ... willen* ‘for the sake of’ and *von ... an* ‘from ... on’ in German, *tussen ... door* ‘between’ and *door ... heen* ‘through’ in Dutch and *from ... on* and possibly *within ... of* in English:

- (17) a. *eine Ehe, die **um** materieller Vorteile **willen** geschlossen worden ist* [Engler, *Die Ostdeutschen*, 2000]
 ‘a marriage entered into for the sake of material advantage’
 b. ***Von** diesem Moment **an** entsteht eine neue Hierarchie.* [Sloterdijk, *Kritik der zynischen Vernunft*, 1983]
 ‘From this moment on, a new hierarchy emerges.’

- (18) a. *Vanmiddag neemt de neerslagneiging af en **tussen** de wolken **door** kan de zon zich heel even manifesteren.* [OpenSoNaR]
 ‘This afternoon the precipitation tendency decreases and between the clouds the sun can manifest itself for a moment’
- b. *Het Oude Stadsplein was **door** de eeuwen **heen** getuige van vele goede en tragische gebeurtenissen.* [OpenSoNaR]
 ‘The Old Town Square has witnessed many good and tragic events throughout the centuries.’
- (19) a. ***From the first day on**, when they had met on the wide steps of the strange school, they had loved one another...* [BNC BN1]
- b. *And after all, what price can you put on the immense satisfaction of taking dew-fresh vegetables to the kitchen **within** minutes **of** harvesting them?* [BNC AOG]

The reason for analyzing such sequences as circumpositions is, first, that the presence of the second adposition depends on that of the first. For example, as (20a) shows, the presence of *of* depends on that of *within* (unlike that of, for example, *after* in (20b)), but not vice versa, as (20c) shows, the same is true of *from* and *on*, as (20d, e) show:

- (20) a. **...taking vegetables to the kitchen **minutes of** harvesting them...*
 b. *...taking vegetables to the kitchen **minutes after** harvesting them...*
 c. *...taking vegetables to the kitchen **minutes after**.*
 d. ****The first day on**, they had loved one another*
 e. ***From the first day**, they had loved one another...*

Second, both parts of the circumposition typically relate semantically to the NP they enclose (this criterion would exclude cases like *within ... of*, where the second part has its own complement). Diachronically, some circumpositions may develop into morphologically complex adpositions through a change in word order between the nominal complement and the second part of the circumposition – for example, *von ... wegen* > *vonwegen* in German, *on ... gemang* > *ongemang* ‘among’ in English or *om ... wille* > *omwille* in Dutch (see further Section 4.3 below).

The case for ambipositions is strongest in Dutch, which has a whole set of primary adpositions that can be used both prepositionally and postpositionally. Examples are *door* ‘through’, *in* ‘in(to)’, *om* ‘(a)round’, *op* ‘on(to)’, *uit* ‘out’ etc. (cf. Broekhuis 2013 for a comprehensive list). Most of these adpositions have different functions in their use as prepositions and postpositions (cf. Broekhuis

2013: 33): as prepositions, they have a stative meaning, referring to a point in space (as in (21a), and (22a)), as postpositions, they refer to a path (as in (21b) and (22b)); the same difference can be seen with some temporal uses of adpositions, where the preposition refers to a point in time and the post position to a temporal path (as in 23a, b):

- (21) a. *heel mijn familie woont **in** de stad Luik* [OpenSoNaR]
 ‘my whole family lives **in** the city of Liège’
 b. *Tien zwaar bepakte jongemannen gingen aan land en liepen de stad **in**.*
 [OpenSoNaR]
 ‘Ten heavily loaded young men went ashore and walked **into** the city.’
- (22) a. *De verdachten komen **uit** Rotterdam en Den Haag.* [OpenSoNaR]
 ‘The suspects are **from** Rotterdam and The Hague.’
 b. *Dan reed ik Rotterdam **uit** over de Kralingse Plaslaan, langs de molens en daar de snelweg **op**.* [OpenSoNaR]
 ‘Then I drove **out of** Rotterdam over the Kralingse Plaslaan, past the windmills and **onto** the highway.’
- (23) a. *Onderwijsminister Luc Vandenbossche belooft **in** het nieuwe jaar te zullen stoppen met roken.* (Open Sonar)
 ‘Education minister Luc Vandenbossche promises to stop smoking in the new year.’
 b. *Ongeveer drieduizend jongeren swingden er het nieuwe jaar **in**.*
 [OpenSoNaR]
 ‘About three thousand young people were dancing **into** the new year.’

The uses of *uit* and *op* in (22b) illustrate a crucial property of Dutch postpositions: typically, they do not denote a mere direction, but a complete trajectory between two points. Therefore, the postpositional use is sometimes called ‘trajective’ (e.g. Draye 1992, van Bree 2014). This special semantics sets the postpositional use apart from the use of prepositions, like *over* (*de Kralingse Plaslaan*) and *langs* (*de molens*) in (22b), which also denote a direction, but lack the semantics of completion (reaching another point).

The semantic distinction between pre- and postpositional uses cannot be observed for all adpositions that alternate between these positions. In particular, there are adpositions that have an inherent path semantics, and can be used as prepositions and postpositions without any difference in meaning, for example, *door* ‘through’ or *voorbij* ‘past’.

- (24) a. *Als moeder Zwarts nu **voorbij** de villa rijdt wordt ze door de huidige bewoners botweg van het erf gejaagd.* [OpenSoNaR]
 ‘When mother Zwarts now drives **past** the villa, the current residents simply chase her away from the yard.’
- b. *Ik kan geen vijf huizen **voorbij** rijden, zonder minstens één keer een politicus te zien die mijn stem wil.* [OpenSoNaR]
 ‘I cannot drive **past** five houses without seeing at least one politician who wants my vote.’

In (24b), the postpositional use is prompted by the negation (*geen*), but semantically there is no difference to (24a).

This systematic alternation between locative and directional uses is not found in German or English, where ambipositions are more marginal in general. German has less than a dozen cases, many of which are limited to highly formal registers (e.g. *entgegen* ‘against’, *ausgenommen* ‘except’, *gemäß* ‘according to’, *entsprechend* ‘according to’, and *ungeachtet* ‘despite’). The more frequent ones are shown in (25):

- (25) a. **wegen** *der*_{GEN} *Kinder* / *der*_{GEN} *Kinder* **wegen**
 ‘because of the children’
- b. **gegenüber** *dem*_{DAT} *Haus* / *dem*_{DAT} *Haus* **gegenüber**
 ‘opposite to the house’
- c. **entlang** *des*_{GEN} *Flusses*, *dem*_{DAT} *Fluss* / *den*_{ACC} *Fluß* **entlang**
 ‘along the river’
- d. **nach** *meiner*_{DAT} *Meinung* / *meiner*_{DAT} *Meinung* **nach**
 ‘in my opinion’

Unlike in Dutch, there is typically no difference in meaning attached to the position. One apparent exception is *entlang* ‘along’, which does seem to display a semantic difference analogous to the Dutch cases:

- (26) a. *In der Nähe des Hauptbahnhofs und **entlang der***_{GEN} ***Prager Straße** waren viele Tausende Menschen versammelt.* [Ingo Schulze, *Neue Leben*, 2005]
 ‘near the central station and along the *Prager* street, many thousand people were assembled’
- b. *Sie gehen **eine***_{ACC} ***enge Straße entlang**, vorbei an einer Zeile zweistöckiger schmaler Häuschen...* [Helmut Krausser, *Eros*, 2006]
 ‘they are walking along a narrow street, past a row of narrow two-storey houses’

As in Dutch, the prepositional use is stative and the postpositional use is dynamic; however, this is not due to the position itself, but to the different cases of the nominal complement: the prepositional use requires the genitive (or, in colloquial usage, dative) NP, the postpositional use requires the accusative. This case alternation is also found with some German prepositions and consistently correlates with the distinction between stative and dynamic readings (see Section 2.5). However, no other German ambiposition displays this case alternation.

There are other formal differences in some cases, but these have no effect on the meaning. For example, the nominal complement of *nach* ‘after’ may occur without a determiner in the prepositional, but not in the postpositional use:

- (27) a. **nach** meiner Überzeugung / meiner Überzeugung **nach**
‘to my mind’
b. **nach** der Überzeugung des Ministers / der Überzeugung des Ministers
nach
‘in the minister’s opinion’
c. **nach** Überzeugung des Ministers / *Überzeugung des Ministers **nach**
‘in the minister’s opinion’

English has three potential ambipositions, *notwithstanding*, *aside* and *apart*:

- (28) a. *Kiwi fruit and pizzas **notwithstanding**, a vast sameness has swept the supermarkets of Europe, North America and Australasia.* [BNC HH3]
b. *Nathan’s illness **apart**, theirs was a home life of high-minded living...* [BNC AOP]
c. *These grumbles **aside**, the Clio is an easily driven, user-friendly car.* [BNC A6X]

As in German, there can be formal differences between the prepositional and postpositional uses: *apart* and *aside* are simple adpositions in their postpositional use, but complex adpositions with *from* in their prepositional use, but neither this nor the position as such has an effect on propositional semantics (although there may be subtle pragmatic and information-structural differences). The examples in (28) can be paraphrased as shown in (29) with the same meanings:

- (29) a. **Notwithstanding** *kiwi fruit and pizzas, a vast sameness has swept the supermarkets of Europe, North America and Australasia.*
b. **Apart from** *Nathan’s illness, theirs was a home life of high-minded living...*
c. **Aside from** *these grumbles, the Clio is an easily driven, user-friendly car.*

In addition, English has two fixed expressions, *all year round* and *all night through* that look as though they contain postpositional uses of the prepositions *round* and *through*.

However, the status of all apparent ambipositions in English can, again, be called into question, as these, too, cannot be stranded (*notwithstanding* cannot be stranded at all, and *apart* and *aside* can only be stranded (marginally), in their prepositional uses with *from*. Thus, again, these words are more straightforwardly categorized as adverbs or adjectives.

2.5 Case assignment

In **German**, in the typical combination with a nominal complement, the adposition assigns case to the noun phrase. Simple prepositions, especially the old, frequent ones, overwhelmingly assign dative and/or accusative case (see Table 4 for the top 15 prepositions).

Table 4: Case assignment of German simple adpositions.

P	DAT	ACC	P	DAT	ACC
<i>in</i> 'in'	+	+	<i>aus</i> 'out'	+	–
<i>zu</i> 'to'	+	–	<i>nach</i> 'after'	+	–
<i>von</i> 'of'	+	–	<i>vor</i> 'before'	+	+
<i>mit</i> 'with'	+	–	<i>durch</i> 'through'	–	+
<i>auf</i> 'on'	+	+	<i>um</i> '(a)round'	–	+
<i>für</i> 'for'	–	+	<i>zwischen</i> 'between'	+	+
<i>bei</i> 'by'	+	–	<i>über</i> 'over'	+	+
<i>an</i> 'on'	+	+			

More recently evolved, less grammaticalized adpositions (especially the complex adpositions discussed below) often assign genitive case. In the course of their further development and with an increasing degree of grammaticalization, a shift from the genitive to the dative is very common (see Section 4.3 for further details). This can be observed as an ongoing process with *wegen* right now: *wegen des Sturms* (genitive) > *wegen dem Sturm* (dative) 'because of the storm'.

As pointed out in Section 2.4, where an adposition assigns either dative or accusative case, this is associated with a systematic difference in meaning, with

dative marked NPs receiving a locative meaning and accusative marked NPs receiving a dynamic reading (30a,b and c,d):

- (30) a. ACCUSATIVE
*Dann lief sie durch das Dorf **auf den** Berg* [Kerstin Jentzsch: Seit die Götter ratlos sind, 1994]
 ‘then she ran through the village up the mountain’
- b. DATIVE
*Der Nebel wurde so dicht, daß er über Nacht **auf dem** Berg bleiben mußte* [Die Zeit, 28.09.1990]
 ‘the fog became so thick that he had to stay on the mountain overnight’
- c. ACCUSATIVE
*Ich bestellte ein Taxi und brachte sie **vor das** Haus ihrer Mutter.* [Luise Rinser, Mitte des Lebens, 1952]
 ‘I ordered a taxi and accompanied her to her mother’s house’
- d. DATIVE
*Beide hatten schon, eine Weile **vor dem** Hause Marias gestanden.* [Die Zeit, 09.08.1951]
 ‘both of them had been standing in front of Maria’s house for some time’

However, with adpositions that do not alternate between the two cases, the dative case may also receive dynamic interpretations (31a) and the accusative stative ones (31b):

- (31) a. DATIVE
*Sie läuft **aus dem** Zimmer.* [Neue deutsche Literatur, 1953]
 ‘she runs from the room’
- b. ACCUSATIVE
*Alle standen **um den** herrlich geschmückten Tannenbaum herum* [Berliner Zeitung, 28.11.1997]
 ‘everyone stood around the beautifully decorated Christmas tree’

Modern Dutch has lost morphological case marking and adpositions no longer assign case to NPs. Remnants of the old system can, however, be found when the complement is a personal pronoun. With pronouns, Dutch still distinguishes between the subject form (the old nominative) and the object forms. Adpositions always assign non-nominative ‘case’ to such pronouns: *met mij, van ons* (= object forms of the pronouns) vs. **met ik, *van wij* (= subject forms). Furthermore, we find a correspondence with the old case system with those spatial or temporal adpositions that can be used as prepositions (which have a stative reading) and

as postpositions (which refer to a direction or path). The former correspond to the use of the dative in German (*in de tuin* ‘in dem/im Garten’) and the latter to the accusative (*de tuin in* ‘in den Garten’), analogous to the case of German *entlang* ‘along’ discussed above.

In Modern English, case is no longer a relevant grammatical category. Where pronouns have different forms for subjective (“nominative”) and objective (“dative/accusative”) case, simple adpositions always assign objective case (as in Dutch). However, a diachronically motivated reflex of case can be seen in the use of the semantically relatively empty and highly grammaticalized adpositions *of* and *to*, which have taken over many of the functions of the genitive case and the dative/accusative distinction. Many simple adpositions that govern the genitive in German have a complex adposition with *of* or *to* as their closest English translation equivalent, for example, *wegen des Regens* (genitive) corresponds to ***because of the rain***, *statt einer Rede* corresponds to ***instead of a speech***, *dank eines Zufalls* corresponds to ***thanks to a coincidence***, etc. Meaningful distinctions between dative and accusative in German are also often expressed in English by *to*, for example, *im Haus* (dative) vs. *ins Haus* (accusative) correspond to ***in the house*** and ***into the house***. Thus, a particular type of complex adposition in English corresponds to German adpositions with exceptional case marking (genitive) or case distinctions (accusative/dative); the functional similarity goes some way toward accounting for the fact that *of* and *to* are so frequently found in English complex adpositions.

3 Complex adpositions in West Germanic

3.1 Overview

Following the definition given in the introductory chapter of this volume, we treat complex adpositions as having the same range of functions and distributional properties as simple adpositions, given in (a)–(c) below, but different formal characteristics, given in (d)–(e) below.

Typical simple adpositions as well as complex adpositions

- a. introduce a complement, typically a noun phrase;
- b. establish a relation between the complement and some higher constituent;
- c. specify the semantic nature of this relationship.

In addition, complex adpositions

- d. are morphologically and/or syntactically complex; and

- e. distribute the functions in (a)–(c) over different components; in the most typical case (referred to as canonical in the introduction to this volume), there are three elements:
- i. an element introducing a complement (P₂)
 - ii. an element establishing the relation to the higher constituent (P₁), and
 - iii. an element determining the semantic nature of the relation ((Det) N).

The canonical form [P₁ (Det) N P₂] is found in all three languages discussed here, an example is given in (32):

- (32) a. *Die Unterschiede **in Bezug auf** den Beginn der Muskelantworten sind beträchtlich* [Rathmeyer, *Motorische Steuerung bei Invertebraten*, 1996]
 ‘The differences with regard to the onset of muscle responses are considerable.’
- b. *Meest opmerkelijk waren de verschillen **met betrekking tot** de personeelkosten.* [OpenSoNaR]
 ‘Most notable were the differences with regard to personnel costs.’
- c. *Cultural differences **in regard to** time have long been of fascination to scholars.* [COCA 4012244]

In German, there is a variant of this canonical case where the nominal complement is not introduced by an adposition, but instead receives structural, typically genitive, case: [P₁ (Det) N NP_{GEN}], as in (33a):

- (33) a. *Anstatt dieser numerischen Darstellung **in Form einer**_{GEN} **Tabelle** kann man auch eine graphische Darstellung [...] wählen.* [Gustav Lienert, *Testaufbau und Testanalyse*, 1961]
 ‘instead of this numeric representation in the form of a table one can also choose a representation in the form of a graph’
- b. *In Aalst kampten we met de gekende problemen **in de vorm van** ziektes en blessures, waardoor we niet op de normale manier konden werken.* [OpenSoNaR]
 ‘In Aalst we had to deal with the known problems in the form of illnesses and injuries, so we could not work in the normal way.’
- c. *But knowledge **in the form of** policy specifics is of less value to the president who hopes to use power to realize policy objectives.* [COCA 4007793]

For Beneš (1974: 36), this variant is actually the canonical form of complex adposition in German, he regards the variant with a P₂ as a special case. English and Dutch both had this variant at earlier stages in their development (see further

Section 4.1 below), but largely lost it along with their case systems and replaced it by the form $[P_1 (\text{Det}) N \text{ van/of}_{P_2}]$. However, in Dutch, the variant with the genitive is still found in formal and archaic language and in fixed expressions like *in de loop der jaren*_{GEN} ‘over the years’ or *in de lijn der verwachtingen*_{GEN} ‘in line with expectations’ (cf. Scott 2014); in English, at least some complex adpositions with *of* alternate with a variant with the prenominal possessive (e.g. *on behalf of my father* ~ *on my father’s behalf*, *for the sake of my father* ~ *for my father’s sake*, cf. Huddleston 1984: 343).

Thus, while $[P_1 (\text{Det}) N \text{ NP}_{\text{GEN}}]$ certainly cannot be regarded as typical for the West Germanic languages in general, we include it as a variant of what is referred to as the ‘canonical’ form in this volume. Note that there is, in fact, also a clear synchronic motivation for doing so: even in German, every complex adposition of the form in (33a) has a variant with *von* that fully conforms to the canonical form, as in (33d):

- (33) d. *Danach schließt sich die Ergebnisdarstellung häufig in Form von einer*_{DAT}
Tabelle an [(Google Books) Regina Krieg, Realisierung von Telearbeit, 2013]
 ‘Frequently, the presentation of results in the form of a table follows’

Given this alternation, it would be very odd to treat the variant with case marking as fundamentally different. By including it and stressing the close analogy between genitive marking and marking with *von*, we also have a natural way of accounting for the fact that *van* and *of* are the most frequent P_2 in Dutch and English respectively. Table 5 shows the two variants of the canonical form.

Our use of the category labels P, N, NP etc. is not meant to suggest any particular stance on the constituency structure of complex adpositions. As they start out (at least in some cases, cf. Section 4.2 below) as free combinations, it is plausible to assume that in these cases they initially have canonical constituency structure, i.e. $[P_1 [(\text{Det}) [N [P_2 \text{ NP}]_{\text{PP}}]_{\text{N}'}]_{\text{NP}}]_{\text{PP}}$. It is also plausible, that at least some complex adpositions end up as completely fixed and, in some cases, semantically opaque expressions (see further below), which has led some to suggest an analysis along the lines of $[[P_1 N P_2]_{\text{P}} \text{ NP}]_{\text{PP}}$, with no internal structure at all in the $[P_1 N P_2]$ sequence.²

Regardless of the specifics of constituent structure, the schema introduced above is an analytical approximation and inevitably glosses over linguistic

² See, for example, the discussion in Seppänen et al (1994). Which of these analyses is closer to the truth remains a matter of contention. Even the authors of this paper were not able to come to a common position, or even agree that it is necessary to do so, and will thus follow Huddleston’s (1984: 344) advice that this “is a tricky area where it ill becomes us to be dogmatic”.

Table 5: Canonical complex adpositions in German, Dutch, and English.

HIGHER CONSTITUENT (A)	ELEMENT ESTABLISHING THE RELATION TO A	ELEMENT SPECIFYING THE SEMANTICS	ELEMENT INTRODUCING B	COMPLEMENT (B)
(a)	[P ₁	N	P ₂]	
<i>(die Unterschiede)</i> 'the differences	<i>in</i> with	<i>Bezug</i> regard	<i>auf</i> to	<i>(den Beginn...)</i> the start'
<i>(de verschillen)</i> 'the differences	<i>met</i> with	<i>betrekking</i> regard	<i>tot</i> to	<i>(de personeelskosten)</i> the personnel costs'
<i>(Soviet policy)</i>	<i>with</i>	<i>regard</i>	<i>to</i>	<i>(Eastern Euope)</i>
(b)	[P ₁	N	<i>von/van/ of</i> _{P₂/NP_{GEN}}	
<i>(eine Erklärung)</i> 'an explanation	<i>in</i> in	<i>Form</i> the form	<i>GENITIVE/von</i> of	<i>(einer Präambel)</i> a preamble'
<i>(water)</i> 'water	<i>in</i> in	<i>de vorm</i> the form	<i>van</i> of	<i>(ijskristallen)</i> ice-crystals'
<i>(knowledge)</i>	<i>in</i>	<i>the form</i>	<i>of</i>	<i>(policy specifics)</i>

reality: In practice, there will be many cases that deviate from such a perfect iconic correspondence, as each complex adposition is more or less idiosyncratic and not strictly compositional. Also, depending on the age and the degree of grammaticalization and/or lexicalization (see Section 5.1 for a discussion of grammaticalization/lexicalization accounts), complex adpositions will display this iconic correspondence to a greater or lesser degree. In general, the more a complex adposition is entrenched as a lexical unit, the less transparent is the structure given in Table 5. And of course, not all complex adpositions distribute these functions across three elements – recall cases like *due to* or *based on*, where they are distributed across two elements, with the first element taking on functions ii) and iii) and the second element taking on function i).

As complex adpositions constitute a relatively open class of expressions which is constantly expanded, it is very difficult if not impossible to give the exact number of the class members. Also, the boundaries to the sub-class of simple adpositions on the one hand and to free syntactic combinations on the other are often blurred. For this reason, we will refrain from giving any exact numbers. That the class of complex adpositions is larger than that of simple ones is a trivial fact. Apart from that, any estimation of the amount of complex adpositions would be connected with a theoretical decision as to what counts as a complex adposition and what not. In this chapter, we do not want to take sides.

In the following, the relevant criteria which apply to complex adpositions in West Germanic will be introduced and discussed in more detail using the example of the canonical pattern $[P_1 (\text{Det}) N P_2/\text{NP}_{\text{GEN}}]$; the criteria are relatively unanimously agreed upon in a variety of sources, e.g. Quirk and Mulholland (1964: 65), Huddleston (1984), Quirk et al. (1985: 671–6729), Huddleston and Pullum (2002) for English and Beneš (1974: 34–35), Lehmann and Stolz (1992); Lindqvist (1994); Di Meola (2000); Szczepaniak (2009) for German. Like most of these authors, we will argue that none of the characteristics may be taken as a defining property of a complex adposition, as different complex adpositions usually display a bundle – but almost never all – of these properties, and individual properties may often be attributed to different complex adpositions to a different degree. This is *inter alia* due to the different degrees of grammaticalization of different items, and due to other – often multiple – factors which may influence a development of a specific linguistic expression. In diachronic terms, the non-homogeneous behavior of complex adpositions with respect to the properties described below captures the dynamic nature of their development. In any case, the main purpose of these formal criteria is to capture the (more or less) unit-like status of complex adpositions, that is, to demonstrate that, despite their (more or less) transparent syntactic structure as multi-word combinations, they are (more or less) fixed and non-separable syntagms.³

To account for the dynamic and variable nature of the system of complex adpositions, and in accordance with the predominant view in functional and cognitive linguistics, we will assume that there is a continuum between free syntactic combinations on the one hand and simple adpositions on the other, with complex adposition being located in between these two extremes. This continuum may be

³ We largely ignore orthographic criteria pointing to the unit status of complex adpositions in the following, although these are sometimes mentioned, e.g., by Beneš (1974: 35), who points out that the noun in complex adpositions is often not capitalized, as would be expected in German, and that the P_1 and the noun are often spelled as a single word. Since orthography is not a direct reflection of mental representations, such observations must be interpreted with extreme care, but they are nevertheless of interest. We may add an additional orthographic argument for the unit status at least of Dutch complex adpositions: many of them have a conventionalized abbreviation that is used quite frequently (although the abbreviations are less frequent than the full form in all cases. For example, *met betrekking tot* ('with respect to') is abbreviated as *m.b.t.* (25 vs. 2 hits p.m.w.), *in verband met* ('in connection with') as *i.v.m.* (27 vs. 2 hits p.m.w.), *in plaats van* ('instead of') as *i.p.v.* (69 vs. 5 hits p.m.w.), *door middel van* ('by means of') as *d.m.v.* (18 vs. 1 hits p.m.w.), *naar aanleiding van* ('as a result of') as *n.a.v.* (30 vs. 1 hits p.m.w.), and *met medewerking van* ('with the cooperation of') as *m.m.v.* (both 1 hit p.m.w.). The frequent use of these abbreviation underlines the fixedness of the corresponding form. In English and German, such abbreviations are extremely rare: German has *z.Hd.* for *zu Händen von* ('c/o', *lit. to hands of*), English has *w.r.t.* for *with regard to*.

interpreted in diachronic terms, meaning that an individual complex adposition may change with respect to its degree of grammaticalization as it progresses “from left to right”. On the other hand, the continuum may also be interpreted in synchronic terms, meaning that there is a lot of variation within the class of complex adpositions at a particular point in time. Hence, all properties introduced further below may be attributed to particular expressions to a lesser or greater degree, and this degree is taken to correspond to the degree of grammaticalization on the scale between a “free syntactic combination” and a “simple adposition” in Figure 1, where each stage may be characterized using a bundle of prototypical features.

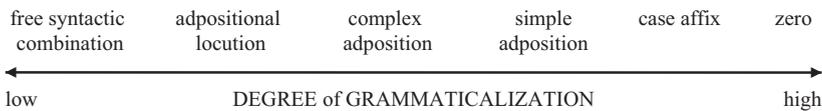


Figure 1: Grammaticalization of nominal relators (based on Lehmann and Stolz 1992: 11).

3.2 Restrictions on P_1

In typical complex adpositions of the form $[P_1 (\text{Det}) N P_2/\text{NP}_{\text{GEN}}]$, P_1 is assumed to be invariable for a given item. Quirk, Svartvik and Leech (1985: 671) illustrate this for the English complex adposition *in spite of*, where a substitution of the P_1 *in* by a different adposition, even a semantically plausible one, produces an unacceptable result: **for spite of*. This contrasts with regularly combined syntactic phrases such as e.g. *on_{P1} the shelf_N by_{P2} (the door)*, where the adposition *on* may be replaced by e.g. *under*, resulting in a fully acceptable grammatical structure.

Indeed, in most complex adpositions in German, Dutch and English, the P_1 is restricted to one particular adposition and cannot be freely replaced by another one, as illustrated in (34):

- (34) a. German
*im/*mit/*bei Zusammenhang mit; im/*unter/*mit Gegensatz zu*
- b. Dutch
*in/*met/*bij verband met; in/*onder/*met tegenstelling tot*
- c. English
*in/*with/*by connection with; in/*under/with contrast to*

However, a small number of complex adpositions seem to allow some variance with respect to P_1 without a noticeable change in semantics. For example, German

im Hinblick auf and *mit Hinblick auf* can be used largely interchangeably, as can English *in regard to* and *with regard to*:

- (35) a. *Das gilt sowohl **im Hinblick auf** das finanzielle Engagement als auch im Hinblick auf die Qualität der Beteiligung.* [Archiv der Gegenwart, 2001 (1967)]
 ‘This applies both **in** regard to the financial commitment and to the quality of the participation.’
- b. ***Mit Hinblick auf** den gegenwärtigen Konflikt sind zwei Punkte des Vertrages von besonderer Wichtigkeit.* [Archiv der Gegenwart, 2001 (1932)]
 ‘**With** regard to the current conflict, two points of the contract are of particular importance.’
- (36) a. *Like the majority of his contemporaries, Levin found himself in the vaguest position **in regard to** religion.* [BNC C8V]
- b. *Yet all these facets also carry negative implications **with regard to** religion.* [BNC HYB]

Further examples of variation in the P₁ position are *in/im/mit Bezug auf* ‘with regard to’ (German), *in/tijdens/gedurende de loop van* ‘in the course of’ (Dutch), and *with/in respect to, in/within the context of, in/by comparison with, in/during the course of* (English).

Of course, closer analysis may reveal subtle differences based on the different semantic contributions made by the P₁. In some cases, such differences are clearly visible. For example, German has **unter Berücksichtigung**, **bei Berücksichtigung** and **in Berücksichtigung** (all roughly translatable as *in consideration of*):

- (37) a. *Im Rahmen eines umfassenden Gedankenaustausches erörterten der Präsident und der Premierminister die gegenwärtige internationale Lage **unter Berücksichtigung** der Verpflichtungen, die ihre Länder gemeinsam mit ihren Verbündeten und Freunden für die Erhaltung von Frieden und stabilen Verhältnissen in allen Teilen der Welt tragen.* [Archiv der Gegenwart, 2001 (1964)]
 ‘In a comprehensive exchange of views, the President and the Prime Minister discussed the current international situation, taking into account the commitments, which their countries, together with their allies and friends, have made in order to maintain peace and stability in all parts of the world.’

- b. *Für die Unterbringung von 90 Stäben mit 26 mm Durchmesser wäre **bei Berücksichtigung** der vorgenannten Mindestabstände eine Fläche von 0,60 m × 2,0 m erforderlich gewesen.* [Alfred Martin, 1971]
 ‘An area of 0.60 m × 2.0 m would have been required for the accommodation of 90 bars of 26 mm diameter, considering the minimum distances mentioned above.’
- c. *Die Kontingente wurden **in Berücksichtigung** des Verbrauchs in Großbritannien drastisch erhöht.* [Archiv der Gegenwart, 2001 (1973)]
 ‘The quotas were drastically increased to take account of consumption in Great Britain.’

Here, the variants are not completely synonymous. In (37a), replacing *unter* by *bei* would be semantically odd, while *in* would be unremarkable. In (37b), however, *bei* could be replaced by *unter* but not by *in*. Finally, in (37c), *in* could be replaced by *unter*, but replacing it by *bei* would yield semantic oddity again. It seems that *bei* can only be used if the consideration of a particular state of affairs is still hypothetical, i.e., if it could also be left out of consideration – a better translation would be *if one were to consider x*. This means that it is not possible in (37a) and (37c), which describe specific situations in the past. The variant *in*, which is the least frequent by far and may in fact be due to contamination from *in Hinblick auf*, seems to refer to situations where something is not just taken into consideration among other things but is in fact the only consideration – a better translation would be *in view of*. Thus, it cannot be used in (37a) or (37b). The variant with *unter* has no such restrictions and can be used in all three contexts.

Thus, we must be careful to distinguish cases of true variation at P₁ from cases of semantically more or less related complex adpositions that differ with respect to their P₁. We are certainly dealing with the latter situation in the case of pairs of antonymic complex prepositions like those in (38), (39) and (40):⁴

- (38) a. *Das Vollstreckungsgericht hebt seinen Beschluß auf Antrag auf, oder ändert ihn, wenn dies **mit Rücksicht auf** eine Änderung der Sachlage geboten ist* [Theo Zimmermann, Der praktische Rechtsberater, 1968]
 ‘the court of execution repeals the decision upon petition, or modifies it if this is appropriate with regard to changed circumstances’

⁴ Note that in each case, the second member of the pair is less obviously a fully grammaticalized complex preposition, but our point here is that these are not cases of variation at P₁, but cases of two separate, albeit semantically related expressions.

- b. *Jederzeit widerruflich ist die Prokura **ohne Rücksicht auf** das der Erteilung zugrunde liegende Rechtsverhältnis.* [Theo Zimmermann, *Der praktische Rechtsberater*, 1968]
 ‘the power of attorney is withdrawable at any time without regard to the legal relation upon which it is based’
- (39) a. *hij vluchtte, **met medeneming van** de kroonjuwelen, met een Brits oorlogsschip* [OpenSoNaR]
 ‘he fled, taking along [lit: with with-taking of] the crown jewels, on a British warship’
 b. *Higgs was ondergedoken voor het verwachte mediacircus, **zonder medeneming van** zijn telefoon* (2013, CHN)
 ‘Higgs was in hiding for the expected media circus, without taking his phone with him’
- (40) a. *This, however, must be looked at **with regard to** aspects of employment law.* [BNC EC7]
 b. *He submitted that paragraph 19 was to be construed **without regard to** paragraph 16(1).* [BNC A49]

In cases of relative synonymy, one variant typically dominates in terms of frequency, see Table 6.

Table 6: Frequency of synonymous variants (p.m.w.).

GERMAN		DUTCH		ENGLISH	
<i>im Hinblick auf</i>	32	<i>met betrekking tot</i>	25	<i>with regard to</i>	16
<i>mit Hinblick auf</i>	0.25	<i>in betrekking tot</i>	0.03	<i>in regard to</i>	3
<i>unter Berücksichtigung</i>	11	<i>in de loop van</i>	33	<i>in comparison with</i>	4
<i>bei Berücksichtigung</i>	0.6	<i>tijdens de loop van</i>	0.03	<i>by comparison with</i>	2

In our view, it is not the fixed and non-substitutable P_1 as such that distinguishes complex adpositions from free syntactic combinations, but rather the degrees of variability with respect to P_1 , which is higher in free syntactic combinations and lower in complex adpositions (with no variation at all as the limiting case).

In light of the differences in frequency mentioned above, it is tempting to assume that the variability of the P_1 decreases in the course of the diachronic development of an adposition. One piece of evidence in favor of such a development is that the less frequent variant is sometimes restricted to older texts

or archaic language use. For example, Dutch translation equivalents of (35a, b) and (36a, b) both exist, but while the dominant form *met betrekking tot* ‘with regard to’ occurs freely, the synonymous variant *in betrekking tot* has a distinctly archaic feel:

(41) *De verhouding van meerdere tot mindere bestaat: a. krachtens hooger en militair rang; b. bij gelijkheid van rang, krachtens meerder en ouderdom daarin, doch alleen in betrekking tot dienst aangelegenheden* (Corpus Hedendaags Nederlands, Juridisch Corpus, 1922)

‘The relationship between superior and inferior exists: a. by virtue of senior military rank; b. in case of equal rank, by virtue of seniority within that rank, but only **in relation to** matters of service’

In general, Dutch seems to show very limited variation in the P1 position of complex adpositions, and when we do find variation, it is usually a remnant of variation in the past.

3.3 Restrictions on P₂

Like P₁, P₂ is also widely described as fixed, which again may be illustrated using the English adposition *in spite of*: whereas *in spite *for* is an unacceptable combination, a free syntactic combination such as *on the shelf by (the door)* may combine with *at*, *on the shelf at (the door)*. And again, this is the case for most complex adpositions in German, English and Dutch as well, see (42).

- (42) a. German
*im Zusammenhang mit/*an/*auf/*zu; im Gegensatz zu/*mit/*auf*
- b. Dutch
*in verband met/*aan/*op/*tot; met betrekking tot/*met/*over*
- c. English
*in relation to/*with/*of, in connection with/*to/*of*

However, similarly to the situation depicted above with P₁ a closer look into the data reveals that some variation is possible with respect to this restriction as well, cf. e.g (43) for German.

(43) a. *Die Mehrkosten sind gering im Vergleich zu den notwendigen häuslichen Vorbereitungen* [Giesder, *Gutes Benehmen*, 1991]

‘The extra costs are low compared to the necessary domestic preparations.’

- b. ***Im Vergleich mit dem Vorjahr ergibt dies ein Mehr von 1200 Mark.***
 [Naumburger Kreisblatt für Stadt und Kreis Naumburg, 26.11.1911]
 ‘Compared with the previous year, this results in an increase of 1200 Marks.’

In English, there is also a handful of cases where P2 seems to allow variation without any substantial difference in meaning, for example in the case of *in contrast to/with* (further examples are *in comparison to/with*, *in respect of/to*):

- (44) a. ***In contrast to the traditional liberal view, they have suggested that during the revolution the masses acted upon the political leaders as much as they were acted upon by them.*** [BNC FB1]
 b. ***In contrast with the views both of Rome and of the Puritans he emphasises the unity of Church and State.*** [BNC CFF]

Again, it is possible that a detailed analysis might uncover subtle differences. It has been suggested that, for the verb *contrast*, complements with *to* may encode a stronger contrast than *with* (Bernstein 1965), and this might be true for the complex adpositions here, too.

The Dutch equivalent of *in contrast to* shows the same variation:

- (45) a. ***In tegenstelling tot de scheidsrechters in Oost-Vlaanderen zijn die in West-Vlaanderen, Henegouwen en Antwerpen een stuk chauvinistischer.***
 [OpenSoNaR]
 ‘***In contrast to*** the referees in East Flanders, those in West Flanders, Hainaut and Antwerp are more chauvinistic.’
 b. ***In totaal hebben we zo’n negentig spelende leden en in tegenstelling met vele andere muziekkorpsen mogen we zeggen dat ons ledenbestand er vrij jong uitziet.*** [OpenSoNaR]
 ‘In total we have about ninety members playing and ***in contrast with*** many other bands we can say that our membership looks quite young.’

Again, one variant is very dominant in terms of frequency. In German, this difference is relatively moderate: *im Vergleich zu* ‘compared to’ has 7 hits p.m.w. in the DWDS Kernkorpus, *im Vergleich mit* ‘compared with’ has 3. In English, the difference is more pronounced, with 9 vs. 2 hits p.m.w. for *in contrast to* vs. *in contrast with* in the BNC; 5 vs. 2 hits p.m.w. for *in comparison with* vs. *in comparison to*; and a very large difference of 29 vs. 1 hit p.m.w. for *in respect of* vs. *in respect to*. In Dutch, the differences are highest on average: OpenSoNaR has 34 hits p.m.w. for *in tegenstelling tot* ‘in contrast to’ vs. 1 for *in tegenstelling met* ‘in contrast with’,

and 24 hits p.m.w. for *in vergelijking met* ‘compared with’ vs. 0.4 for *in vergelijking tot* ‘compared to’. Again, it seems that Dutch is the least flexible language with respect to complex adpositions.

A plausible working hypothesis is, once again, that the variability of the P_2 decreases in the course of the diachronic development of an adposition. However, it might well be the case – at least for some complex adpositions such as *im Zusammenhang mit* ‘in relation to’, *im Gegensatz zu* ‘in contrast to’ etc. – that the P_2 has been fixed right from the onset, as a part of the (inherited) valency properties of the verb underlying the deverbal nominal *Zusammenhang – mit etw. zusammenhängen* ‘relate to sth.’. It goes without saying that careful diachronic investigations are needed to further substantiate this hypothesis.

3.4 Restrictions on N

Even more pronounced restrictions are assumed to apply to [(Det) N], which makes sense given that the noun is the main contributor of meaning in complex adpositions. These restrictions are two-fold, concerning, first, the possibility of exchanging it for another, semantically related noun, and second, the modifiability of the noun.

With respect to the first type of restriction, note that there are cases of near-synonymous complex prepositions that differ only with respect to the noun. For example, English has *in breach of*, *in contravention of* and *in violation of*:

- (46) a. *A UN report on the Iran-Iraq war [...] found that Iraq’s invasion of Iran on Sept. 22, 1980, [...] had been unjustified and **in breach of** international law.* [BNC HLE]
- b. *The Wine Committee (three to six members strong) was to be elected by all categories of membership, so as not to be **in contravention of** the law.* [BNC AMY]
- c. *[T]he Second Circuit thought that ‘one who misappropriates non-public information in breach of a fiduciary duty and trades on that information to his own advantage’ was **in violation of** the law.* [BNC ECD]

However, this does not mean that there are no restrictions on the choice of the noun, as shown by the non-existence of other potential synonyms like **in infraction of*, **in rupture of*, **in infringement of*, **in transgression of*, **in abuse of*, etc. There are only a few such cases (other examples are *in conjunction/combination with*, *as a result/consequence of*, and *with regard/a view to*), for most complex

adpositions in English no such synonyms exist. As in the case of variation at the P₁ or P₂ position, there is typically a dominant variant in terms of frequency (e.g., *in breach of* occurs 5 times p. mil., in contravention of and *in violation of* occur less than once p.m.w. in the BNC).

In Dutch, the word *behulp* only occurs in the complex preposition *met behulp van* ‘with the help of’. In this expression, it can be replaced with its modern equivalent *hulp*: *met (de) hulp van*. The two variants can be used *synonymously* in most cases, but *met behulp van* is much more frequent (it occurs almost three times as often as its counterparts in the OpenSoNaR corpus (where *met behulp van* gets 17 hits p. m.w., *met hulp van* 3, and *met de hulp van* 5). Other complex prepositions can be used interchangeably in many, but probably not all contexts (*ten aanzien van*, *ten opzichte van*). They occur with about the same frequency in the OpenSoNaR corpus.

In German, the situation is very similar to English, there are cases of near-synonymous complex prepositions that differ only with respect to the noun, cf. e.g.:

- (47) a. *Er scheint sehr hoffnungslos **in Hinsicht auf** schweizerische Gerechtigkeit*
[Frisch, Stiller 23]
‘He seems very hopeless with regard to Swiss justice’
b. *Und wie würden Sie Ihre Tätigkeit **im Hinblick auf** die Akten nennen?* [Der Nürnberger Prozeß, 1946]
‘And what would you call your activity with regard to the files?’

However, similarly to English, the noun cannot be exchanged freely, as shown by the non-existence of other potential synonyms like *im Laufe/*Gang/*Rennen der Zeit* ‘in the run/*gear/*race of time’. Instead, there are only a few cases of noun alternation which seem to be restricted to a small number of fixed and conventionalized combinations (other examples are *im Zuge GEN/von* ‘in the course of’ – *im Wege GEN/von* ‘in the way of’ – *im Verlauf(e) GEN/von* ‘in the course of’, *im Vergleich zu* ‘in comparison to’ – *im Kontrast zu* ‘in contrast to’).

With respect to the second type of restriction, most of the sources cited above note that the N in complex adpositions cannot be modified syntactically (e.g. by adjectives and determiners), or morphologically (e.g. by plural marking). Let us begin with adjectival modification. Di Meola provides the following judgments (2000: 105–109) to exemplify this constraint:

- (48) a. ***Im Zug** der Verhandlungen hat sich eine Kompromisslösung abgezeichnet.*
‘in the course of the negotiations, a compromise solution has emerged’
b. ****Im schnellen Zug** der Verhandlungen hat sich eine Kompromisslösung abgezeichnet.*
‘in the quick course of negotiations...’

It is frequently pointed out, however, that this criterion is not an absolute one, and there are cases that seem to allow some modification (Quirk et al 1964, Di Meola 2000: 105, Klégr 2002: 30). The set of adjectives permitted is heavily constrained in these cases by the semantics of the whole complex adposition, and in this respect, it differs considerably from the original modifying potential of the lexical noun; frequently, the adjectives are conventionalized collocations of complex adpositions in question (Klégr (2002: 30) cites *within a wide area of, from a broad range of*).⁵

Consider for example the German complex adposition *im Rahmen GEN/von* ‘in the context of’, in which some adjectival modification is possible, see (49a). Note however that the adjectives allowed in the prenominal position comply with the general semantics of the complex adposition itself and specify further the ‘context’ or ‘scope’ as *eng* ‘narrow’, *größer* ‘wider’, or *beschränkt* ‘limited’. Other adjectives which can obviously combine with the noun *Rahmen* outside the complex adposition, see (49), such as *privat* ‘private’, *international*, etc., are not found in this pattern. This roughly corresponds to what has been called *lexical split* in the traditional research on language change and grammaticalization. Also, the distinct behavior of N within and outside the complex adpositions may be associated with different contexts of use (“switch contexts” in Heine 2002, “isolating contexts” in Diewald 2002) whose existence provides us with clear evidence of reanalysis:

- (49) a. **im engen Rahmen** der Problemstellung [Luhmann, *Soziale Systeme*, 1984]
 ‘in the narrow context of the problem’
im größeren Rahmen einer Atlantischen Gemeinschaft. [Archiv der Gegenwart, 2001 (1953)]
 ‘in the wider context of an Atlantic community’
- b. **im privaten/internationalen/gleichen/bisherigen/nationalen/europäischen/allgemeinen ... Rahmen**
 ‘in a private/international/same/present/national/European/general ... setting’

⁵ Klégr (2002: 30) notes one potential systematic exception for English, however: the complex adposition *within ADJ distance of*. Here, the adposition actually provides a relatively open slot that can be filled by a wide range of adjectives indicating distance (the BNC has *walking, striking, spitting, easy, touching, commuting, traveling*, and many others). However, note that this expression also occurs freely without the P2 *of* (the BNC has 95 cases with, and 71 cases without *of*); this is highly untypical of complex adpositions (Beneš 1974: 34–35 and Quirk and Mulholland 1964: 65 take the non-deletability of P2 as a defining feature of complex adpositions), which suggests that we might be dealing with an idiomatic expression of the form [*within ADJ distance*], and that the cases with *of* show regular complementation of the noun *distance*.

Similar examples can be found for Dutch (50a, b) and English (51a, b):

- (50) a. reading as complex adposition possible:
*De wraakzuchtige houding staat **in direct verband met** de politieke moorden op Pim Fortuyn en Theo van Gogh.* [OpenSoNaR]
 ‘The vengeful attitude is **in direct connection** with the political murders of Pim Fortuyn and Theo van Gogh.’
- b. reading as complex adposition not possible:
*Finnen werken **in internationaal verband** met anderen samen bij de mijnopruiiming.* [OpenSoNaR]
 ‘Finns work together with others in an international context in mine clearance.’
- (51) a. reading as complex adposition possible:
*any steps taken **in direct connection with** the preparation and service of a Schedule of dilapidations* [BNC J77]
- b. reading as complex adposition not possible:
*the ipsilateral medulla externa and lobula should be **in nervous connection with** the thorax* [BNC EVW]

Let us turn to modification by determiners next. It has frequently been claimed that nouns in complex prepositions cannot be modified by determiners and that they cannot occur in inflected forms. While this is true if we regard it as a general tendency, there is some variation with respect to these restrictions, too.

As pointed out in Section 2.1, German allows (and sometimes requires) a cliticization of the determiner of the nominal complement to the prepositional head. In many complex prepositions, the determiner is preserved and fossilized as such a clitic, rather than being lost:

- (52) a. [*in + dem*]: ‘in + the’
***im** Interesse* ‘in the interest’, ***im** Rahmen* ‘in the context’, ***im** Wege* ‘in the way’, ***im** Zuge* ‘in the course’, ***im** Umfang* ‘in the amount’, ***im** Verlaufe* ‘in the course’, ***im** Gefolge* ‘in the wake’
- b. [*an + dem*]: ‘on + the’
***am** Rand(e)* ‘on the brink’
- c. [*zu + dem*]: ‘to + the’
***zum** Wohl(e)* ‘for the good’, ***zum** Trotz* ‘in defiance’, ***zum** Zeitpunkt* ‘at the time’, ***zum** Preis* ‘for the price’, ***zum** Nachteil* ‘to the disadvantage’, ***zum** Zeichen* ‘for the indication’, ***zum** Vorteil* ‘to the advantage’

- d. [*bei + dem*]: ‘by + the’
beim *Anblick* ‘at the sight’

German also has a relatively systematic exception to the constraint on inflection: complex adpositions often preserve the option to use dative inflectional suffix *-e*, which has been lost in the language as a whole except for some fixed expressions:

- (53) *im Zug(e)* ‘in the course’, *im Gefolg(e)* ‘in the wake’, *im Fall(e)* ‘in the event’,
im Lauf(e) ‘in the course’, *im Weg(e)* ‘in the way’, *zum Zweck(e)* ‘for the purpose’

In Dutch, the absence of a determiner is usually a good indication of the status as complex adposition.

- (54) a. ***Naar aanleiding van de rellen in Parijs besliste de stichter van Axa dat diversiteit een hot topic moest worden*** [OpenSoNaR]
 ‘As a result of the riots in Paris, the founder of Axa decided that diversity should become a hot topic.’
 b. *Voorlopig blijft het gissen naar de aanleiding van de schietpartij.* [OpenSoNaR]
 ‘For the time being, we are still guessing at the reason of the shooting.’

In (54b) the adposition *naar* belongs to the verb (*gissen naar iets* ‘to guess at sth.’), while in (a) *naar aanleiding van* has to be analyzed as a complex adposition (‘because of, as a result of’). In its adpositional use the construction seems to be fixed; in the OpenSoNaR corpus, there are no examples of modification of the noun *aanleiding*. But a quick Google search reveals that even in such cases, the fixation is not 100%:

- (55) *Bijna 400 mensen hebben tot nu toe een gesprek gevoerd met SOS Telefonische Hulpdienst naar directe aanleiding van de rampen in Amerika.* [gic.nl]
 ‘Up to now, almost 400 people have had a conversation with SOS Telephone Help Service in direct response to the calamities in America.’

While the absence of a determiner indicates the adposition status, the presence of a determiner does by no means exclude this analysis. We even find many cases where a determiner is obligatory in a complex adposition (*aan de hand van* ‘on the basis of’, *in de loop van* ‘in the course of’, *in de trant van* ‘in the style of’, *in het kader van* ‘within the framework of’), while others have an optional determiner (*uit (het) oogpunt van* ‘from the viewpoint of’), which might indicate a transi-

tional phase (cf. Hüning 2014). Also, as in German, the determiner may cliticize to P1 instead of disappearing (cf. *ter* or *ten*); as mentioned in Section 2.1, cliticized forms of some prepositions survive only in fixed expressions – crucially, the latter include many fixed complex adpositions:

- (56) a. *ter gelegenheid van* ‘on the occasion of’, *ter aanvulling van* ‘as a supplement of’, *ter beschikking van* ‘at the disposal of’, *ter hoogte van* ‘at the height of’
 b. *ten aanzien van* ‘with regard to’, *ten behoeve van* ‘for the benefit of’, *ten gevolge van* ‘as a result of’, *ten zuiden van* ‘to the south of’, *ten nadele van* ‘at the expense of’

In English, the constraints against determiners and inflection hold as a general tendency, but there are a number of complex adpositions where the noun occurs in the plural obligatorily (*in terms of*) or optionally (*in case(s) of*, *for the purpose(s) of*), and there are complex adpositions where a determiner is obligatory (*in the course of*) or optional (*in (the) case of*, *at (the) risk of*, *in (the) light of*) (Quirk and Mulholland 1964 find that more than a third of the expressions they consider complex adpositions have a determiner).

It appears that the modification of N by determiners, adjectives and plural marking may be accounted for by at least two different tendencies. On the one hand, the decreasing ability to be modified by an adjective is in line with a more general diachronic tendency of the original syntactic structure to become more fixed and less analyzable (cf. our comments concerning the decreasing variability at P₁ and P₂ above). Newly evolving adpositions of the type [P₁ (Det) N P₂] often remain – at least for some period of time – partially transparent and analyzable. Within such new constructions, the semantic contribution of the noun is still apparent and more or less predictable from its lexical semantics outside the construction.

As the construction grammaticalizes, the noun becomes an integral part of the complex adposition, losing its autonomy and freedom to behave like an independent linguistic item. The corresponding changes, including the decreasing ability of the N to be modified by adjectives (but also by determiners or plural marking) and the decreasing ability to refer independently, correspond to what has been known as *decatégorialization* from the grammaticalization research. In extreme cases, a lexical split and loss of the original lexical noun may occur, as in *in Anbetracht GEN/von* ‘in view of’, for which there is no corresponding noun **Anbetracht* in present-day German.

Since grammaticalization can take a long time and its different stages can exist alongside each other, a more flexible view must be taken with respect to all

formal criteria discussed in this section. The class of complex adpositions is an open one in the West Germanic languages, which entails that there will always be variation within it as productive diachronic processes unfold – processes to which we now turn.

4 Diachronic considerations

4.1 The emergence and development of (complex) adpositions: Overview

4.1.1 German

Like the other West Germanic languages, German has inherited most of its simple adpositions from Proto-Germanic (e.g. *in*, *an*, *auf*, *zu*, *bei*). The most relevant diachronic sources of simple adpositions that can be recovered from the eldest available text records are adverbs (e.g. *ab* ‘up’, *an* ‘on’, *bei* ‘by’, *vor* ‘before’, *hinter* ‘behind’, *unter* ‘under’, *seit* ‘since’) and adjectives (e.g. *mitten*_{adj} > *inmitten* ‘in the midst of’, *twisc*_{adj} > *zwischen* ‘between’, see Behaghel 1924: 29–33).

Most of the more recent and often more formally transparent simple adpositions like *statt* ‘instead of’, *kraft* ‘by virtue of’, and *laut* ‘according to’ go back to complex adpositions of the form [P₁ (Det) N P₂ / CASE]: *an Statt* > *statt*; *in/aus/durch/bei Kraft* > *in Kraft* > *kraft*; *nach Laut* > *laut*. Most present-day German complex adpositions show no sign (yet) of this reduction, although a number of them show orthographic signs of univerbation, e.g. *anhand* ‘by means of’ < *an Hand*. Orthography by itself is, of course, an uncertain guide to univerbation, although Beneš (1974: 35) suggests we may treat it as a type of ‘sociolinguistic informant judgment’.

The majority of complex adpositions in present-day German are full-fledged realizations of the canonical pattern as e.g. *im Vergleich zu*, *in Bezug auf*, *in Folge von*/GEN (see Section 3 above). In any case, the canonical pattern [P₁ N P₂ / CASE] is the most frequent and productive formation pattern of complex adpositions in present-day German. Most “newcomers” to the category of complex adpositions are formed according to this structure, such as e.g. *im Sinne von*/GEN ‘in terms of’, *im Interesse von*/GEN ‘on behalf of’, *vor dem Hintergrund von*/GEN ‘against the background of’, etc. Often, their status is disputable between so-called “free adpositional combinations” and complex adpositions (see e.g. Lehmann and Stolz 1992: 17, Duden 2009: 600, Meibauer 1995, Eisenberg 1979).

Within the general canonical pattern [P₁ (Det) N P₂ / CASE], the schema with the genitive case and the alternative P₂ *von*, i.e. [P₁ (Det) N *von*/GEN], is the most productive one, whereas sub-patterns with other P₂ are less productive. The following examples from the corpus illustrate the pattern [P₁ (Det) N *von*/GEN] with the alternative genitive or *von*-phrase expression.

- (57) a. [P₁ N GEN]
Auf Grund dieser Notizen könnte ich jetzt ein eingehendes Bild entwickeln.
 [Der Nürnberger Prozeß, 1999 (1946)]
 ‘On the basis of these notes, I could now develop an in-depth picture.’
- b. [P₁ N *von*]
 Wir haben **auf Grund** von Erfahrungen beim Müllereibetrieb erkannt,
 daß ... [Vossische Zeitung, 05.03.1912]
 ‘On the basis of our experience in milling, we have recognized that...’

Table 7 gives some frequency counts for the five most frequent complex adpositions from these two sub-groups.

Table 7: Most frequent complex adpositions in present-day German (DWDS Kernkorpus).

SUB-PATTERN	EXAMPLES	FREQ. P.M.W.
[P ₁ N GEN / <i>von</i>]	<i>auf Grund</i> / <i>aufgrund</i> ‘on the basis of’	118
	<i>in Folge</i> / <i>infolge</i> ‘as a result of’	66
	<i>im Rahmen</i> ‘within the framework’	64
	<i>(an)statt</i> ‘instead of’	51
	<i>im Laufe</i> ‘in the course of’	50
[P ₁ N P ₂]	<i>im Gegensatz zu</i> ‘in contrast to’	46
	<i>in/im Zusammenhang mit</i> ‘in connection with’	33
	<i>in Bezug auf/ in bezug auf</i> ‘in regard with’	32
	<i>im Anschluss an</i> ‘following’	32
	<i>im Hinblick auf</i> ‘in view of’	31

A look at the diachronic situation of German reveals a relative stability, on the one hand, as the canonical pattern [P₁ N P₂ / CASE] has remained dominant since the earliest periods. On the other hand, there is variability as well, as concerns other patterns which seem to pop up and stay productive during relatively short periods of time (see Table 8 below).

Other patterns emerge later than the canonical pattern, and generally, do not survive until today. Table 8 summarizes the formation patterns of non-simple German adpositions and their periods of productivity.

Table 8: Major formation patterns of complex prepositions in German and their periods of productivity.

PATTERN	PRODUCTIVE PERIOD	EXAMPLES
[P ₁ N P ₂ /CASE]	at least 8th C. – PDG	<i>auf Grund / aufgrund von</i> /GEN ‘on the basis of’; <i>mit Hilfe / mithilfe von</i> /GEN ‘by means of’, <i>(an)statt von</i> /GEN ‘instead of’, <i>im Anschluss an</i> ‘following’, <i>im Hinblick auf</i> ‘in view of’, <i>in Bezug auf</i> ‘in regard with’, <i>im Vergleich zu</i> ‘in comparison to’, <i>im Einklang mit</i> ‘in accordance with’
[ADJ / ADV] [P ADV]	12th – 18th C.	<i>nebst</i> ‘along with’, <i>nah</i> ‘close to’, <i>fern</i> ‘far from’ <i>binnen</i> ‘within’, <i>mitsamt</i> ‘together with’, <i>inmitten</i> ‘amidst’
[N-s] _{ADV}	14th – 17th C.	<i>angesichts</i> ‘in the face of’, <i>mittels</i> ‘by means of’, <i>betreffe</i> ‘regarding’, <i>mangels</i> ‘for lack of’, <i>zwecks</i> ‘for the purpose of’
[V _{pres.part.}] [V _{past.part.}]	17th – 18th C.	<i>entsprechend</i> ‘according to’, <i>betreffend</i> ‘concerning’ <i>ausgenommen</i> ‘except’
[N _{dev-lich}]	18th – 19th C.	<i>abzüglich</i> ‘excluding’, <i>einschließlich</i> ‘including’, <i>bezüglich</i> ‘concerning’

In the Early New High German period (ca. 14th–17th C.), the adverbial pattern [N-s]_{ADV} gains in popularity (see Table 7). Later on, another adverbial formation pattern with the suffix *-lich* becomes popular. Both patterns are no longer productive in Present-Day German, but many adpositions formed after these patterns are still in use (e.g. *zwecks*, *betreffe*, *mangels*). Importantly, though the base nouns are still recognizable in the internal structure of these adpositions, the adpositions themselves are mostly perceived as non-separable and non-compositional units, partly due to the fact that the corresponding word formation patterns have fallen out of use.

The verbal pattern [V_{pres./past.part.}] was productive during the 17th and 18th centuries (Lehmann and Stolz 1992: 25). Deverbal adpositions based on the past participle form are usually prefixed with the negation prefix *un-*: *ungeachtet*, *unbeschadet*, *unerachtet* ‘notwithstanding’. Adpositions based on the present participle form are for example *betreffend* ‘concerning’ and *entsprechend* ‘according to’. Interestingly, most of the participial adpositions assign genitive case, irrespective of the original valency pattern of the base verb.

4.1.2 Dutch

Like in German, most of the simple adpositions in the basic vocabulary of Dutch (like *aan* ‘on’, *bij* ‘by’, *met* ‘with’, *na* ‘after’, *op* ‘at’, *over* ‘about’ etc.) developed

out of Germanic local adverbs (Stoett 1923; Van der Horst 2008). This adverbial origin is often still clearly visible in the older stages of Dutch:

- (58) *ende als die dieren hem sijn bi* (NB – Jacob van Maerlant’s *Naturen Bloeme*, 13th century Middle Dutch, quoted in Van der Horst 2008: 461)
 ‘and when the animals are with him’ (lit. ‘and when the animals him are by’)

The postpositional use of simple adpositions starts in the 16th century. Van der Horst (2008: 943) suspects a correlation with the deflection (=degeneration of inflectional marking) process and the reduction of the case system from the late Middle Dutch period onwards. The old distinction between dative (location) and accusative (direction) gets replaced by the use of simple adpositions like *in* or *op* as adposition (location) or postposition (direction).

An early example of *door* used as a postposition is here:

- (59) *Zoo ghaende de stat deurre zyn (wij) ghecommen tot den pallaisen van conynck Mathias van Hungheryen* (RC 154, 16th century Dutch, quoted in Van der Horst 2008: 944)
 ‘Walking through the town, we reached the palace of king Mathias of Hungary’

Incidentally, such uses of *door* are already found in late Middle Dutch (15th century), but the analysis is not always clear; relevant examples could also show the adverbial use of *door* (Van der Horst 2008: 695). From the 16th century onwards, however, it is quite obvious that Dutch has developed postpositions and the postpositional use spreads across other contexts and other adpositions in the centuries to come.

Circumpositions, too, emerged from the 16th century onwards. Earlier examples are confined to constructions with *te* or *tot* ‘to’ and another adposition/adverb in Middle Dutch. These constructions often indicate a direction.

- (60) a. *dat mi dat bloet ter noesen uut barstede* (JC – Jhesus collacien 173,128, quoted in Van der Horst 2008: 695)
 lit. ‘that me the blood to nose out spattered’ (that the blood spattered out of my nose)
- b. *om v te segghen enen anderen wech, die tote baylonien waert gaet* (RM – De reis van Jan van Mandeville 44, quoted in Van der Horst 2008: 696)
 lit. ‘to tell u another way that to Babylon ward goes’ (to show you another way leading to Babylon)

From the 16th century onwards, the number of circumpositions gradually increases and they are used in new contexts, like *van ... aan* with a temporal function ('from ... on').

- (61) *ende van sijnre ioncheit an leert hem wel te doen* (MA – Dialoog Salomon and Marcolphus (1501) 7, quoted in Van der Horst 2008: 944)
'and from his youth on teach him to behave well'

If we now turn to other complex adpositions, the situation is in some respects very similar to German and English. The canonical form [P₁ N P₂ / CASE] is quite old and it has always been productively used. Many of the formally complex adpositions of Dutch go back to this canonical structure. Table 8 lists the most frequent adpositions of this type in Present-Day Dutch. Most of them are quite old and have lost the determiner, some still have the contracted form [P₁ + inflected determiner] as their first element (*ten*), and *met het oog op* ('with an eye to, in view of') might be used without the determiner *het* in Present Day Dutch, too. There is, however, a very clear preference for the version with the determiner present.

As Stoett (1923: 82) points out, many early cases of (orthographic) univerbation can still be analyzed as 'nominal constructions' of the canonical type. The Middle Dutch adposition *inmidden* (now obsolete) is an example of such a process of univerbation. According to the MNW ('Middelnederlandsch Woordenboek'), *in midden* was an adverbial that could be used with a genitive complement or with a *van* PP. Then it became univerbated and could only be used as an adposition.

- (62) a. *In midden van desen paradise_{DAT} daer steet een boom* (MNW, Blisc. v. M. 183, Brabant, 1450-1470)
'In the middle of this paradise there stands a tree'
b. *Also hi quam in midden dier riuier_{GEN/DAT}* (1290, 'Vroegmiddelnederland woordenboek')
'When he came in the middle of the river' (when he reached the middle of the river)
c. *In midden den paradise_{DAT}* (MNW, Blisc. v. M. 254, Brabant, 1450-1470)
'In the middle of the paradise'

The example in (62a) shows the adverbial use of *in midden* with the adposition *van* and a dative NP. The status of example (62b) is less clear: *in midden* could be an adverbial or an adposition, the complement most probably is a genitive (equivalent to the modern German construction: *inmitten des Flusses_{GEN}* 'in the middle of the river'), but formally it could also be dative. For (62c), the interpretation is unambiguous: *inmidden* is an adposition, followed by a dative NP. This

Table 9: Most frequent complex adpositions in Present-Day Dutch (CHN, Corpus Hedendaags Nederlands).

SUB-PATTERN	EXAMPLES	FREQ. P.M.W.
[P1 (Det) N van]	<i>in plaats van</i> ‘instead of’	91
	<i>op basis van</i> ‘on the basis of’	82
	<i>ten aanzien van</i> ‘with regard to’	45
	<i>ten opzichte van</i> ‘in relation to’	44
	<i>op grond van</i> ‘on the basis of’	43
	<i>naar aanleiding van</i> ‘as a result of’	37
[P ₁ (Det) N P ₂]	<i>in verband met</i> ‘in connection with’	38
	<i>met betrekking tot</i> ‘with regard to’	38
	<i>in tegenstelling tot</i> ‘in contrast to’	36
	<i>in vergelijking met</i> ‘in comparison with’	29
	<i>in samenwerking met</i> ‘in cooperation with’	21
	<i>met het oog op</i> ‘in view of’	19

adposition became obsolete when the case system eroded. As a replacement, a new version of the canonical construction developed: *in het midden van* ‘in the middle of’, a complex adposition that is still in use today.

Early univerbation processes can also be found with the Proto-Germanic adposition *bi* ‘by’: *bezijden* ‘beside’, *bezuiden* ‘south of’, *beoosten* ‘east of’ etc. In Middle Dutch, these adpositions governed the dative case. They are a bit outdated nowadays, but they can still be used. Case markings got lost after the Middle Dutch period: *bezuiden de grote rivieren* means ‘south of the big rivers’ (with no case marking on determiner, adjective or noun). An even more reduced form can be found in adverbs like *binnen* ‘inside’ (from *be* + *inne*), *buiten* ‘outside’, *boven* ‘above’. They are used as adpositions from the 12th century to the present day.

Compounding is found in many univerbated combinations of two adpositions or of an adverb and an adposition. An example is *bovenop* ‘on top (of)’, a compound of *boven* ‘above’ and *op* ‘on, at’. Similar compounds are: *achteraan* ‘behind’, *achterin* ‘in the back’, *binnenin* ‘inside’, *rondom* ‘(a)round’, *tegenover* ‘against’, *voorbij* ‘past’ etc. This group is quite old: the first examples for *voorbij* (from *voor* ‘before’ + *bij* ‘near to’), for example, are found in the 13th century. It can be used as a preposition in contexts like *ze liepen voorbij het huis* ‘they walked past the house’. Univerbation on the basis of (fixed) word groups is also possible, e.g. *dankzij* ‘thanks to, lit. thank be’ (1764, WNT).

The adverbial *-s* is found not only in German adpositions, but in Dutch adpositions, too. It is used not only denominal (*ondanks* ‘in spite of’, from *ondank* ‘ingratitude’), but also on the basis of adjectives (*langs* ‘along’, from *lang* ‘long’) and deverbal (*volgens* ‘according to’). *Volgens* is most probably derived from

the present participle *volgende* (> *volgen(d)s*) of the verb *volgen* ‘follow’. Adpositions with the adverbial *-s* occur from early Middle Dutch onwards. Similar is the by now almost obsolete adposition *nopens* ‘with regard to’ (from *nopen(d)* ‘impel(ling)’). Past participles are also possible: *behoudens* ‘except for’, from *behouden* ‘preserve, preserved’. *Wegens* ‘because of’ is derived with the adverbial *-s* from a dative plural form of the noun *weg* ‘way’.

Participles could be used as adpositions ever since the Middle Dutch period. Stoett (1923) and Van der Horst (2008) list many examples, some of which are outdated nowadays, like *durende* ‘during’, replaced by *gedurende* in the 16th century (*gedurende* is a present participle of the now obsolete verb *geduren* ‘last, continue’). In some cases, the past participle can be used as an adposition (*gezien* ‘in view of’, *ongeacht* ‘irrespective of’), but it is especially the present participle that has gained some productivity in Dutch: *aangaande* ‘concerning’, *belangende* ‘concerning’ (obs.), *betreffende* ‘regarding’, *hangende* ‘pending’, *staande* ‘during, lit. standing’, *niettegenstaande* ‘notwithstanding’ etc. Recent newcomers in this category are *gaande* ‘during, lit. going’ and *lopende* ‘during, lit. running’ (Van der Horst and Van der Horst 1999: 190).

- (63) *Maar gaande de avond worden de supporters stiller.* (2003, OpenSoNaR)
 ‘But during the evening (or: as the evening progresses) the supporters become quieter.’

The pattern has, however, surely never been as productive as its equal in English (see below for the English participles in *-ing*). Table 10 summarizes the formation patterns of complex Dutch adpositions and their periods of productivity.

Table 10: Major formation patterns of complex adpositions in Modern Dutch and their periods of productivity (based on dictionaries and on Van der Horst 2008).

PATTERN	PRODUCTIVE PERIOD	EXAMPLES
[P ₁ (Det) N P ₂]	from 12th century onwards, increasingly productive;	<i>aan de hand van</i> ‘on the basis of’, <i>in (de) plaats van</i> ‘instead of’
[P ₁ (Det) N CASE]	Middle Dutch; used until 19th century	<i>in midden</i> NP _{GEN/DAT} ‘in the midst of’ <i>met behulp</i> NP _{GEN} ‘with the help of’
[N. -s] [V/Part. -s] [A -s]	13th century – 17th century	<i>ondanks</i> ‘in spite of’, <i>middels</i> ‘by means of’ <i>volgens</i> ‘according to’, <i>nopens</i> ‘with regard to’ <i>langs</i> ‘along, past’
[be- N] [be- P/Adv]	12th century – 16th century	<i>bezuiden</i> ‘south of’, <i>bezijden</i> ‘beside’ <i>binnen</i> ‘inside’, <i>buiten</i> ‘outside’
[V _{pres.part}] [V _{past.part}]	from 13th century on	<i>aangaande</i> ‘regarding’, <i>betreffende</i> ‘concerning’ <i>gezien</i> ‘in view of’, <i>ongeacht</i> ‘irrespective of’

4.1.3 English

English, like the other West Germanic languages, has inherited most of its simple adpositions from Proto-Germanic, sharing wide-spread cognates like *of(f)*, *in*, *for*, *to*, *with*, etc. with the other languages in the family. In some cases, the shared adpositions or adverbs have been modified by (pseudo-)prefixes, as in the case of *about*, from the shared Germanic *but* ‘outside’ (cf. Middle Low German *būten*) or *between* from the shared Germanic root for ‘two’. A much smaller set has developed from complex adpositions more recently (see further Section 4.3 below).

With respect to the latter, $[P_1 \text{ (Det) N } P_2 / \text{GEN}]$ is the most productive formation pattern in English today (as in the other West Germanic languages), and the one that has been productive the longest. Table 11 shows the most frequent cases in Present Day English.

Table 11: Most frequent complex adpositions in present-day English (BNC).

SUB-PATTERN	EXAMPLES	FREQ. P.M.W.
[P ₁ N of]	<i>in terms of</i>	100
	<i>at the end of</i>	94
	<i>in front of</i>	61
	<i>as a result of</i>	51
	<i>as part of</i>	48
[P ₁ N P ₂]	<i>in relation to</i>	47
	<i>in addition to</i>	34
	<i>in accordance with</i>	20
	<i>in connection with</i>	15
	<i>with respect to</i>	13

Early cases of this type can already be observed in Early Old English, where the genitive-marked NP preceded the relational noun forming the core of the expression, e.g. [*bi* NP_{GEN} *sidan*] ‘by ... side’ (cf. Rissanen 2006) or [*on* NP_{GEN} *gemang*] ‘on ... crowd’ (see further Section 4.3 below); from the 15th C. onwards, an increased productivity of the modern (analytic) form $[P_1 \text{ (Det) N } P_2]$ can be observed. Another pattern still productive today is [Adverbial + P], where the adverbial element can be an adjective/adverb or a participle (see Table 1 for examples). This pattern has been productive at least since the 14th C., i.e. after English had lost its case system and most other remnants of syntheticity.

Again, as in the other West Germanic languages, formation through affixation has always been the exception in English. There are a number of cases formed with the adverbial genitive -s between the 11th and 17th century, and the suffix *-ward*, that still productively derives adverbs today, is responsible for a few

adpositions of which only *toward* survives in Present Day English (PDE). From Old English times to the 19th century, there were two productive pseudo-affixes, *a-* and *be-*, which attached to stems of all kinds to produce adpositions like *above*, *again*, *below*, *before*, etc; these are discussed further in Section 4.3 below.

In contrast, the formation of adpositions by conversion from present participles, which started in the 14th century, with *notwithstanding* (1325) and *during* (1385) as the earliest documented cases, continues to be highly productive in PDE (unlike in German). Conversion from past participles or adjectives, on the other hand, is marginal in English. Examples of the former are *given* and *granted*, examples of the latter are *near* and *opposite*. Of these, *near* is an interesting, but highly exceptional case in that it has retained its ability to be inflected for comparative and superlative forms even in its adpositional uses:

- (64) a. *We stayed in a hotel **near** the railway station.* [BNC FEM]
 b. *Mr Ozberk has now been moved to Pentonville Prison which is **nearer** the airport.* [BNC A59]
 c. *The building **nearest** the jetty was the post office-cum-shop.* [BNC CKF]

This case suggests that conversion from adjectives could be a highly productive process in PDE, which raises the question why it is not – especially given the high productivity of conversion between word classes that characterizes English in general.

Table 12 summarizes the formation patterns of complex English adpositions and their periods of productivity.

Table 12: Major formation patterns of complex adpositions in Modern English and their periods of productivity (based on the OED).

PATTERN	PRODUCTIVE PERIOD	EXAMPLES
[P ₁ (Det) N P ₂]	at least 9th C. – PDE (current set since 15th C.)	<i>in terms of, in relation to, in the case of, in accordance with, with regard to</i>
[<i>be-X</i>], [<i>a-X</i>]	11th C. – 19th C.	<i>afore, around, beyond, below</i>
[AD]/ADV P [V _{pres.part} P] [V _{past.part} P]	14th C. – PDE	<i>due to, prior to, apart from, near to according to, consisting of, arising from based on, compared with, related to</i>
[ADV -s(<i>t</i>)]	11th – 17th century	<i>alongst (obs.), beyonds (obs.), amongst, besides, towards</i>
[P/ADV -ward]	11th C. – 16th C. (for prepositions), 11th C. – PDE (for adverbs)	<i>downward (now adv. only), overward (obs.), toward</i>
[V _{pres.part}]	14th C. – PDE	<i>including, concerning, regarding, pending</i>

4.2 The formation of complex adpositions of the type [P (Det) N P]

The properties distinguishing complex adpositions from free syntactic combinations, discussed in Section 3 above, are all related to formal (syntagmatic) fixedness. They are generally (and quite naturally) interpreted as the result of a gradual diachronic process leading from free combinations to fixed expressions. During this diachronic process, the individual elements of the syntagm successively become less independent, gradually coalescing into a holistic and idiosyncratic structure. This loss of autonomy of the individual elements is usually accompanied by decategorialization. Evidence of decategorialization is difficult to diagnose in the case of the P₁ and the P₂, as they already start out as monomorphemic function words; it is easier to find if one looks at the N, as the N loses the ability to inflect and be modified (see Section 3). Table 13 summarizes the relevant properties of this process, most of them have been discussed in more detail in Section 3.

For German, it can be said that generally, all recently emerged complex adpositions follow the canonical pattern [P₁ N P₂ / CASE] and originate from a regular syntactic structure. In the most productive sub-pattern [P₁ N *von* / GEN], the genitive case and its alternative with the adposition *von* can be traced back to the originally postnominal genitive attribute. This attributive genitive noun phrase is then reanalyzed as a genitive complement of a new complex adposition (see Lehmann 1998 for the detailed account of the reanalysis).

Most of the characteristics given in Table 13 also apply to Dutch, the result of the process being a lexical unit that functions as an adposition. But there are also differences compared to the properties of complex adpositions in German. First of all, the case system has been gradually abandoned in Dutch over the past 500 years. As a consequence, the [P N P] structure became more dominant and the degree of fixedness of the elements also increased. The adposition *van* replaced the genitive case in the P₂ position; and the morphological distinction between dative and accusative got lost, too. While German retained the possibility of alternation of genitive with an adposition, the second adposition became obligatory in Dutch.

A prototypical example of the development summarized in Table 13 is *met behulp van* ('with the help of'). This complex adposition emerged in the 17th century (Vranjes 2012). The noun *behulp* 'help' became obsolete as an independent lexeme in Dutch, but it still exists in the complex adposition *met behulp van*.

Table 13: From free syntactic combination to complex adposition.

FREE SYNTACTIC COMBINATION [P ₁ (DET) N P ₂] _{PP}	LEXICAL UNIT [P ₁ N P ₂] _P
P ₁ / P ₂ can vary	P ₁ / P ₂ are fixed
N can be modified	N cannot be modified
N can have determiners	N cannot have determiners
N can inflect	N cannot inflect
N can be replaced	N cannot be replaced
N exists as an independent lexeme	N exists only in the CA
P ₂ /CASE + complement may be omitted	P ₂ /CASE + complement are obligatory

In the 17th century, we still find variation. (65) shows an early example of the construction:

(65) *Hy perst de misdaad uit d' onschuldigen ... **Met dat vervloekt behulp** van pyningen* (OUDAAN, H. Broederm. 54 [1673])

'He presses the crime out of the innocent... with the cursed help of torture.'

Here, the noun *behulp* is still autonomous. It is accompanied by an article (or a demonstrative) and modified by an evaluative adjective, which is no longer possible in modern Dutch. In the early stages there is also variation with regard to the first adposition: *met* or *door* (we find the same variation in German; *mit Hilfe von*, *durch Hilfe von*). And there is variation with regard to the form of the noun, too: *behulp* and *hulp* are both used. In the early 20th century, *hulp* became the only form of the noun for 'help'. The form *behulp* became outdated, with one exception: it is still used in the complex adposition (see above, Section 3.4, for *met (de) hulp van*). The second adposition (*van*) competed with the genitive for a long time. Even in the 19th century, we still observe the use of the genitive instead of the second adposition.

(66) *Weldra ontdekten wij, **met behulp der lantaren**, den Duitscher ..., die ... in een moddersloot was gezakt* (V. LENNEP, Rom. 4, 83 [1840])

'We found, with the help of the lantern, the German, sunk into a mud ditch.'

We can assume this genitive (*der lantaren*) to be a bit archaic and limited to written Dutch, at that time. With respect to *met behulp van*, it can be seen that the form without a determiner has been prominent from the very beginning and it is the only possibility in modern Dutch.

As mentioned above, the loss of the determiner is a good indicator for complex adpositions, but it definitely is not a condition. Many complex adpositions are constructed with the determiner and there are no signs that this will change. Examples are: *aan de hand van* ‘on the basis of’, *in de loop van* ‘in the course of’, *in het kader van* ‘in the framework of’. Other adpositions have a clear preference for the omission of the determiner: *naar aanleiding van* ‘as a result of’, *in verband met* ‘in connection with’, *met betrekking tot* ‘with regard to’. For most of the complex adpositions the preferences are quite stable, but for some (less frequent) units, we do find variation. An example is *uit het oogpunt van* ‘from the viewpoint of’. The OpenSoNaR corpus has the following distribution:

- (67) *uit het oogpunt van* – 182 hits
uit een oogpunt van – 86 hits
uit oogpunt van – 106 hits

The tempting interpretation of such numbers is one in terms of directional change (in accordance with Table 10 or Table 12), but corpus data do not allow us to predict whether the variant without determiner will ‘win’ in the future.

In English, the process summarized in Table 13 above is also plausible in broad outline, but with one minor and one major caveat. The minor caveat is that not all formal changes can be observed for all cases. In particular, there are units widely agreed upon to be complex adpositions that contain a determiner and show no signs of losing it, such as *on the part of*, *at the expense of*, *with the exception of*, *in the name of*, etc. These cases suggest that the loss of the determiner is a common, but by no means an obligatory stage in the development of complex adpositions.⁶ There are also cases where variation between a variant with and one without a determiner has lasted for quite some time – such as *at (the) risk of*, *in (the) light of*, *by (the) force of*, and others. Of course, this does not invalidate the model in Table 13 in principle – the cases of variation might still resolve themselves in favor of the determiner-less variant sometime in the future, but this is pure speculation. Even if the alternation between cases with and without a determiner persist, however, this would not be a major problem for the model in Table 13 (or any other model of diachronic change), as we would not

⁶ As one of the editors has pointed out (and we fully agree), the loss of a determiner is likely to be indicative of a greater degree of grammaticalization for a given sequence – *at risk of* is more grammaticalized than *at the risk of* (Borillo 1997); however, the same does not necessarily hold across sequences – *at risk of* is not necessarily more grammaticalized than *in the light of* (Fagard & De Mulder 2007).

expect diachronic processes to operate without exception (cf. our discussions of variation in Section 3 above).

The major caveat is that – as already pointed out by Hoffmann (2005) – for many complex adpositions there is no plausible evidence of a development from a free combination to a fixed complex adposition – they appear in the written record abruptly in the form [P1 N P2]. At the very least, this suggests that complex prepositions may emerge by analogy to existing cases that did emerge in the way we outlined above (Hoffmann 2005: 86); we will expand on this possibility in Section 5. It should be noted, additionally, that the 15th century, when the current set of complex adpositions in English began to emerge, is characterized by a number of loan calques from French, such as *in lieu of* < *au lieu de* (cf. Traugott and Schwenter 1995). These did not always have a transparent determiner in French at the time of borrowing and were calqued accordingly, establishing the pattern [P₁ N P₂] directly.⁷

4.3 The further development of adpositions of the type [P (Det) N P]: From complex to simple(r)

Most of the complex adpositions of the type [P1 (Det) N P2] are relatively young in the languages under discussion here, so we must be careful with claims about preferred paths of change that they might take. Still, there is enough evidence for a tentative assessment of the importance that possible paths have in German, Dutch and English.

Let us begin with a strong claim about a preferred grammaticalization path. For English, Allerton (2008: 53–55) predicts a regular path along four stages:

(68) A hypothetical grammaticalization path from complex to simple prepositions (English)

I	P ₁	NP(Det N ₁)	P ₂	
II	P ₁	NP(N ₁)	P ₂	(Loss of Det in the NP)
III		NP(N ₁)	P ₂	(Loss of P ₁)
IV		NP(N ₁)		(Loss of P ₂)

In the broader context of the West Germanic languages, we have to amend this hypothetical path insofar as the function of P2 could also be served by geni-

⁷ *au lieu de* (Modern French) originally appears as *en lieu de* in Medieval French (see e.g. Fagard & De Mulder 2007, Fagard to appear).

tive case marking on the nominal complement, as is typical for German. In this case, Stage IV would consist in a loss of the genitive case in favor of the dative or accusative case typically assigned by German adpositions (see Section 2.5). We also have to take into account the more flexible word order of German. In particular, in the first two stages the genitive-marked nominal complement can precede N1 in free combinations (e.g. *meiner Mutter Auto* ‘my.GEN mother car’), but this would be expected to change as speakers start treating the combination as an adposition. Thus, we get the following (still hypothetical) path:

(69) A hypothetical grammaticalization path from complex to simple prepositions

(West Germanic) (word order is indicated by arrows)

- | | | | | | |
|-----|-------------------|--------------------|-------------------|------------------------|------------------------------------|
| I | $P_1 \Rightarrow$ | Det N ₁ | \Leftrightarrow | P_2/gen | |
| II | $P_1 \Rightarrow$ | N ₁ | \Leftrightarrow | P_2/gen | (Loss of the determiner in the NP) |
| III | | N ₁ | \Rightarrow | P_2/gen | (Loss of P1) |
| IV | | N ₁ | | \emptyset or acc/dat | (Loss of P2/genitive case) |

In German, where the ‘canonical’ pattern with an abstract, often deverbal noun as lexical nucleus has been attested since the earliest documented stages of the language, there are a number of adpositions whose development has roughly followed this amended grammaticalization path. Let us illustrate this in detail for the adposition *wegen* (‘because of’), following the reconstruction in Szczepaniak (2009: 98–101).

The original structure in which *wegen* occurs is $[P_1 [NP_{GEN} N]_{NP}]$, corresponding roughly to Stage I/II above (this and all following examples are from Szczepaniak 2009):

(70) *gebruder von der mutter wegen* [DWB, Makrelius, Pommern 1,65; 17th C.]
 gebruder [*von*_P [[*der Mutter*]_{NP,GEN} *Wegen*]_{NP,DAT}]
 ‘brothers from the mother.GEN way.PL.DAT’

Note that the original syntactic structure accounts for the fact that the case assigned by the grammaticalized adposition is the genitive: the original possessive genitive of the attributive noun phrase (*der Mutter*) is reanalyzed in later stages as a genitive of the complement of the adposition.

In the source structure exemplified in (70) above, the noun *Weg* ‘way’ refers to the line of descent; the genitive noun phrase specifies the ancestor (here *Mutter*). In the course of the Middle High German period, the expression undergoes context expansion and semantic bleaching: the structure comes to be used with N_{GEN} referring to human referents as sources of communicative contents as well

as legitimation sources, whereby the noun *Weg* loses its original semantics, cf. e.g. (71), where N_{GEN} refers to the person who originally expressed his gratitude:

- (71) *(die boten) dem marggraffen von seinen (ihres herren) wegen vil dinst*
 (the messenger) the.DAT margrave.DAT **von** he.GEN (their lord.GEN) **wegen**
 much service
und danck sagten [DWB, Decamerone 104, Keller]
 and thanks said
 ‘(the messengers) said to the margrave in the name of their lord much
 service and thanks’

The crucial step of structural reanalysis likely occurred in such contexts with human referents (Szczepaniak 2009: 99, Braunmüller 1982: 205): The noun *Weg* is reanalyzed as a part of a fixed unit, i.e. the circumposition *von ... wegen*, and the N_{GEN} is reanalyzed as a complement of this adposition:

- (72) $[von_P [N_{\text{GEN}} \textit{Weg}_N]_{\text{NP}}]_{\text{PP}} > von_P [N_{\text{GEN}}]_{\text{NP}} \textit{wegen}_P]_{\text{PP}}$

This still corresponds to Stage II in the schema above. During the Early New High German period, the complex adposition *von ... wegen* takes on a more abstract causal meaning; the original restriction of N_{GEN} to human referents is gradually lifted and all kinds of noun phrases come to combine with the adposition *von ... wegen*, cf. e.g. (73):

- (73) *gesche ez, daz kene der vorgenannten stete **von brandes wegin** ... des vorgeschriebenen geldes neit ingebe* [DWB, Waldeck. Urkundenbuch 135]
 ‘it may happen that none of the aforementioned cities **von fire.GEN wegen**
 [= because of fire] will give the prescribed money’

Additionally to this semantic change, the word order changes from a circumposition to a complex adposition, i.e., $[von_P [N_{\text{GEN}}]_{\text{NP}} \textit{wegen}_P]_{\text{PP}}$ becomes $[[\textit{von wegen}]_P [N_{\text{GEN}}]_{\text{NP}}]_{\text{PP}}$

- (74) **von wegen** *syner koniglichen durchluchtikeit* [DWB, Urkundenbuch d.st. Arnstadt 249]
von wegen his.GEN royal.GEN highness.GEN
 ‘from his royal Highness’

This is a crucial step in the grammaticalization of this adposition. In the next step, the P1 *von* disappears (Stage III in the schema above). This is the form still found in present-day German, in particular in formal registers. In colloquial usage, and

increasingly in formal registers, *wegen* is used with dative-marked complements (as would be expected for a simple adposition), showing that it is the process of moving from Stage III to Stage IV):

- (75) *Ich habe hier einige Fachleute, die möchten das auch gern wissen,*
 I have here some specialists, who also would like to know this,
schon wegen dem Preis.
 ‘yet **wegen** the.DAT price.DAT’
 [Brief von Ernst G. an Irene G. vom 27.11.1939, Feldpost-Archive mkb-fp-0270]

The developmental path of *wegen* seems to be one typical possibility for German. Other examples are the adposition *neben* (‘next to, beside’) which emerged in the Old High German period (ca. 750–1150) from the complex adposition *in eban(i)* ‘in (the same) plane’ (cf. Lehmann and Stolz 1992; Behaghel 1924), as well as several adpositions that developed during Middle High German, such as *in kraft* > *kraft* ‘by virtue’, *nach laut(e)* > *laut* ‘according to’, and *an ... stat/stete* > *anstatt* > *statt* ‘instead of’, which includes a stage of (orthographic) univerbation between P1 and N1 observed fairly regularly in German but not part of Allerton’s schema (we will return to these cases below). However, it should be noted that even many of the old complex adpositions show no signs of moving beyond Stage II – for example, the circumposition *um ... willen* ‘for ... sake’ started out just like *von ... wegen* but has retained its original structure to the present day.

In Dutch, we find a development quite parallel to that of German *wegen*, both formally and semantically:

- (76) a. *van NP_{GEN} wege(n)_{DAT}*
 Middle Dutch: *van des graven weggen* – ‘on behalf of the count’, later also causal:
 ‘because of the count’
- b. *van wege(n) NP_{GEN/DAT}*
Als ’er een Spanjaarde Poza geheeten, van weeghe des Landtvooghds, komt roepen, dat zy niet schieten zouden [WNT, HOOFT, N.H. 364 (1642)]
 ‘When a Spaniard named Poza, on behalf of the bailiff, announces that they should not shoot.’
 (the complement of *van weeghe* has genitive case here, but dative is also possible with this meaning)
- c. *vanwege NP*
De twee koeikoopers ..., die goei zin hebben vanwege de winst [WNT, DE MAN, *Kl. Wereld* 171 (1932)]
 ‘The two cow buyers are in a good mood because of the profit’

- d. *wegens* NP (from the 16th century on)
*De Eland ..., hebbende een huydt **wegens** de hardigheyt ondoorsnijdelijk*
 [WNT, COMENIUS, Deure d. Taalen 44 (1666)]
 ‘The elk ..., having a skin that cannot be cut because of its hardness.’

According to the WNT (‘Woordenboek der Nederlandsche Taal’), the form *wegens* emerged in the course of the 16th century next to *vanwege* to which it corresponds in meaning. Nowadays they are still used next to each other. The OpenSoNaR corpus has 50,418 hits for *wegens* and 34,326 for *vanwege*.

There are more examples of circumpositional structures becoming a (complex) adposition in Dutch. Take the German circumposition *um ... willen* (*um der Liebe willen* ‘for love’s sake’). This construction had an equivalent in Dutch: *om* NP_{GEN} *wille* that was replaced by the complex adposition *omwille van* NP from the 19th century onwards. In Present-Day Dutch, *omwille van* is the only possibility, but even in the first half of the 20th century, the combination with a genitive complement (77c) is still found:

- (77) a. **omwille van de kinderen** bleven zij samen
 ‘for the children’s sake, they remained together’
 b. **omwille van het weer** bleef hij thuis
 ‘because of the weather he stayed at home’
 c. *De nationaal-socialistische antipapist ds. A. van Hoogenhuyze wekt voortdurend de suggestie, alsof in Oostenrijk de protestanten **omwille des** geloofs vervolgd zouden worden* [WNT, *Maasbode* 6 Juli 1937]
 ‘The National Socialist anti-papist Rev. A. van Hoogenhuyze constantly suggests that in Austria Protestants would be persecuted for religious reasons’

In example (77a) the PP indicates a purpose or a goal; in (77b) it is used to indicate the cause. The shift towards the causal usage is typically found in the south of the Dutch language area (in Flanders). De Vogelaer and Vandenberghe (2014) present a detailed analysis of the history of this complex adposition in which they focus on the semantics and on differences in use between Flanders and the Netherlands.

While in German and English (see below), this development seems to have occurred only with structures of the type [P [[NP_{GEN}] N]_{NP}], Dutch also has complex adpositions that emerged from typical circumpositions consisting of two adpositions (cf. Section 2.4 above):

- (78) local:
 a. *Hij rijdt **van de weg af**.* (direction/path)
 ‘He’s driving off the road.’

- b. **Vanaf** de berg kun je het strand goed zien. (place)
 ‘From the mountain you can clearly see the beach.’
- temporal:
- c. **Van** morgen **af** gaat hij met de trein.
 ‘From tomorrow he will be taking the train.’
- d. **Vanaf** morgen gaat hij met de trein.
 ‘From tomorrow he will be taking the train.’

There seems to be a slight difference in meaning between the circumpositional use and the preposition when used to indicate locality. While the circumposition indicates a direction, the preposition refers to a place. It is this second possibility that was extended to the temporal domain. However, semantically, there is no systematic difference between the circumposition and the preposition when used temporally. They both refer to a (starting) point/period on the timeline.

A more recent example from Dutch whose initial and final stages correspond to Allerton’s path shows that matters are more complex. It concerns the recently emerged adposition *richting* (‘towards’, lit. ‘direction’) shown in (79), whose development is described in some detail by Van der Horst and Van der Horst (1999: 187-189; see also Stathi forth. for the development of the German analogue *in Richtung auf*):

- (79) a. *de president heeft wat recht te zetten* **richting** het leger (NRC 2003; Van der Horst 2008: 1833)
 ‘the president has some rectification to make **towards** the army’
- b. *het boek is bedoeld als een uitdaging* **richting** historici (Leids Dagblad 1989; Van der Horst (2008: 1833))
 ‘the book is intended as a challenge **to** historians’

This adposition started out as a canonical complex adposition, *in de richting van* (80a), corresponding to Stage I. Evidence of the loss of the determiner (Stage II) appears in the historical record in the late 20th century (80b). At this point, the development diverges from the path in (69) above: it is not P1 that is lost first, but P2 (80c), and although a variant with P2 but without P1 also emerged, it is extremely rare (one example from a Surinam newspaper is shown in 80d). Additionally, we find uses that retain the determiner but have lost P2 (80e). Example (80e) also has the bare noun used adpositionally (Stage IV).

- (80) a. *van daar gaat de grens westwaarts* **in de richting van** die scheiding tot aan het perceel sectie B [CHN, Juridisch corpus, 1895]
 ‘from there the border goes westward in the direction of that partition up to the lot section B’

- b. *Bijna alle economische aanwijzingen gaan al lange tijd in richting van een veel hogere dollarkoers.* [CHN, NRC Handelsblad, 1994]
‘Almost all economic indicators have been moving towards a much higher dollar exchange rate for a long time.’
- c. *VS overwegen weer stapje in richting Cuba.* [CHN, NRC Handelsblad, 1994]
‘The United States considers taking another step in the direction of Cuba’
- d. *De vrachtwagen kwam van de Commissaris Weytinghweg en ging richting van de Lalarookhweg.* [CHN, Omgekomen bromfietser is 19-jarige Aroen Sabar, 2012]
‘The truck came from the Commissaris Weytinghweg and went towards the Lalarookhweg’.
- e. *Een wagen die in de richting binnenland reed, raasde er om nog onduidelijke reden door de middenberm en reed frontaal in op het verkeer richting kust.* [CHN, Weer auto door middenberm E40, 2008]
‘A car driving towards the back-country crashed through the median strip for unclear reason and drove head-on into traffic towards the coast’

Absent a systematic study of the development of complex adpositions in Dutch, it remains unclear what the typical path from Stage I to Stage IV looks like – certainly, the existence of adpositions at both ends of the path should not be taken as evidence for the path itself.

In English, evidence for the grammaticalization path proposed by Allerton is extremely difficult to find. He cites a possible development of *north of* from *the north of* (Stage I to Stage III), and possible developments of [*inside/outside* N] from [*inside/outside* of N] and [*opposite* N] from [*opposite to* N] (Stage III to Stage IV). The only case he mentions as evidence for all four stages – though without providing data – is the potential adposition *courtesy*.

In the case of *courtesy*, a corpus analysis bears out this claim, albeit only in broad terms. The earliest citation we were able to find is from 1652 and does not have a determiner (81a), the earliest citation with a determiner is from 1701 (81b) (both via the Google Books search):

- (81) a. *Where a man shall be tenant by courtesy of a rent...*
- b. *An Estate for Life, by the Courtesy of England.*

This suggests that the version without the determiner (Stage II) may in fact have preceded the one with a determiner (Stage I) – at the very least, it seems that the two “stages” coexisted from the inception of the expression on (they can

frequently be found within one and the same source, as in (81c, d) (from an 1822 volume of bankruptcy reports):

- (81) c. *There shall be a tenancy **by the courtesy of** a trust...*
 d. *...he shall hold the land **by courtesy of** England during his life...*

The variant with the determiner seems to disappear gradually, however – it is not found in the Corpus of Historical American English (COHA) after 1942, while the variant *by courtesy of* is the dominant variant to the present day, which is at least compatible with the idea of a reduction from Stage I to Stage II over time.

The variant *courtesy of* (Stage III) first appears 1903 in an attribution of an image to a copyright holder (82a), and as part of full clauses from 1927 onward (82b, c):

- (82) a. *From a drawing by Lucius Hitchcock **Courtesy of** Collier's Weekly.* [1903, COHA]
 b. *After a fitting period of exhibition in the Ministry of Public Education at Rome, the duke will be restored to his duchy, **courtesy of** Signor Raphael Sanzio.* [1927, COHA]
 c. *They are devoted to the standard heroine of romance, **courtesy of** De Mille.* [1930, COHA]

The fully reduced variant (Stage IV) is first found in 1933, again, in an attribution of an image to a copyright holder (83a), and in the following remains restricted to the telegraphic style typical for this context. However, very rare uses in regular grammatical contexts can be found more recently in the Corpus of Current American English (COCA), both in this copyright-attributing function (83b) and with a more general meaning “due to” (83c, d):

- (83) a. *Henry Harris, examining the weekly issue of the Sansbury Times ..., observed beneath the cut the minute italics: **Courtesy** Interstate Light & Power and allowed himself to smile.* [1933, COHA]
 b. *[I]mages in this article are **courtesy** the Jet Propulsion Laboratory* [COCA 2015529]
 c. *According to Pop, my Flanagan blood – Catholic as Communion wine – was corrupted at the cellular level by an infusion of Protestant DNA **courtesy** the Lowells, my mother's northern Anglo-Irish tribe.* [COCA 4040946; Elisabeth Kelly, Apologize, Apologize!, 2010]
 d. *It's an amazing string of winter fishing, all **courtesy** the hump of high pressure that had been sitting over the state.* [COCA 3049478; Brian Hoffmann, The Fishing Report, SFGATE, 1999]

Thus, the grammaticalization path proposed by Allerton is possible in English. However, unlike in German, it appears to be extremely marginal. This is suggested not only by the marginal status of *courtesy*, but by the fact that there are at most a handful of adpositions in English that have potentially grammaticalized along this path – namely the words *inside*, *outside* and *opposite* mentioned by Allerton. We can almost certainly exclude *opposite* from this group, as expressions like *on/at (the) opposite of* are not documented. Since the word was borrowed simultaneously as a noun and an adjective at the end of the 14th century, it is thus likely that the adposition *opposite* derived from the sequence $[[opposite]_{ADJ} to]$. For *inside/outside*, the developmental path described by Allerton is at least possible, although further research would be necessary to confirm this (they, too, could have developed from adjectival or adverbial uses).

Even if *inside* and *outside* did develop along this path, they would be the only two well-established cases in English. Whether the path will become more important in the future, i.e. whether other complex adpositions follow the development of *courtesy*, remains to be seen, but it seems doubtful at the moment.

A number of phenomena characteristic from many grammaticalization paths are missing in Allerton's schematic proposal, most importantly the possibility of a gradual reduction in form (e.g. from an adposition to a prefix or a clitic), of univerbation (in this case, of P_1 and N, and/or of N and P_2) and of phonological erosion.

As already hinted at in connection with the adposition *statt*, which developed from a complex adposition *an ... stat/stete* via the intermediate stage *anstatt* (which is also still found in Present-Day German), univerbation is quite typical for German. Other cases are *in Folge* > *infolge* 'as a result of', *an Hand* > *anhand* 'on the basis of', *von ... seiten* > *vonseiten* 'on the part of', *zu Liebe* > *zuliebe* 'for the sake of'.

This fits nicely into the general tendency in German, described by Erben (2000: 133): the relationship between syntactic constructions and word formation has shifted in the course of German language history in favour of word formation. German is, according to Erben, characterized by a trend towards univerbation.

In Modern English, univerbation in complex adpositions is extremely rare, with *by cause of* > *because of* the only clear case, i.e., the only case where the orthographic univerbation is accompanied by reduced stress (as in a number of simple adpositions of complex origin, cf. the discussion of the pseudo-prefixes *a-* and *be-* in Table 11 above, and in some more detail below). *By cause* is still widely spelled as *by cause* at the end of the 14th century, then the spelling *because* takes over, presumably reflecting a phonological weakening of the P_1 and thus univerbation, which then results in phonological erosion, with the form *cause of* first documented in the early 16th C.

In English, orthography is probably an even worse guide than in German and Dutch, when it comes to univerbation – certainly, the fact that a sequence

of morphemes is not spelled as a single orthographic word does not preclude the possibility that it is a single phonological word. Still, it is interesting that in addition to *because of*, there is only one other case of orthographic univerbation of a complex adposition in English: *in (the) stead of* > *instead of*. This adposition starts out as a form with a determiner (84a), but the variant without is documented shortly after (84b), the spelling *instead* first appears at the turn of the 17th century (84c), although the two spellings coexist until the late 17th century (occasionally, the variant *in the stead of*, where the determiner shows that no univerbation has taken place, can be found as late as the 18th century, cf. 84d). The use as an adverb, which provides clear evidence that univerbation has taken place, is first attested in the late 16th century, see (84e):

- (84) a. *Requiescant inpace in þe stude of benedicamus.* [c1225, Ancrene Riwe (Cleo. C.vi) (1972) 21]
 ‘Rest in peace in the stead of Praise the Lord’
- b. *Is dragon arerd of golde As in stude of is baner.* [1297 R. Gloucester’s Chron. (Rolls) 4358]
 ‘his dragon of gold raised as (lit. as instead of) his banner’
- c. *The translation leadeth the schollar as by the hand, or instead of his Master.* [1612 J. Brinsley Ludus Lit. viii. 101]
 ‘the translation leads the scholar, as though by the hand, or instead (or in place) of his master’
- d. *In the Stead of the Sanction was established the famous Concordat.* [1718, Free-thinker No. 56. 2]
 ‘instead (or in place) of the sanction, the famous concordat was established’
- e. *To rase Quite out thir Native Language, and instead To sow a jangling noise of words unknown.* [Milton Paradise Lost xii. 54, 1667]
 ‘to erase their native language completely, and instead to sow a jangling noise of words unknown’

Evidence of further phonological erosion is found in deliberate non-standard spellings like (85a,b):

- (85) a. *It’s the hard knock life for us. / Steada treated, we get tricked. / Steada kisses, we get kicked.* [Jay-Z, Hard Knock Life, 1998]
- b. *Steada’ talking and ranting and infighting, and dancing to a “one note samba”, it’s time to examine what you can come together around.* [Daily Kos, 2009]

Of course, a systematic analysis of such non-standard spellings or of large corpora of spoken colloquial language may well reveal further evidence for univerbation that has not made its way into the orthography, but it is certainly not a widespread phenomenon at this point.

In earlier stages of English, univerbation (both orthographic and phonological) can be observed more frequently, in contexts that are otherwise very similar to the case of German *wegen* above. For example, the (now simple) adposition *among* started out as an adpositional circumlocution [[on]_P [[]_{NP,GEN} *gemang*]_{NP,DAT}]_{PP} (86a), where the nominal complement of *gemang* stands in the genitive; as the circumlocution is reanalyzed as a complex preposition, the order of *gemang* and its complement is reversed, *gemang* forming a compound with *on* and the complement now being marked for dative, as expected for adpositions (86). In the next steps, there is phonological reduction to *onmang* (86c) and then to the modern *among* (86d):

- (86) a. ...*ond þæt halige treo him beforan ferian on feonda gemang* [Cynewulf
Elene, Early Old English]
and that holy tree him before carry in enemy.PL.GEN crowd
'and to carry that holy tree before him among the enemies'
- b. *Ne forleos mine sawle ongemang þam arleasum* [c1000, Metr. Ps. xxv. 9]
neither destroy my soul in.crowd the wicked.PL.DAT
'neither destroy my soul among the wicked ones'
- c. *Swa scep onmang wulfen.* [c1160, Hatton G. Matt. x. 16]
'as sheep among wolves'
- d. *and com vt forþrihtes; among alle his cnihtes* [1250, La3amon Brut 17742]
'and come out forthright, among all his knights'

Other adpositions that emerged in this or similar ways are *on middan* > *amid*, *at foran* > *afore*, *at hindan* > *athinden* (obs.), *on loft* > *aloft*, *be [by] æftan* > *baft* 'behind' (obs.), *be healfe* > *bihalve* 'beside' (obs.), and *bi sidan* > *beside* in Old English, and *on cross* > *across*, *in/on brode* > *abroad*, *on length* > *alength* 'along-side' (obs.), and *on side* > *aside* in Middle English. The (now obsolete) phonologically reduced form of the adposition *on* gave rise to further examples well into the Early Modern English period, for example *a breast* > *abreast*, *a stern* > *astern*, *a top* > *atop*, and *a stride* > *astride*.

While not exactly exceptional, this development was not wide-spread in the early history of English either, although this is perhaps due to the fact that the pattern P N P itself was not as productive as it became later. Instead, beginning in the Old English period, we can observe the emergence of two (pseudo-)prefixes for the formation of adverbs, many of which doubled as or developed into

adpositions: [*a*-[]_{N/ADV}] and [*be*-[]_{N/ADJ/ADV}]. This development seems to be based, on the one hand, on the emergence of unstressed variants of the adpositions *on* ‘in’ and *bi* ‘by’, i.e. *a* and *be*, and on the other hand on the existence of a number of adpositions of the type just mentioned. These prefixes each yielded a number of adverbs/adpositions, such as *afore*, *again*, *anunder* (obs. ‘under’), *anuppe* (obs. ‘on, upon’), *about*, *above*, *aninne* (obs. ‘within’), *around*, *atwixt* (obs. ‘between’), *alow* (obs. ‘below’), *atween* (obs. ‘between’), *ahind* (obs. ‘behind’), *aneath* (obs. ‘beneath’), *betwixt*, *be-east* (obs. ‘to the east of’), *besouth* (obs. ‘to the south of’), *benorth* (obs. ‘to the north of’), *beyond*, *below*, *bewest* (obs. ‘to the west of’) and *bedown* (obs. ‘below’). Most of these adpositions are derived from simpler adverbs, with the prefixes adding the meaning *at/near*, for example *about* from the adverb *but* ‘outside’, *alow* from *low* or *besouth* from *south*; however, some are also derived from nouns such as *around* from *round* (‘circular band’), or from stems found with the other prefix, such as *bestride* from (*a*)*stride*. Both prefixes ceased to be productive around the middle of the 19th century.

Dutch does have univerbation of adpositions, but, like English, only incidentally. An example is *inzake*. This word was univerbated on the basis of the former complex preposition *in (de) zake van* ‘in the case of’ around 1900 (another case is *ingevolge* ‘in accordance with, by virtue of’).

- (87) a. *een van de opvallendste beslissingen **inzake** mensenrechten in jaren*
[2006, OpenSoNaR]
‘one of the most striking decisions **on** human rights in years’
- b. *werknemers die werkloos zijn gesteld **ingevolge** het faillissement van hun onderneming* [2009, OpenSoNaR]
‘employees who have been made redundant **as a result of** the bankruptcy of their company’

The examples from the OpenSoNaR corpus illustrate the formal character of these adpositions in Modern Dutch. Like English, Dutch also has univerbation with the Proto-Germanic adposition *bi* ‘by’ + N: *bezijden* ‘beside(s)’, *bezuiden* ‘(to the) south of’, *beoosten* ‘(to the) east of’ etc. They were formed as adverbs, but could (and still can) also be used as prepositions.

We can conclude that univerbation is a recurring feature in the development of complex adpositions and needs to be taken into account. Why the three languages discussed here differ with respect to their tendency towards univerbation is an interesting question. While it seems to have happened quite frequently in German, it is rare, and restricted to earlier stages of the languages in Dutch and English. We can assume that the reluctance is related to the generally higher degree of analyticity of the present stage of the two languages.

Another, perhaps related, interesting question concerns the fact that the final stage of Allerton's grammaticalization path, the reduction of the expression to what was originally the noun, is relatively frequent in German but very rarely taken in English. Here, the analyticity of English should not be an obstacle (it might even be argued that it should favor such a reduction). It is possible that the availability of a productive derivation of adpositions from present participles stands in the way of this step – that the reduction of, for example, *with regard* to *regard* is preempted by the availability of the form *regarding*, etc. This does not, however, explain why this kind of reduction is rare in Dutch, too. Above, we already discussed the example of *richting*, which is arguably the most prominent of the few cases of adpositions resulting from this reduction process.

Taking into account everything discussed here, we can assume the following, more fine-grained model of the development of canonical complex adpositions from free combinations, and the subsequent development into simple adpositions, see Figure 2. As should have become clear from our discussion, none of the subsequent changes are predictable or even obligatory. Which, if any, of the paths in our model are preferred in general, or in specific languages, is not currently known.

5 Conclusion

5.1 Lexicalization or grammaticalization

So far, we have remained largely agnostic about the issue whether the diachronic development of complex adpositions should be considered a process of grammaticalization or rather one of lexicalization. In this concluding section, we would like to take up this issue and discuss some alternative views on the phenomenon.

Without going into much detail, grammaticalization is usually seen as a

... the change whereby in certain linguistic contexts speakers use parts of a construction with a grammatical function. Over time the resulting grammatical item may become more grammatical by acquiring more grammatical functions and expanding its host-classes.

(Brinton and Traugott 2005: 99)

whereas lexicalization is defined as

... the change whereby in certain linguistic contexts speakers use a syntactic construction or word formation as a new contentful form with formal and semantic properties that are not completely derivable or predictable from the constituents of the construction or the word formation pattern. Over time there may be further loss of internal constituency and the item may become more lexical.

(Brinton and Traugott 2005: 96)

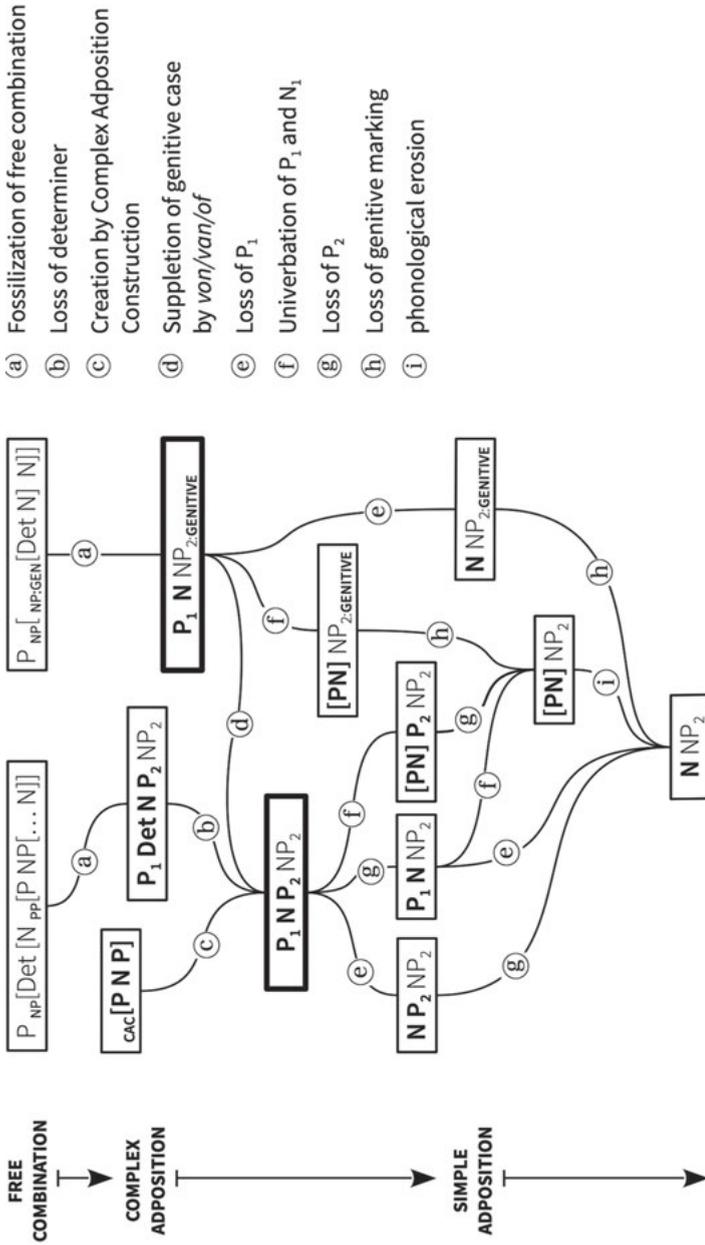


Figure 2: From free combination to complex preposition and on to simple preposition.

Grammaticalization and lexicalization are constituted by diachronic processes which have a lot in common; but the received view is that they differ from each other with respect to some crucial aspects, which are summarized in Table 14.

Table 14: Similarities between lexicalization and grammaticalization (Brinton and Traugott 2005: 110).

	LEXICALIZATION	GRAMMATICALIZATION
Bonding	+	+
Coalescence	+	+
Idiomatization	+	+
Unidirectionality	+	+
Gradualness	+	+
Functional shift	–	+
Decategorialization	–	+
Type frequency/ productivity	–	+
Token frequency	–	+
Typological generality	–	+

With respect to complex adpositions, the question has often been posed whether the development of these elements constitutes a case of lexicalization or grammaticalization. Very generally and intuitively, the discussion may be roughly summarized as follows. The addition of a new expression to the lexicon of a language is usually considered the result of lexicalization. As the development of complex adpositions crucially involves the creation of a completely new and idiosyncratic linguistic string which has not been there before, there are reasons to call their emergence lexicalization. On the other hand, the newly emerged expression behaves similarly to primary or simple adpositions (see Section 3). On these grounds, the emergence of complex adpositions should be considered a case of grammaticalization, as the outcome of the process is a grammatical item, a gram.

Some linguists see complex adpositions as lexical units and attribute their development to the diachronic process of lexicalization. Huddleston (1988) and Huddleston and Pullum (2002) explicitly exclude PNP sequences, which we have described as complex adpositions above, from the grammatical class of adpositions/prepositions, and call them “idiomatic expressions” with varying degrees of fossilization, cf. e.g.:

[m]ost lexical items are words or lexemes but they can also be larger: these are idioms, and with idioms there may be conflict instead of the usual congruence between what counts as a unit from a lexical point of view and what counts as a unit from a grammatical point of

view. [...] Such a mismatch is found in the above Prep ([Determiner]) N Prep sequences. In *for the sake of the premier*, for example, *for the sake of* belongs together lexically, but grammatically the immediate constituents are not *for the sake of + the premier* but *for + the sake of the premier*. (Huddleston 1988: 126–7)

It should be said at this point, however, that the majority of linguists agree on the idea that the development of complex adpositions cannot be fully accounted for without a reference to grammaticalization. Somewhat intuitively, this idea is already contained in the term “complex adposition”, as adpositions are traditionally conceived of as a grammatical class. That is, when a combination of words, as e.g. *for the sake of*, develops towards a more or less fixed and idiomatic linguistic expression with a clearly grammatical function usually dedicated to (simple, primary) adpositions, it lends itself to the interpretation in terms of grammaticalization. At the same time, the process of fixation, idiomaticization etc., that is, the process of formation of a new linguistic expression in a language, has been usually considered lexicalization, understood as an addition of a new expression to the inventory of a language.

In the literature, one finds accounts in which both lexicalization and grammaticalization are taken to be at work – simultaneously – during the development of complex adpositions, cf. e.g.:

In the evolution of this preposition, lexicalization and grammaticalization are intertwined: the construction *du côté de* becomes a new lexical unit, but during this very evolution, the noun *côté* is decategorized, which constitutes a grammaticalization process. Thus, the formation of the complex preposition *du côté de* and even the transformation from *côté* to a preposition is a lexicalization process, but the change of category implied by this evolution can be considered as a case of grammaticalization.

(Fagard and De Mulder 2007: 27, our translation)⁸

Others, however, argue for a clear distinction between these two processes, considering them two different stages in the development of complex adpositions. Lehmann (1998, 2002), for example, reserves the term lexicalization for the development of fixed and idiomatic constructions, and proposes to use the term

8 “Dans l’évolution de cette préposition, lexicalisation et grammaticalisation sont imbriquées l’une dans l’autre: la construction *du côté de* devient une nouvelle unité lexicale, mais lors de cette évolution même, le nom *côté* est décatégorisé, ce qui constitue plutôt un processus de grammaticalisation. [...] Ainsi, la formation de la préposition complexe *du côté de* et même la transformation de *côté* en préposition est une lexicalisation, mais le changement de catégorie impliquée par cette évolution peut être considéré comme un phénomène de grammaticalisation.”

grammaticalization only for the further stages in the development of complex adpositions, which may (but not need to) follow, cf. e.g.:

We have seen that prepositions [...] come about not by grammaticalization, but by lexicalization. Once they have come into existence, they may then be grammaticalized. [...] From among all the new prepositions [...], only a fraction is grammaticalized. All the others are abandoned and replaced by other neologisms. (Lehmann 2002: 18)

The genesis of members of minor word classes, in particular adpositions and conjunctions, has often been treated as an instance of grammaticalization. However, minor word classes are not necessarily classes of grammatical formatives. In particular, there are more lexical and more grammatical adpositions. For instance, before *auf Grund (von)* ‘on the basis of’ can ever get grammaticalized to a grammatical preposition, it must first be lexicalized to the lexical preposition *aufgrund (von)*. In this sense, grammaticalization presupposes lexicalization. (Lehmann 2002: 1)

In this scenario, the formation of a complex adposition, as it has been described in Section 4.2 above, is considered a process of lexicalization, whereas further possible changes, focused on in Section 4.3 above, are considered grammaticalization. Lexicalization is thus essentially constituted by the reanalysis of the internal syntactic structure, that is, a compositional syntactic string formed according to the regular rules of syntax is reanalyzed as a unit. Grammaticalization affects then this new item or construction newly created by reanalysis at the stage of lexicalization.

This neat distinction between two different diachronic processes is very appealing, at least from a theoretical point of view. However, one should bear in mind that it is an analytical distinction, which may be very useful on the conceptual level, but is, however, very hard or nearly impossible to maintain in practice. That is, if we want to observe and trace back a development of a complex adposition using diachronic corpus data, it will be hardly possible to determine where the lexicalization process ends and the grammaticalization process starts. Trivially, change is always gradual and involves a lot of variation in a particular synchronic segment. That is, the boundary between lexicalization and grammaticalization may in fact be less clear and neat than suggested by Lehmann (1998, 2002).

5.2 An alternative view: Constructionalization

Both lexicalization and grammaticalization are gradual processes (or, more specifically, take place within gradual processes of linguistic change). However, as hinted at in Section 4.2 above, at least with respect to English, evidence for such gradual processes is often lacking, with complex adpositions appearing instantaneously. Hoffmann suggests, plausibly, that such cases may be formed by what he

calls ‘grammaticalization by analogy’ (Hoffmann 2005: 86). He does not expand on this possibility extensively, but a close reading of his discussion suggests that he has in mind a process that actually goes beyond exemplar-based analogical formation:

... by the beginning of the seventeenth century, *the sequence ‘preposition + noun + preposition’ was already well established as a potential unit-like structure* via a number of combinations. [...] [T]he structure exists in an abstract, grammaticalized state but may – when necessary – be filled with new lexical items. (Hoffmann 2005: 86, our emphasis)

And:

... the sequence ‘preposition–noun–preposition’ appears to be available as a *grammaticalized yet abstract construct* which under certain circumstances can be filled by new lexical entities to form a new complex preposition. (Hoffmann 2005: 171, our emphasis)

In other words, what Hoffmann is suggesting is what we could call, in terms of construction morphology (cf. Booij 2010; 2016), a constructional schema. To introduce the notion, consider (88a), which is Booij’s analysis for Dutch compounds (Booij 2016:431, which is also accurate for right-headed English compounds). In Booij’s notation, the entire constructional schema is enclosed in angled brackets, the part on the left side of the double arrow specifies the form, that on the right the meaning, with the double arrow indicating the general relation between form and meaning and the indices showing more specific correspondences. The schema thus specifies a construction consisting of two nouns, where the entire construction has the general meaning of the right noun, modified by the meaning of the left noun. Our rough suggestion for a (canonical) complex adposition schema is shown in (88b):

- (88) a. $\langle [N_i N_j]_{N_k} \longleftrightarrow [SEM_j \text{ with some relation to } SEM_i]_k \rangle$
 b. $\langle [P_i N_j P_k]_{P_m} \longleftrightarrow [RELATION \text{ characterized by } SEM_{ij}]_m \rangle$

The formal side specifies that a canonical complex adposition consists of a sequence involving a preposition (P_i), a noun (N_j), and another preposition (P_k), that is different from the first, and that these three constituents together form a new preposition (P_m). The semantic side specifies that the meaning of this new preposition is a relation that is somehow characterized by the first preposition and the noun (the second preposition is structurally governed by the noun and does not contribute any specific semantics). Of course, once a pattern is established on the basis of this schema, it may undergo semantic change (such as the semantic bleaching of the noun mentioned in Section 4.2).

This characterization of the complex adposition schema can no doubt be improved on. Our point here is that the existence of such a schema means that speakers can create new complex prepositions spontaneously by filling the variable slots. Although we cannot argue conclusively for the existence of such a schema, it is certainly plausible from a cognitive-functional perspective. Once a sufficient number of adpositions has emerged via regular processes of grammaticalization and/or lexicalization, their formal and functional similarities will be abstracted in the minds of the speakers, leading to a schematic representation. Hoffmann (2005: 86) suggests that a number of complex adpositions had emerged at the beginning of the 17th century that could have formed the basis of just such a schematization process. In addition, as mentioned in Section 4.2 above, a number of French loan calques of the relevant form had entered the English language. Also, the Old and Middle English complex adpositions discussed in Section 4.3, although fully lexicalized by the 17th Century, might still have been transparent enough to contribute to such a schema.

For Dutch, Van der Wouden (2006: 201) suggests abstract schemata for certain subtypes of complex prepositions. A (sub)schema like <ter [V-ing]_N van> seems to account for a set of complex prepositions:

- (89) a. **ter behandeling van de nasleep van 't letsel**
lit. 'for treatment of the aftermath of the injury'
- b. **ter beoordeling van de moties**
lit. 'for consideration of the motions'
- c. **ter bepaling van de gedachte**
lit. 'for determining of the idea'
- d. **ter bereiking van win-win-situaties**
lit. 'for reaching of win-win-situations'

The text corpora of Dutch contain dozens of prepositional constructions corresponding to this pattern. The schema has, apparently, been used productively for the formation of new complex prepositions.

In the context of constructionalization, it is striking that not all simple adpositions and their logically possible combinations are represented equally among the modern complex adpositions – just five combinations of P1 and P2 account for more than two thirds of the 81 commonly recognized complex adpositions in English: [*in N of*] (26 types), [*on N of*] (11 types), [*in N with*] (10 types), [*in N to*] (6 types), and [*by N of*] (6 types). German and Dutch are very similar in this respect.

The dominance of *of*, *with* and *to* in the P2 position can be explained in terms of the fact that these are the adpositions most commonly occurring in NP-internal

PPs in general. With respect to P1, however, general frequency cannot account completely for the dominance of certain adpositions. Here are the type frequencies for P1 among the 81 complex adpositions: *in* (42), *on* (11), *at* (8), *by* (7), *with* (6), *for* (4), *under* (2), *from* (1). Although *in* is the most frequent adposition in English after *of*, it is only about twice as frequent as *for*, which almost never occurs as P1, and *on*, which does, but much less frequently than *in*. Other frequent English adpositions, such as *about*, *over* or *through*, never occur in the P1 slot of complex adpositions.

This may suggest the existence of more specific subschemas of a general Complex Adposition schema, where the slots of P₁ and P₂ are fixed – Quirk and Mulholland in effect suggest as much when they observe that ‘since P1 – P2 are expounded more often by *in* – *of* than by any other pair of prepositions, this feature may well have analogical power’ (Quirk and Mulholland 1964: 71, cf. also Klégr 2002: 28).

This possibility is shown for the combinations *in* – *of* and *in* – *to* in (90a, b):

- (90) a. < [*in* N_i *of*]_{P1} ↔ [RELATION of SEM_i]_j >
 b. < [*in* N_i *to*]_{P1} ↔ [RELATION of SEM_i]_j >

Here, we assume that the prepositions, being a fixed part of the schema, no longer contribute their usual, independent meaning. Tentative evidence for the existence of such subschemas is that all English complex adpositions with variation at the P1/P2 slots have one variant corresponding to *in N of* or *in N to* in addition to a more established variant with other adpositions, as shown in Table 15.

Table 15: Evidence for the existence of a construction [*in N of/to*]_P in English.

Major variant	Minor variant with [<i>in N of/to</i>] _P
<i>on behalf of</i> (27.1 p. Mill.)	<i>in behalf of</i> (0.003 p. Mill.)
<i>by reference to</i> (6.6 p. Mill.)	<i>in reference to</i> (0.7 p. Mill.)
<i>with reference to</i> (3.9 p. Mill.)	
<i>in respect of</i> (29.3 p. Mill.)	<i>in respect to</i> (0.58 p. Mill.)
<i>with respect to</i> (13.3 p. Mill.)	
<i>with regard to</i> (16.5 p. Mill.)	<i>in regard to</i> (2.8 p. Mill.)

Of course, a more detailed investigation is needed, also with respect to whether such evidence also exists for Dutch and German.

5.3 Emergence of complex adpositions in European Languages: A hypothesis

A more general question with respect to the emergence of complex adpositions not just in the West Germanic languages, but in all European languages, is where and why they became popular (cf. Van der Horst 2013). With respect to the where, a long-standing observation concerning their distribution across text types is intriguing: complex adpositions typically occur in formal registers. This was first pointed out (for German) by Beneš (1974: 45), who specifically mentions technical, non-fiction and journalistic writing. Similar observations (also pertaining to German) are found in Lehmann (1991: 501), who sees complex adpositions as a feature of ‘bureaucratic and journalistic jargon’ as opposed to ‘colloquial’ usage, and Di Meola (2000: 183) who mentions academic and newspaper language. For Dutch, Van der Wouden (2006: 201) claims, likewise, that such prepositions are found especially in formal language, and Hoffmann (2005, Ch. 6) reports that they are rare in imaginative prose and largely absent from conversational registers.

This distribution across genres may simply be a consequence of the higher need for explicitness in the registers in question (cf. Beneš 1974: 45, Stefanowitsch, forthcoming), i.e., complex adpositions may have emerged in everyday language to be taken up especially in such formal contexts. An alternative hypothesis that suggests itself, however, is that many of these complex adpositions were actually created in these formal registers and that their growing productivity and popularity in the language in general is due to the fact that these registers and the corresponding text types became increasingly important in the course of the 19th and 20th centuries (Hüning 2014). This hypothesis would also account for the strikingly parallel development in all European languages as well as for the specific timing of the emergence of the current set of complex prepositions in all these languages.

Abbreviations

ACC	accusative	NOM	nominative
ADJ	adjective	NP	noun phrase
ADV	adverb	P	preposition
DAT	dative	PP	prepositional phrase
DET	determiner	SEM	semantics
GEN	genitive	V	verb
N	noun		

References

- Allerton, David J. 2008. “Over the hills and far away” or “far away over the hills”: English place adverb phrases and place prepositional phrases in tandem? In Dennis Kurzon & Silvia Adler (eds.), *Adpositions: pragmatic, semantic and syntactic perspectives*. (Typological Studies in Language 74), 37–65. Amsterdam: John Benjamins.
- Augustin, Hagen. 2018. *Verschmelzung von Präposition und Artikel: eine kontrastive Analyse zum Deutschen und Italienischen* (Konvergenz Und Divergenz 6). Berlin: De Gruyter Mouton.
- Baker, Mark C. 2003. *Lexical categories: verbs, nouns, and adjectives*. (Cambridge Studies in Linguistics 102). Cambridge: Cambridge University Press.
- Baayen, R. Harald, Richard Piepenbrock & Leon Gulikers. 1995. *The CELEX Lexical Database* (CD-ROM). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.
- Beliën, Maaïke Laura. 2008. *Constructions, constraints, and construal: Adpositions in Dutch*. Diss. VU Amsterdam. (LOT Dissertation Series 194). Utrecht: LOT.
- Behaghel, Otto. 1924. *Deutsche Syntax. Eine geschichtliche Darstellung*. Bd. 2. Heidelberg: Winter.
- Beneš, Eduard. 1974. Präpositionswertige Präpositionalfügungen. In Ulrich Engel & Paul Grebe (eds.), *Sprachsystem und Sprachgebrauch: Festschrift für Hugo Moser zum 65. Geburtstag* (Sprache der Gegenwart 33), vol. 1, 33–52. Düsseldorf: Schwann.
- Bernstein, Theodore M. 1965. *The careful writer: A modern guide to English usage*. New York: Atheneum.
- Booij, Geert. 2010. *Construction morphology*. Oxford: Oxford University Press.
- Booij, Geert. 2016. Construction morphology. In Andrew Hippisley & Gregory Stump (eds.), *The Cambridge handbook of morphology*, 424–448. Cambridge: Cambridge University Press.
- Borillo, Andrée. 1997. Aide à l'identification des prépositions complexes de temps et de lieu. *Faits de langue* 9, 173–184.
- Braunmüller, Kurt. 1982. *Syntaxtypologische Studien zum Germanischen*. (Tübinger Beiträge zur Linguistik; 197) Tübingen: Narr.
- Brinton, Laurel J. & Elizabeth Closs Traugott. 2005. *Lexicalization and Language Change*. Research Surveys in Linguistics. Cambridge University Press.
- Brugman, Claudia. 1988. *The story of 'over': Polysemy, semantics and the structure of the lexicon*. New York: Garland.
- Broekhuis, Hans. 2013. *Syntax of Dutch: adpositions and adpositional phrases*. (Comprehensive grammar resources). Amsterdam: Amsterdam University Press.
- Crössmann, Helga. 1973. Präposition oder Konjunktion. In Ulrich Engel & Paul Grebe (eds.), *Sprachsystem und Sprachgebrauch: Festschrift für Hugo Moser zum 65. Geburtstag* (Sprache der Gegenwart 33), vol. 2, 16–38. Düsseldorf: Schwann.
- Cruse, Alan. 2011. *Meaning in language: An introduction to semantics and pragmatics* (Oxford Textbooks in Linguistics). Third edition. Oxford: Oxford University Press.
- Culicover, Peter W. 1999. *Syntactic nuts: hard cases, syntactic theory, and language acquisition* (Foundations of Syntax 1). Oxford: Oxford University Press.
- Cuyckens, Hubert. 1993. The Dutch spatial preposition ‘in’: A cognitive-semantic analysis. In Cornelia Zelinsky-Wibbelt (ed.), *The semantics of prepositions in natural language processing*, 27–71. Berlin: Mouton de Gruyter.
- Cuyckens, Hubert. 1995. Family resemblance in the Dutch spatial prepositions *door* and *langs*. *Cognitive Linguistics* 6. 183–208.

- Di Meola, Claudio. 2000. *Die Grammatikalisierung deutscher Präpositionen* (Studien Zur Deutschen Grammatik 62). Tübingen: Stauffenburg.
- Di Meola, Claudio. 2003. Grammaticalization of postpositions in German. In Hubert Cuyckens, Thomas Berg, René Dirven & Klaus-Uwe Panther (eds.), *Motivation in language: studies in honor of Günter Radden* (Current Issues in Linguistic Theory 243), 203–222. Amsterdam & Philadelphia: John Benjamins.
- Diewald, Gabriele. 2002. A model for relevant types of contexts in grammaticalization. In Ilse Wischer & Gabriele Diewald (eds.), *New Reflections on Grammaticalization*, 103–120. Amsterdam & Philadelphia: Benjamins (TSL 49).
- Draye, Luc. 1992. Zum Trajektiv. Ein Kapitel aus einer kognitiv orientierten niederländisch-deutschen Kontrastivgrammatik. *Leuvense Bijdragen* 81. 163–203.
- Dudenredaktion. 2009. *Duden. Die Grammatik (Duden 4)*. 8th edition. Mannheim, Leipzig, Wien, Zürich: Dudenverlag.
- Eichinger, Ludwig M. 1995. Von der Valenz des Verbs und den Abhängigkeiten in der Nominalgruppe. In Eroms, Hans-Werner, Ludwig M. Eichinger (eds.), *Dependenz und Valenz*, 37–52. Hamburg: Buske.
- Eisenberg, Peter. 1979. Syntax und Semantik denominaler Präpositionen des Deutschen. In Harald Weydt (ed.), *Die Partikeln der deutschen Sprache*, 518–527. Berlin: de Gruyter.
- Erben, Johannes. 2000. *Einführung in die deutsche Wortbildungslehre*. 4th edition. (Grundlagen der Germanistik 17). Berlin: Erich Schmidt.
- Fagard, Benjamin & Walter De Mulder. 2007. La formation des prépositions complexes : grammaticalisation ou lexicalisation ? *Langue française* 156, 9–29.
- Fagard, Benjamin (to appear). Prépositions et locutions prépositives. In Marchello-Nizia, C. et al. (eds), *Grande grammaire historique du français*. Berlin: De Gruyter.
- Heine, Bernd. 2002. On the role of context in grammaticalization. In Wischer, Ilse & Gabriele Diewald (eds.) 2002. *New reflections on grammaticalization*. (Typological Studies in Language, 49), 83–101. Amsterdam, Philadelphia: John Benjamins.
- Hoffmann, Sebastian. 2005. *Grammaticalization and English complex prepositions: a corpus-based study* (Routledge Advances in Corpus Linguistics 7). New York: Routledge.
- Huddleston, Rodney D. 1984. *Introduction to the grammar of English*. Cambridge: Cambridge University Press,
- Huddleston, Rodney D. 1988. *English grammar: An outline*. Cambridge: Cambridge University Press.
- Huddleston, Rodney D. & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Hüning, Matthias. 2014. Over complexe preposities en convergentie [About complex prepositions and convergence]. In Freek Van de Velde, Hans Smessaert, Frank Van Eynde & Sara Verbrugge (eds.), *Patroon en argument. Een dubbelfestbundel bij het emeritaat van William Van Belle en Joop van der Horst* [Pattern and argument. A double festschrift on the occasion of the retirements of William Van Belle and Joop van der Horst], 433–445. Leuven: Universitaire Pers Leuven.
- Klégr, Aleš. 2002. *English complex prepositions of the type in spite of and analogous sequences: a study & dictionary*. Praha: Karolinum.
- König, Ekkehard & Bernd Kortmann. 1991. On the reanalysis of verbs as prepositions. In Gisa Rauh (ed.), *Approaches to prepositions* (Tübinger Beiträge Zur Linguistik 358), 109–125. Tübingen: Gunter Narr.

- Kurzon, Dennis & Silvia Adler (eds.). 2008. *Adpositions: pragmatic, semantic and syntactic perspectives* (Typological Studies in Language 74). Amsterdam, Philadelphia: John Benjamins.
- Lakoff, George. 1987. *Women, fire, and dangerous things: what categories reveal about the mind*. Chicago: The University of Chicago Press.
- Lehmann, Christian. 1991. Grammaticalization and related changes in contemporary German. In Elizabeth Closs Traugott & Bernd Heine (eds.), *Approaches to grammaticalization*, vol. 2, 493–535. (Typological Studies in Language 19). Amsterdam & Philadelphia: John Benjamins.
- Lehmann, Christian. 1998. *German abstract prepositional phrases*. In Iørn Korzen & Michael Herslund (eds.), *Clause combining and text structure*, 87–106. (Copenhagen Studies in Language 22). Frederiksberg: Samfundslitteratur.
- Lehmann, Christian & Christel Stolz. 1992. *Bildung von Adpositionen im Deutschen* (Arbeitspapiere des Seminars für Sprachwissenschaft der Universität Erfurt 6). Erfurt: Seminar für Sprachwissenschaft der Universität.
- Lindqvist, Christer. 1994. *Zur Entstehung von Präpositionen im Deutschen und Schwedischen* (Linguistische Arbeiten 311). Tübingen: Niemeyer.
- Lindstromberg, Seth. 2010. *English prepositions explained*. 2nd edition. Amsterdam & Philadelphia: John Benjamins.
- Meex, Birgitta. 2002. Die Wegpräposition über. In Hubert Cuyckens & Günter Radden (eds.), *Perspectives on prepositions*, 157–176. Berlin: De Gruyter.
- Meibauer, Jörg. 1995. Komplexe Präpositionen – Grammatikalisierung, Metapher, Implikatur und *division of pragmatic labour*. In Frank Liedtke (ed.), *Implikaturen* (Linguistische Arbeiten 343), 47–74. Tübingen: Niemeyer.
- Nübling, Damaris. 1998. Wann werden die deutschen Präpositionen flektieren? Grammatisierungswege zur Flexion. In Ray Fabri, Albert Ortman & Teresa Parodi (eds.), *Models of inflection*, 266–292. Berlin: De Gruyter.
- Paul, Hermann. 1920. *Prinzipien der Sprachgeschichte*. Studienausgabe. 9., unveränderte Auflage, 1975 [Nachdruck der 5. Auflage, 1920]. (Konzepte der Sprach- und Literaturwissenschaft 6). Tübingen: Max Niemeyer Verlag.
- Quirk, Randolph & Joan Mulholland. 1964. Complex prepositions and related sequences. *English Studies* 45(1–6). 64–73.
- Quirk, Randolph, Jan Svartvik & Geoffrey N. Leech. 1985. *A comprehensive grammar of the English language*. London; New York: Longman.
- Rauh, Gisa. 1991. Prepositional forms in the lexicon: Problems and suggestions. In Gisa Rauh (ed.), *Approaches to prepositions*, 169–223. (Tübinger Beiträge zur Linguistik 358). Tübingen: Gunter Narr.
- Rissanen, Matti. 2004. Grammaticalisation from side to side: On the development of *beside(s)*. In Hans Lindqvist & Christian Mair (eds.), *Corpus approaches to grammaticalization in English* (Studies in Corpus Linguistics 13), 151–170. Amsterdam: John Benjamins.
- Santorini, Beatrice. 1990. *Part-of-speech tagging guidelines for the Penn Treebank project*. Technical Report. Philadelphia: University of Pennsylvania, Dept. of Computer and Information Science.
- Schiering, René. 2005. Flektierte Präpositionen im Deutschen? Neue Evidenz aus dem Ruhrgebiet. *Zeitschrift für Dialektologie und Linguistik* 72(1). 52–79.
- Schwenter, Scott A. & Elizabeth Closs Traugott. 1995. The semantic and pragmatic development of substitutive complex prepositions in English. In Andreas H. Jucker (ed.), *Historical*

- pragmatics: Pragmatic developments in the history of English*, 243–273. (Pragmatics & Beyond New Series 35). Amsterdam & Philadelphia: John Benjamins.
- Scott, Alan. 2014. The survival and use of case morphology in Modern Dutch. In Ronny Boogaart, Timothy Coleman & Gijsbert Rutten (eds.), *Extending the scope of construction grammar*, 107–137. (Cognitive Linguistics Research 54). Berlin: De Gruyter Mouton.
- Seppänen, Aimo, Rhonwen Bowen & Joe Trotta. 1994. On the so-called complex prepositions. *Studia Anglica Posnaniensia* 24. 3–29.
- Stathi, Katerina. forthc. Die Entstehung der komplexen Präposition in Richtung im Deutschen: Zur Emergenz und Verfestigung von Konstruktionen. In Alexander Lasch & Alexander Ziem (eds.), *Konstruktionsgrammatik VI. Historische Konstruktionsgrammatik: Konvergenzen und Divergenzen im Sprach- und Konstruktionswandel*. Tübingen: Stauffenburg.
- Stefanowitsch, Anatol. forthc. Complex prepositions of contrast and opposition in English. In Marcin Grygiel and Barbara Lewandowska-Tomaszczyk (eds.), *Contrast and analogy in language: Perspectives from cognitive linguistics*. Amsterdam & Philadelphia: John Benjamins.
- Stoett, Frederik A. 1923. *Middelnederlandsche Spraakkunst: Syntaxis* [Middle Dutch Grammar: Syntax]. 3rd edition. Den Haag: Martinus Nijhoff.
- Szczepaniak, Renata. 2009. *Grammatikalisierung im Deutschen. Eine Einführung* (Narr Studienbücher). Tübingen: Gunter Narr.
- van Bree, Cor. 2014. Adpositionele configuraties in het Nederlands [Adpositional configurations in Dutch]. *Acta Universitatis Wratislaviensis. Neerlandica Wratislaviensia* 24. 11–25.
- van der Horst, Joop. 2008. *Geschiedenis van de Nederlandse syntaxis* [Dutch historical syntax]. Leuven: Universitaire Pers Leuven.
- van der Horst, Joop. 2013. *Taal op drift: Lange-termijnontwikkelingen in taal en samenleving* [Language adrift: long-term evolutions in language and society]. Amsterdam: Meulenhoff.
- van der Horst, Joop & Kees van der Horst. 1999. *Geschiedenis van het Nederlands in de twintigste eeuw* [History of Dutch in the twentieth century]. Den Haag & Antwerpen: Sdu Uitgevers.
- van der Horst, Joop & Freek Van de Velde. 2008. Het voorzetsel diachronisch. In Irena Barbara Kalla & Bożena Czarnecka (eds.), *Neerlandistische ontmoetingen: Trefpunt Wrocław* [Dutch encounters: Meeting point Wrocław], 54–62. Wrocław: Oficyna Wydawnicza Atut.
- van der Wouden, Ton. 2006. Nederlandse voorzetsels tussen Duitse en Engelse [Dutch prepositions between German and English.]. In Matthias Hüning, Ulrike Vogl, Ton van der Wouden & Arie Verhagen (eds.), *Nederlands tussen Duits en Engels* [Dutch between German and English] (SNL-Reeks 15), 183–206. Leiden: SNL.
- Villada Moirón, María Begoña. 2005. *Data-driven identification of fixed expressions and their modifiability* (Groningen Dissertations in Linguistics 52). Groningen: Rijksuniversiteit Groningen.
- Vranjes, Jelena. 2012. *Voorzetseluitdrukkingen in het Nederlands sinds de 16de eeuw: een diachroon corpusonderzoek*. [Prepositional expressions in Dutch since the 16th century: a diachronic corpus study] Unpublished M.A. thesis. Leuven: Katholieke Universiteit Leuven.
- Zifonun, Gisela, Ludger Hoffmann, Bruno Strecker & Joachim Ballweg. 1997. *Grammatik der deutschen Sprache* (Schriften des Instituts für Deutsche Sprache 7). Berlin: De Gruyter.

Dejan Stosic

3 Complex adpositions in Slavic languages: The case of Serbian

Abstract: The chapter offers an illustration of the adpositional systems of Slavic languages through both an overview of simple adpositions and an in-depth study of complex adpositions in Serbian, a South-Slavic language. Using a broad definition of complex adpositions, which can be considered as multi-word functional equivalents of simple adpositions, I first provide an inventory of four hundred expressions which, while displaying variable degrees of fossilization, constitute the empirical basis for this work. This then allows me to point out that their formation follows up to twenty different patterns, and that the vast majority of them involve four construction schemas, with one highly productive pattern forming about 70% of complex adpositions in Serbian, namely [P₁ N] (e.g. *u okviru* ‘in the framework/scope of’). Finally, a quantitative, corpus-based analysis confirms that, as in other European languages, the class of complex adpositions in Serbian is to be thought of as a continuum between sequences that are formally coalesced and semantically rather opaque, and those that exhibit a certain degree of transparency.

Keywords: Slavic languages, Serbian, Prepositions, Adpositions, Complex adpositions

1 Introduction

Even though most Slavic languages retained the nominal declension with rich, six- or seven-case systems, during the three last centuries, they also developed a wide range of complex adpositions (CAs), that are generally used as prepositions, and rarely as postpositions, and that complement the existing paradigms of simple adpositions (SAs). The aim of this chapter is to provide an illustration of the adpositional systems of Slavic languages through an in-depth study of complex adpositions in one of them, namely Serbian, a South-Slavic language. The chapter consists of four sections. After this brief introduction, the second section gives an overview of the main sociolinguistic and typological

Dejan Stosic, University of Toulouse Jean Jaurès & Laboratory CLLE (CNRS & UT2J), Toulouse, France. dejan.stosic@univ-tlse2.fr

<https://doi.org/10.1515/9783110686647-004>

characteristics of Slavic languages. The third section opens with a general description of the adpositional system of Serbian, providing a survey of its role in marking relations between syntactic constituents of various types. Further, working definitions of both simple and complex adpositions are established, and a set of possible criteria for identifying the latter is proposed. The fourth section deals specifically with complex adpositions in Serbian. It first tackles the methodological issue of how to index all the sequences that meet the definitional requirements of complex adposition (§ 4.1), and then describes their construction patterns and the way in which complex adpositions interact with case marking (§ 4.2). The last part of this section (§ 4.3) reports the preliminary results of a corpus-based study of the frequency of occurrence of complex adpositions in a 477-million-word Serbian corpus, comparing it to the frequency of use of simple adpositions in the same corpus.

2 Slavic languages

Slavic languages descend directly from the Proto-Slavic language and are very close to the Baltic branch of Indo-European languages. They are spoken by some 390 million people in central and eastern Europe, and in Russia.

2.1 Sociolinguistic and geographical features

There are more than twenty Slavic languages split into three sub-groups: West Slavic (e.g. Polish, Czech), South Slavic (e.g. Serbian, Croatian, Bulgarian), and East Slavic (e.g. Russian, Ukrainian). Table 1 gives the number of speakers for nineteen Slavic languages, as well as their affiliation to one of the three sub-groups:

Table 1: Number of speakers for the main Slavic languages (source: *Ethnologue*).

Sub-group	Variety	Number of speakers in the country of origin:	Total number of speakers: all users (L1+L2)
East (4)	Russian	138 000 000	258 227 760
	Ukrainian	30 800 000	33 082 790
	Belarusian	3 510 000	3 843 110
	Rusyn	6 730	82 960

Table 1 (continued)

Sub-group	Variety	Number of speakers in the country of origin:	Total number of speakers: all users (L1+L2)
West (7)	Polish	36 711 000	40 378 030
	Czech	10 532 000	13 386 850
	Slovak	5 115 000	7 227 240
	Kashubian	107 000	117 000
	Silesian	522 000	522 000
	Lower Sorbian	6 670	6 670
	Upper Sorbian	13 300	13 300
South (8)	Serbian	6 340 000	8 779 176
	Croatian	4 100 000	6 741 420
	Slovenian	1 995 000	2 226 650
	Bosnian	2 010 000	2 623 190
	Montenegrin	230 000	266 130
	Slavomolisano	1 000	1 000
	Bulgarian	6 840 000	8 144 880
	Macedonian	1 380 000	1 620 290
TOTAL		248 219 700	387 290 446

2.2 Major typological characteristics

Slavic languages display a quite complex morphology, conserving an array of Indo-European morphological features. For instance, they generally retained the three-way gender opposition (m, f, n), a nominal declension with six to seven cases applying to nouns as well as to pronouns and adjectives, and a very rich verbal inflexion, sensitive to the perfective vs imperfective aspectual opposition. The development of the latter led to the loss of the imperfect and aorist forms in most Slavic languages, their values being recovered by an original analytical perfect. According to Comrie and Grimes (2013), “the basic vocabulary shared by all Slavic languages reflects both IE origins and Common Slavic innovations”. However, one can find significant differences in the lexicons of Slavic languages, mainly due to different cultural traditions and to the influence of different foreign languages, which “have led to surprising dissimilarity in more abstract and technical vocabulary” (ibidem). As for syntax, modern Slavic languages mainly have an SVO structure, but the word order is very flexible. Two main devices are generally used in expressing syntactic relations, namely the case system and prepositions (rather than postpositions) (see Comrie & Corbett 1993).

3 General description of the Serbian adpositional system

The following sections will be devoted to the Serbian adpositional system, as illustrative of other Slavic languages. For more extensive studies on adpositions in some other Slavic languages, the reader is referred to McGranahan (1971), Šiganova (2001), Hill (2012) for Russian, Hittery (2000) for Slovenian, Jelínek (1964), Dušková (1994), Čermák (1996, 2007), Melíšková (2008) for Czech, and to Stosic (forth. b) for a general overview.

As in other European languages, Serbian has a paradigm of adpositions, which are mainly prepositions. They can be morphologically simple (1–2) or complex (3–4):

- (1) *devojka iz grada*
girl.NOM from city.GEN
'a city girl'

- (2) *kuća kod crkve*
house.NOM at church.GEN
'a house near the church'

- (3) *tišina za vreme koncerta*
silence.NOM during concert.GEN
'a silence during the concert'

- (4) *percepcija nasilja od strane dece*
perception.NOM violence.GEN on behalf of children.GEN
'a perception of violence by children'

Some adpositions can also function as postpositions (see Blagus Bartolec and Matas Ivanković 2013), but these cases are very limited in number and should rather be considered as inversions of the canonical construction:

- (5) *Učiću je da se ne laže da se
žrtvuje ljubavi radi.* (<http://mahlat.rs/lajavi-krelac>)
teach.FUT.1sg her.ACC that herself not lie.PRS.3sg that herself
sacrifice.PRS.3sg love.GEN for the sake of
'I will teach her to not lie to herself that she makes sacrifices for the sake of love'

- (6) *Nas* **što se tiče**, *to je sasvim u redu.*
 us.ACC as for this be.PRS.3sg entirely all right
 ‘As for us, this is perfectly all right’
- (7) *Došli smo vama za ljubav.*
 come.PST.PTCP.m.sg be.AUX.PRS.1pl you.DAT for love of
 ‘We came for (love of) you’

It is difficult to draw up an exhaustive list of adpositions in Serbian, especially complex ones. While grammars usually enumerate many simple prepositions, very little attention is devoted to complex prepositions. Popović’s (1966) paper provides, to the best of my knowledge, the most encompassing list of CAs in Serbian by indexing 155 “phrasal prepositions”. Nevertheless, one can legitimately wonder whether this list is exhaustive, and it also contains a series of items which, without being “primary” prepositions (see below), are not morphologically complex in contemporary language (e.g. *zahvaljujući* ‘thanks to’, *shodno* ‘following’, *tokom* ‘during’, *povodom* ‘on the occasion of’, *krajem* ‘by the end of’). This means that in order to have an overview of the adpositional system of Serbian, it is first necessary to draw up an inventory of the adpositions as far as possible, independently of their internal structure, which will be one of my main goals in this chapter. To do this, a working definition of adposition is required. In the following sub-sections, I first offer a brief survey of nominal relation marking in Serbian (see § 3.1), and secondly, I present the definitions of simple and complex adpositions that I adopt here following those proposed in the Introduction to this volume, and further investigate their formal and functional aspects on the basis of data that the adopted definitions allowed me to compile (see §§ 3.2 and 3.3).

3.1 Nominal relation marking and its morphological and syntactic realizations in Serbian

Even though adpositions have several syntactic roles, nominal relation marking is often considered as one of their main functions in the sentence. In Serbian, there are two main types of nominal relation marking: pure case marking (8–10), and the use of adpositions, that is prepositions, in combination with case markers (11–12):

- (8) *deo* *prsta*
 part.NOM finger.GEN
 ‘a part of a finger’

- (9) *šetnja ulicom*
 walk.NOM street.INS
 ‘a walk along the street’
- (10) *profesor istorije*
 professor.NOM history.GEN
 ‘professor of history’
- (11) *puding od vanile*
 pudding.NOM of vanilla.GEN
 ‘vanilla pudding’
- (12) *poklon za Marka*
 gift.NOM for Marko.ACC
 ‘a gift for Marko’

Given the extensive use of both case markers and prepositions in expressing nominal relations, Serbian grammarians and linguists usually talk about *preposition-case constructions* (“predložko-padežene konstrukcije”) in order to stress the distributive nature of the way functional and semantic relations between syntactic constituents are established in the sentence. Thus, it would be wrong to say that in (11) and (12) the prepositions are the only nominal relation markers: both the prepositions and the case morphemes, namely *od + GEN* and *za + ACC*, encode a specific semantic relationship between the given nouns. This situation can be seen as a result of the overlapping of two concurrent sub-systems of relation marking: an older one inherited from Indo-European, primarily based on the use of cases, and a more recent one that developed a wide range of adpositions that play a more or less identical role as cases, but that still require inflected NPs as complements (see among others Belić 1998, t.1–2). Consequently, and as opposed to what generally happens in the process of grammatical change (cf. Lehmann 1991, 2002a), one cannot claim that, in Serbian, the rise of prepositions – either simple or complex ones – has ousted case marking, nor that the possible attrition of case inflection stimulated the prepositions to fill the gap. What seems to have happened is a certain loss of the capacity for some cases to express certain relationships autonomously (cf. Belić 1998, t.1: 92–93).

Combining the two types of relation markers in Serbian offers a broad spectrum of possibilities for meaning construction, certainly far larger than the (seven) cases alone – although highly polysemous –, or prepositions alone – although much more numerous and some of them very polysemous –, could provide. As for the combination of cases and prepositions within prepositional phrases, a large

majority of prepositions govern a single case,¹ but there are also series of mainly simple prepositions governing two (e.g. *na* ‘on, onto’, *s(a)* ‘off, with’, *pred* ‘in front of’ – see 13 and 14) or three cases (e.g. *u* ‘in, into, at’, *za* ‘for, behind, at’ – see 15).

- (13) a. *Opazio je mačku kako sedi*
 see.PST.PTCP.m.sg be.AUX.PRS.3sg cat.ACC how sit.PRS.3sg
na drvetu u dvorištu.
 in(to) tree.LOC in(to) garden.LOC
 ‘He saw a cat sitting in the tree in the garden’
- b. *Komšijin pas je oterao našu mačku na drvo.*
 of neighbor dog.NOM be.AUX.PRS.3sg chase.PST.PTCP.m.sg our.ACC
 cat.ACC in(to) tree.ACC
 ‘The neighbor’s dog chased our cat up the tree’
- (14) a. *Sklonite hleb sa stola.*
 clear.IMP.2pl bread.ACC off table.GEN
 ‘Clear the bread off the table’
- b. *Dolaze sa decom.*
 come.PRS.3pl with children.INS
 ‘They come with children’
- (15) a. *Šta to kriješ u džepu?*
 what it hide.PRS.2sg in(to) pocket.LOC
 ‘What are you hiding in your pocket?’
- b. *Šta to stavljaš u džep?*
 what it put.PRS.2sg in(to) pocket.ACC
 ‘What are you putting into your pocket?’
- c. *U Milice duge trepavice.*
 at Milica.GEN long.NOM eyelashes.NOM
 ‘Milica has long eyelashes.’

As seen in these examples, different governing rules license the construction of different meanings. In (13), the preposition *na* ‘on, at, onto’ describes atelic scenes when combined with NPs in Locative case (see 13a), and telic

¹ There are seven cases in Serbian: Nominative, Genitive, Dative, Accusative, Vocative, Instrumental and Locative. Only Nominative and Vocative cases are never governed by adpositions (cf. Ivić 1983, Stanojčić & Popović 2011).

ones expressing the goal of a spontaneous or caused motion when combined with NPs in the Accusative (see 13b). The preposition *na* can thus appear in two different preposition-case constructions, each of which conveys a distinct meaning: combined with *na*, NPs in the Locative describe sites occupied by a given entity (13a), whereas NPs in the Accusative, combined with the same preposition, denote sites reached by another, necessarily moving, entity (13b). The same meaning alternation is observed for the preposition *u* ‘in, into’, and there are several other meaning alternations of this type in Serbian, as in German and Albanian, for example (see respectively Hüning et al. in this volume, Fagard & Bonnet in this volume). In (14), one can find another, much more radical meaning alternation, in that the preposition *sa* conveys an allative (spatial) meaning in (14a) with an NP in the Genitive, and a comitative meaning with an NP in the Instrumental (14b).

Complex prepositions governing two cases are very rare, but this is the case for some of them, such as *u prilog* ‘in favor of’ and *za ljubav* ‘for love of, for’ in (16) and (17), which can both combine with NPs in the Genitive (examples in a) and Dative (examples in b) cases:

- (16) a. *Ova asocijacija objedinjava aktivnosti u prilog*
 this.f association.NOM gather.PRS.3sg activity.pl.ACC in favor of
mira, tolerancije i ravnopravne saradnje
 peace.GEN tolerance.GEN and equitable.GEN cooperation.GEN
preko 130 velikih gradova. (<http://www.beoforum.rs/>)
 over 130 big.m.pl.GEN city.pl.GEN

‘This association groups activities in favor of peace, tolerance and equitable cooperation for more than 130 big cities’

- b. *Najjači i najpopularniji dokazi*
 strong.m.pl.sup.NOM and popular.m.pl.sup.NOM proof.pl.NOM
u prilog evoluciji imaju ozbiljne
 in favor of evolution.DAT have.PRS.3pl serious.m.pl.ACC
probleme. (<http://www.eden.rs/>)
 problem.pl.ACC

‘The strongest and the most popular proof supporting evolution presents serious problems’

- (17) a. *Dobile smo pravo na (...) rad u*
 obtain.PST.PTCP.f.pl be.AUX.PRS.1pl right.ACC on(to) work.ACC in(to)
firmi za platu i rad kod kuće za ljubav
 firm.LOC for salary.ACC and work.ACC at home.ACC for love of

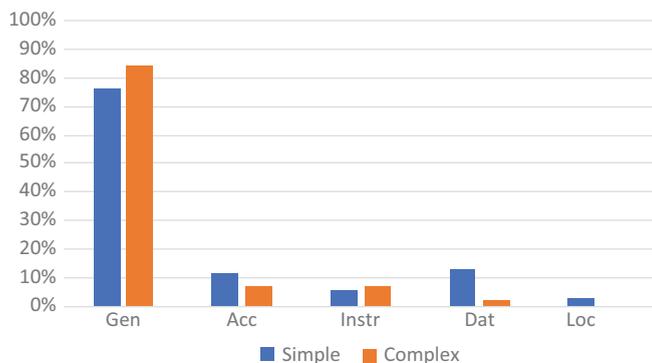
porodice. (<http://domacica.blog.rs>)

family.GEN

‘We acquired the right to work in the firm for a salary and to work at home for the family’

- b. *Ukratko, Atila je učinio sve*
 briefly Attila.NOM be.AUX.PRS.3sg do.PST.PTCP.m.sg all.ACC
za ljubav svom storičaru kako bi
 for love of his.DAT historian.DAT in order to be.AUX.COND.3sg
ličio na Mongola. (<http://www.svevlad.org.rs/>)
 look.PST.PTCP.m.sg on(to) Mongol.ACC
 ‘Briefly, Attila did everything in his power to look like a Mongol, for the sake of his historian’

Among the five cases likely to be combined with simple and complex adpositions in Serbian (see note 1), the Genitive is the one most frequently selected.² The following graph shows a global view of case governing by the simple and complex adpositions collected for the purposes of this study (see Sections 3.2 and 4 for more details about the corpus of adpositions this study is based on):



Graph 1: Adpositions and case governing in Serbian (based on type counts).

By adding up the selection of cases by simple and complex adpositions, it is found that 81% of them combine with the Genitive.

² I will use here the terms ‘selection’ and ‘governing’ when talking about the relation between adpositions and cases, which means that adpositions function as the governing element. Note however that several Serbian grammarians and linguists claim that the role of adpositions is rather to modify or to specify meanings conveyed by cases, which is not wholly improbable (see among others Belić 1998, Stevanović 1991).

The nature of constituents that are likely to be introduced by adpositions in Serbian is another feature to be mentioned when introducing the way they behave. As in many other European languages, in Serbian, most often, adpositions take non-propositional expressions, particularly noun phrases, as complements (see examples 1–4 or 13–17 above), but some other types of syntactic constituents can be introduced by adpositions. Among them, one can find adverbial phrases as in (18), prepositional phrases, as in (19) (cf. Šekrst 2011), or finite and, much more marginally, non-finite subordinate clauses, as shown respectively in (20a–b) and (20c–d):

- (18) a. *Odvedi nas do gore.*
 take.IMP.2sg us.ACC to/till up there
 ‘Take us up there’
- b. *Prva utakmica je zakazana za sutra.*
 first.f match.NOM be.AUX.PRS.3sg scheduled.PASS.PTCP.f.sg
 for tomorrow
 ‘The first match is scheduled tomorrow’
- c. *Od koliko godina sviraš klarinet?*
 from how many year.pl.GEN play.PRS.2sg clarinet.ACC
 ‘How long have you been playing the clarinet?’
- (19) a. *Ostavi bananu za posle ručka.*
 leave.IMP.2sg banana.ACC for after lunch.GEN
 ‘Keep the banana for after lunch’
- b. *Čekala ga je na prozoru do u mrklu noć.*
 wait.PST.PTCP.f.sg he.ACC be.AUX.PRS.3sg on(to) window.LOC until
 in(to) dark.f.ACC night.ACC
 ‘She was waiting for him at the window until late in the dark night’
- (20) a. *Ne treba mi ništa osim da me ostavite na miru.*
 not need.PRS.3sg I.DAT nothing.ACC except that I.ACC
 leave.PRS.2pl on(to) peace.ACC
 ‘I don’t need anything except for you to leave me alone’
- b. *Usluga je odlična s obzirom da je sajt besplatan.*
 service.NOM be.PRS.3sg excellent.f.NOM given that be.PRS.3sg
 site.NOM free.m.NOM
 ‘The service is excellent given that the site is free’

- c. *Nudimo kuvana jela za poneti.*
propose.PRS.1pl cooked.f.ACC dish.pl.ACC for take away.INF
'We propose cooked dishes to take away'
- d. *Stvarno je za razmisliti šta oni tamo rade.*
really be.PRS.3sg for think.INF what they.NOM thither
do. PRS.3pl
'One can really ask what they actually do out there'

The last type of usage raises the question of whether *osim*, *s obzirom* and *za* still function as prepositions or whether one should rather consider them as conjunctions *osim da*, *s obzirom da* and *za* (see the introductory chapter for a discussion). It is worth mentioning that some prepositions can appear without any complement, in which case they display an adverbial distribution, as seen in (21).

- (21) a. *Oni stanuju veoma blizu.*
they.NOM live.PRS.3sg very near/nearby
'They live very nearby'
- b. *Dodji malo posle da mi objasniš.*
come.IMP.2sg little after in order to I.DAT explain.PRS.2sg
'Come back later to explain it to me'

3.2 Simple adpositions in Serbian: A debatable issue

As stated in the Introduction to this volume, simple adpositions can be described as monomorphemic relational expressions whose main function in language is to introduce nouns or noun phrases. Following what is said in the introductory chapter, I assume that simple adpositions are characterized by four main features:

- a. they consist of a single morpheme
- b. they introduce a lower syntactic – *a priori* nominal – constituent
- c. they establish a relation between this constituent and a higher constituent, or the sentence as a whole
- d. they contribute to determining the semantic nature of this relationship

In example (22) below, the preposition *za* 'for' is a single autonomous morpheme that is used to both introduce the NP *životno delo* 'lifetime achievement' and to link it to the noun *nagradu* in order to form the wider NP *prestiznu nagradu za životno delo*. The preposition also plays a very important role in constructing the semantic relationship between the two nouns.

- (22) *Slavni* *reditelj* *dobio* *prestižnu*
 famous.m.NOM director.NOM receive.PST.PTCP.m.sg prestigious.f.ACC
nagradu *za* *životno delo*.
 award.ACC for lifetime achievement.ACC
 ‘The famous director received the prestigious award for lifetime
 achievement’

Potentially, this definition can lead to a larger list of simple prepositions than usually found in the literature or in grammars and dictionaries. This may mean that, from the traditional point of view, the inclusion of some items in the list is somewhat problematic. I nevertheless assume that their functional properties, namely (b–d), fully legitimize their membership in the class. My corpus thus comprises 137 simple prepositions, all of which are listed below.

The feature (a), which introduces a very strict constraint on the internal structure of simple adpositions, makes it necessary to distinguish canonical (or “true”, cf. Klajn 2002) simple adpositions (consisting of a single morpheme) and non-canonical ones, which either violate this morphological constraint or originate from other word classes. Nevertheless, I will consider the latter here as simple because nowadays, graphically, they are unitary forms.

The opposition between canonical and non-canonical SAs, as introduced in the Introduction to this volume, partially matches the distinction between so-called “primary” and “secondary” prepositions (see, among others, Lehmann 2002a, 2019, Zwarts 1997, Fleischer 2002, Hopper and Traugott 2003, Šarić 1999, Bučar 2009, Hill 2012, Dragičević 2012). The two oppositions match only partially because all complex prepositions should also be considered as secondary ones, and it is not certain that all canonical simple adpositions fit into the category of primary adpositions (see Schema 1, § 5). According to Hopper and Traugott (2003: 110), “primary adpositions are thought of as the restricted set of adpositions, often monosyllabic, that indicate purely grammatical relationships, such as *of*, *by*, and *to*”. While most canonical simple adpositions are monosyllabic, they are far from expressing purely grammatical relationships; even the most grammaticalized simple prepositions in Serbian such as *u* ‘in, into, at’, *na* ‘on, onto, at’, *za* ‘for, behind’, *od* ‘from, of’, *sa* ‘off, from, with’ undoubtedly have a series of lexical meanings. The question arises as to whether their grammatical role is not attenuated by the coexistence of the case sub-system, as opposed to languages without any case marking such as French and English where so-called primary prepositions display a resolutely grammatical behavior (for further details, see the Discussion chapter that concludes this volume). As for SAs, the opposition between canonical vs non-canonical SAs proposed here, and that between primary and secondary adpositions match much better if one takes as

the main criterion the transparency of both their morphological structure and their meaning, as suggested by Lehmann (2019: 95):

At a given synchronic stage, the former [primary adpositions] are those which have no – synchronically or diachronically – transparent structure, while the latter [secondary adpositions] are those whose formation is – again synchronically or diachronically – transparent. They are called secondary because they are presumably more recent.

For all these reasons, I maintain in what follows the opposition between canonical and non-canonical adpositions at the expense of the distinction between primary and secondary prepositions.

Canonical SAs are old Slavic adpositions, essentially monomorphemic elements, which evolved from two main sources: adverbs, and nouns in their adverbial usages (see Belić 1998: 92–93). There are about 30 of them, they have no transparent structure and *a priori* display the highest frequency of occurrence among all adpositions (see § 4.3 below), meaning that their function in the language system is fundamentally structuring in nature. Some of them are also used as prefixes (labeled by ^P in the following list), which supports the claim that they originate from adverbs (for a justification, see Belić 1998, vol.1: 94, Stevanović 1991). The following items can be considered as canonical simple prepositions in Serbian:

bez^P ‘without’, *do*^P ‘up to, until’, *iz*^P ‘from, out of’, *ka* ‘toward’, *kod* ‘at, near’, *kroz* ‘through’, *mimo* ‘by, next to’, *na*^P ‘on, onto’, *nad*^{P3} ‘above, overhead’, *niz* ‘down’, *o*^P ‘about’, *od*^P ‘from, of’, *osim* ‘except’, *po*^P ‘around’, *pod*^P ‘under’, *pre*^P ‘before’, *pred* ‘in front of’, *prema* ‘toward’, *pri*^P ‘at’, *protiv* ‘against’, *radi* ‘for’, *sa*^P ‘with, off’, *sem* ‘except’, *spram* ‘opposite’, *suprot* ‘opposite’, *u*^P ‘in, into, at’, *uz*^P ‘up, close to, with’, *van* ‘outside’, *za*^P ‘behind, for’.

Non-canonical SAs are more recent creations; they always have a transparent structure, and generally correspond to polymorphemic expressions. Some of them are obtained by conversion of lexical items, while others are formally compounds but appear nowadays as single words.⁴ This means first that their formation is still transparent, and second that they perform the three aforementioned

³ Although some of these prepositions originate from Indo-European adverbs (e.g. *nad* ‘above, overhead’, *pod* ‘under’, *pred* ‘in front of’ – cf. Belić 1998: 94), I consider them as canonical prepositions because they have no transparent structure, and, in synchrony, their adverbial uses no longer exist. They are thus clearly opposed to such prepositions as *preko* ‘over, across, by’, *suprotno* ‘unlike’, *niže* ‘below, lower down’, whose relation to the category of adverb is still noticeable.

⁴ Some exceptions are to be noted: *bojeći se* ‘for fear of’ and *plašeći se* ‘for fear of’, as they are originally verb forms, keep the reflexive pronoun included in the lexical entry (*bojati se*, *plašiti se* ‘to be afraid of, to fear’). Consequently, they are strictly speaking made of two morphemes but I will keep them in the sub-class of simple adpositions because they originate from a single lexical unit.

functions (b, c, d) in a cumulative way, contrary to what complex adpositions do (see § 3.2.2 below).

As for converts, there are about 57. These adpositions are obtained from nouns, adverbs and verbs (see Belić 1998, t.1/t.2, Klajn 2002, t. 2: 386–388, Bučar 2009):

- there are 30 simple prepositions obtained from nouns, mainly in the instrumental case⁵ (see Šarić 1999, Glušac & Rišner 2016): *dobrotom* ‘by the goodness’, *dužinom* ‘alongside’, *igrom* ‘by the play of’, *krajem* ‘at/by the end of’, *početkom* ‘at the beginning of’, *polovinom* ‘in the middle of’, *pomoću* ‘by means of’, *posredovanjem* ‘by the help of’, *posredstvom* ‘by means of’, *povodom* ‘about’, *prilikom* ‘on the occasion of’, *putem* ‘by’, *silom* ‘by force of’, *sredinom* ‘by the middle of’, *stopama* ‘following’, *širom* ‘throughout, all around’, *tokom* ‘during’, *tragom* ‘following’, *tragovima* ‘following’, *zaslugom* ‘thanks to, by means of’, and marginally in the accusative (see Dragičević 2012, Glušac and Rišner 2016): *čelo* ‘at the top of’, *dno* ‘at the bottom of’, *duž* ‘along’, *kraj* ‘next to’, *među* ‘among’, *mesto* ‘instead of’, *put* ‘towards’, *sred* ‘in the middle of’, *vrh* ‘at the top of’.
- next, there are 17 simple prepositions formed from adverbs: *blizu* ‘near, close to’, *niže* ‘below, lower down’, *oko* ‘around’, *posle* ‘after’, *pozadi* ‘behind’, *preko* ‘over, across, by’, *proporcijalno* ‘in proportion to’, *protivno* ‘contrary to’, *razmerno* ‘in proportion to’, *saglasno* ‘in compliance with’, *saobrazno* ‘in accordance with’, *shodno* ‘according to’, *sledstveno* ‘according to’, *slično* ‘similarly to’, *srazmerno* ‘in proportion to’, *suprotno* ‘unlike’, *više* ‘above’.
- finally, there are 10 prepositions obtained by conversion of verbal forms, that is of the present participle such as *blagodareći* ‘thanks to’, *bojeći se* ‘for fear of’, *čekajući* ‘waiting for’, *izuzimajući* ‘excluding’, *plašeći se* ‘for fear of’, *uključujući* ‘including’, *zahvaljujući* ‘thanks to’, and of the past participle *isključivši* ‘apart from’, *izuzev* ‘except for’, *izuzevši* ‘apart from’.

Regarding simple adpositions that are considered here as compounds, they all result from a kind of fusion of two words into a single one. I indexed 45 adpositions of this type in Serbian. They are formed according to three patterns and almost all of them occur with NPs in the genitive case:

- [P N] (35): *dovrh* ‘up to the top of’, *između* ‘among, between’, *nadno* ‘at the bottom of’, *nadohvat* ‘within the range of’, *nadomak* ‘within the range of’,

⁵ According to Glušac and Rišner (2016), the process of fossilization of NP in the Instrumental case conveying either a spatial or a temporal meaning occurs in Old Slavic. Their rise can, however, be traced back to the 19th century, which makes them more recent than converts formed from nouns in the accusative.

nakon ‘after’, *nakraj* ‘at the end of’, *namesto* ‘instead of’, *naoči* ‘in front of’, *naočigled* ‘in front of’, *nasred* ‘in the middle of’, *nauštrb* ‘to the detriment of’, *navrh* ‘at the top of’, *podno* ‘at the foot of’, *pokraj* ‘next to’, *poput* ‘like, as’, *posred* ‘midst’, *pored* ‘next to’, *potkraj* ‘by the end of’, *povrh* ‘above, in addition to’, *prekoputa* ‘across, opposite’, *skraj* ‘next to’, *udno* ‘at the bottom of’, *udnu* ‘at the bottom of’, *uime* ‘on behalf of’, *uinat* ‘in spite of, against’, *ukraj* ‘beside, next to’, *umesto* ‘in place of’, *uoči* ‘on the eve of, on the brink of’, *uprkos* ‘despite’, *usled* ‘subsequent to’, *usred* ‘in the midst of’, *ususret* ‘towards’, *uvrh* ‘atop’, *zbog* ‘because’

- [P P] (11): *ispod* ‘below’, *ispred* ‘in front of’, *iza* ‘behind’, *iznad* ‘above’, *naspram* ‘opposite’, *nasuprot* ‘opposite’, *otpre* ‘since’, *poradi* ‘for the sake of’, *usuprot* ‘despite’, *uzduž* ‘along’, *zarad* ‘for the sake of, for’
- [P Adv] (2): *izvan* ‘outside’, *poviše* ‘above’.

These forms are considered by many linguists as “complex adpositions” (see Klajn 2002: t 1: 139). They thus raise the question of their morphological formation, because they seem to stand between univerbation and composition. According to Lehmann (2002a: 135), the difference between these two processes can be explained as follows:

Univerbation is restricted to the syntagmatic axis and may affect, in perhaps idiosyncratic ways, any two particular word forms which happen to be habitually used in collocation. Composition, as a schema of word-formation, presupposes a paradigm in analogy to which it proceeds and affects a class of stems according to a structural pattern.

Given the wide range of possibilities for fusing syntagma into single words, the formations that are considered here as compound (non-canonical or secondary) SAs are restricted in number, and could be rather seen as occasional, and hence as resulting from univerbation. On the other hand, their formation presents some degree of regularity: they all display a preposition in initial position and follow three patterns of formation ([P N], [P P] and [P Adv]), which suggests a process of composition. This hypothesis is difficult to substantiate, however, because of the very limited scope of paradigms that are formed in this way, and because these formations originate from syntactic constructions, which is not characteristic of composition according to Lehmann (2019: 102). The process of fusion at issue thus appears as occasional rather than systematic, which further supports the hypothesis of a more advanced degree of lexicalization and/or grammaticalization of these expressions. This hypothesis is all the more likely given that the most productive pattern of formation [P N] is also the main schema of formation for complex adpositions (see 4.2.1 below), which explains why some scholars taking rather a diachronic perspective consider them as complex adpositions (see, for

instance, Popović 1966). Another argument in favor of this hypothesis is that for a series of so-called compound simple prepositions, there exists a corresponding analytical form: *na dohvat / nadohvat* ‘within range of’, *u inat / uinat* ‘in spite of, against’, *u sred / usred* ‘in the middle of’, *u susret / ususret* ‘towards, to go to meet’, *u ime / uime* ‘on behalf of’, and so on:

- (23) a. *ona [srpska privreda] trenutno prolazi kroz fazu u susret otvaranju modernizacije EU tržišta.* (https://www.b92.net/)
 she.NOM (Serbian economy) currently pass.PRS.3sg through phase.ACC towards opening.GEN modernization.GEN EU market.GEN
 ‘it (Serbian economy) is going through a phase of modernization in view of the opening of the EU market’
- b. *poslali su jednu legiju ususret njima* (http://forum.arheo-amateri.rs)
 send.PST.PTCP.m.pl be.AUX.PRS.3pl one.f.ACC legion.ACC towards they.DAT
 ‘They sent one legion to go to meet them’
- (24) a. *ne može Vučić da odlučuje u ime suda.* (http://www.rts.rs)
 not can.PRS.3sg Vučić.NOM that decide.PRS.3sg on behalf of court.GEN
 ‘Vučić cannot decide on behalf of the court’
- b. *gostima je (...) zahvalio jedan od domaćih novinara uime navijača.* (http://www.balkanmagazin.net/)
 gest.m.pl.DAT be.AUX.PRS.3sg thank.PST.PTCP.m.sg one.NOM of national.pl.GEN journalist.pl.GEN on behalf of supporter.pl.GEN
 ‘The guests were thanked by a national journalist on behalf of supporters’

This sub-group of SAs thus clearly acts in the system as an active switching area between simple and complex adpositions: their diachronic analyzability pushes them towards CAs, but by the loss of boundaries and the coalescence of their formants in synchrony, as well as by their frequency of use (see § 4.3 below), they side with SAs.

3.3 Defining complex adpositions

Dealing with complex adpositions implies a preliminary selection of items belonging to this word sub-class. Such a selection however needs as precise a working definition as possible of which sequences deserve the class membership, and which do not. This section aims to provide some categorial, semantic and formal cues for a comprehensive definition of CAs.

3.3.1 Towards a general definition of complex adpositions

Many Slavic linguists use the term complex prepositions to denote what are considered here as non-canonical compound prepositions (*umesto* ‘in place of’, *ispred* ‘in front of’), and oppose them to multiword expressions such as *u odnosu na* ‘with respect to’, *u toku* ‘during’, considered as prepositional locutions. In this paper, I do not differentiate between complex prepositions and prepositional locutions. I shall rather follow the definition given in the Introduction to this volume, and define complex adpositions as relators having the same functions and distributional properties as simple adpositions (b–d) (cf. Hoffmann 2005, Fagard *et al.* this volume), but with different formal characteristics (e–f):

- e. they are MORPHOLOGICALLY complex, multiword expressions
- f. canonically, functions (b-d) are performed by different components:
 - i. a simple adposition (or case marker) links the lower syntactic constituent
 - ii. another simple adposition (or case marker) establishes the relation with the higher constituent
 - iii. a lexical nucleus determines the semantic nature of the relation

While the fulfilment of the three aforementioned functions can be assigned to the expression as a whole, as done for simple prepositions (cf. e.g. 22 above), the definition above distributes them onto its different parts. Admittedly, the distributiveness of this definition is fundamentally contrary to the strong version of the definition of complex adpositions, as proposed for instance by Quirk *et al.* (1985: 671), claiming that they are sequences “indivisible both in terms of syntax and in terms of meaning”. However, this particularly categorical conception of CAs does not adequately account for the reality of the way these expressions function in languages, because they generally display different degrees of cohesiveness and fossilization, which directly calls into question their supposed indivisibility.

Example (25) below with the sequence *u skladu sa* ‘in accordance with’ illustrates, in an iconic way, the hypothesized distributional nature of complex

prepositions in Serbian. Consequently, it is suggested that the role of the final preposition *sa* ‘with’ is to integrate – together with the case marking – the lower constituent, namely the noun *priroda* ‘nature’ into the PP (function i.), the role of the initial preposition *u* ‘in, into, at’ is to integrate the PP into the higher NP constituent by linking it to the head noun *dizajn* ‘design’ (function ii.), whereas the nominal nucleus *sklad* ‘accordance, harmony’ serves to specify the semantic relationship between the two NPs (function iii.).

- (25) *moderan dizajn u skladu sa prirodom*
 modern.NOM design.NOM in accordance with nature.INS
 ‘the modern design in accordance with nature’

In the absence of the final preposition, the division of labor is somewhat modified in that its role is recovered by the case marking on the lower constituent, as shown in the following example, where the linking of the complement is performed by the structural Genitive case:

- (26) *međunarodni samit mladih u oblasti pozorišta*
 international.m.NOM summit.NOM young.pl.GEN in domain theater.GEN
 ‘the international summit of young people in the field of theater’

Given the overlapping roles of prepositions and cases in nominal relation marking in Serbian (see § 3.1 above), the variant [P₁ N] can be considered as canonical as the variant with P₂ in the previous example. Moreover, this pattern forms many more CAs than the pattern [P₁ N P₂] as will be shown in Section 4.2.1 below, suggesting that cases have a strong linking power, hence their structural vitality in nominal relation marking. Further, there are several other variants that more or less match with the iconic representation of the distribution of functions onto different formants of CAs. For example, one can observe the lack of the initial preposition, as illustrated in (27)-(28), or the complete absence of prepositions, as in example (29):

- (27) *još jedna planeta nalik na zemlju*
 else one.f.NOM planet.NOM similar to Earth.ACC
 ‘one more planet like (similar to) Earth’

- (28) *revizija računa počev od januara*
 revision.NOM account.pl.GEN beginning from January.GEN
 ‘the auditing of accounts from January’

- (29) *imamo dobre vesti što se tiče posla*
 have. PRS.3sg good.f.pl.ACC news.pl.ACC concerning work.GEN
 ‘we have good news about work’

Because of these accommodations, the iconic projection of the three functions performed by CAs (i–iii), as sketched above, should be rather considered as an idealization of the way they operate in the language system, even though it applies to the vast majority of CAs in Serbian (see Hüning *et al.* this volume for similar considerations on West Germanic languages). In order to distinguish prototypical CAs from those that are in some way deviant, the opposition between *canonical* and *non-canonical* formations can be used again. Thus, a canonical CA can be defined as one in which each of the three functions of adpositions enumerated in (f) (i.–iii.) is performed by a different component (including case marking), while a non-canonical complex adposition is one in which there is no such one-to-one mapping (see, for a similar treatment of CAs in French, Fagard *et al.* b this volume). I will come back to the issue of canonical CAs in Serbian in section four below (see § 4.3).

3.3.2 Some more criteria

Defining complex adpositions, however, is a much more complex task than suggested by the proposed definition, as reflected by the considerable amount of research on this topic in linguistics. Generally considered as part of the grammatical system, in line with and under the influence of what is done for simple adpositions, complex adpositions remain an ill-established class because of the extreme difficulty of delineating them from other formally similar or identical complex sequences (see among others Hoffmann 2005: 25–35 for an overview of the relevant literature, Quirk *et al.* 1985, Borillo 1997, Fagard and De Mulder 2007, Melis 2003, Stosic and Fagard 2019). A whole range of diagnostics has thus been applied to assess, for various languages, whether a given sequence is a free construction or an indivisible unit. Since in European languages CAs are mainly formed by combining an initial preposition, a noun (with or without a determiner), and a final preposition, most of these criteria are meant to study the so-called “PNP-constructions”, and more precisely to assess the extent to which their individual elements have lost their syntagmatic autonomy. In what follows, I give a brief overview of the most frequently used diagnostics, illustrating with Serbian data, whenever possible, the way in which they are used. These criteria are collected from different studies on the topic, particularly from Popović (1966), Quirk *et al.* (1985), Heine (1989), Seppänen, Bowen, and Trotta (1994), Borillo

(1997, 2002), Adler (2001), Melis (2003), Hoffmann (2005), Fagard and De Mulder (2007), Leeman (2007), Kurzon and Adler (2008), Stosic and Fagard (2019).

First of all, there is a set of formal – either morphological or syntactic – criteria:

- variation of the initial relator, which is generally a preposition (P₁) – compared to *u visini od* ‘of up to’ (30a) which allows variation, *u skladu sa* ‘in compliance with’ (30b) is a better candidate for inclusion in the class:

- (30) a. *u visini od* ‘of up to’ / *na visini od* ‘to the height of, at the altitude of’ / *pri visini od* ‘by/toward the height of’
 b. *u skladu sa* ‘in accordance with’ / **na skladu sa* / **pri skladu sa*

- variation of the final relator, which can be either a preposition (P₂) or a case – compared to *na putu ka* ‘on the path towards’ (31a), *u odnosu na* ‘with respect to’ (31b) appears as more fossilized:

- (31) a. *na putu ka* ‘on the path towards’ / *na putu za* ‘on the path to’ / *na putu iz* ‘on the path from’ / *na putu prema* ‘on the path towards’
 b. *u odnosu na* ‘with respect to’ / **u odnosu za* / **u odnosu ka*

- presence / absence of the determiner in languages with this word class – structures without a determiner suggest a kind of syntactic deviation, as in the French examples in (32a–c),⁶ and hence a higher degree of coalescence, compared to their respective variants which bear/include a determiner, and are syntactically regular, in (32d–f):

- (32) a. *à côté de* la route ‘next to the road’
 b. *à base de* fruits ‘fruit based’
 c. *sous contrôle de* la police ‘under police control’
 d. *sur le côté de* la route ‘at the side of the road’
 e. *à la base de* l’accord ‘at the basis of the agreement’
 f. *sous le contrôle de* la police ‘under the control of the police’

- variation of the determiner if it is present in the construction – the French expression *le long de* in (33b) does not license any variation of the determiner, and can be considered as an indivisible unit whose parts have entirely lost their autonomy, as opposed to *au bout de*, which allows this variation, as illustrated in (33a):

⁶ Examples (32) and (33) are from French because Serbian does not have determiners.

- (33) a. *au bout de* (=à le *bout de*) ‘at the end of, after’ / *à un bout de* ‘at one end of’
 b. *le long de* ‘along’ / **un long de*
- variation of the nominal nucleus in number – one can oppose *u toku* (*tok* n.sg ‘flow’) used only with a singular noun or *za potrebe* (*potreba* n.pl ‘need’) used only with a plural noun, and *na margini* or *u okviru* that display variation in number. This does not mean that one of the two variants is not a CA:
- (34) a. *na margini* ‘at the margin of’ / *na marginama* ‘at the margins of’
 b. *u okviru* ‘within the scope of’ / *u okvirima* ‘within the scopes of’
 c. *u toku dana* ‘during the day’ / **u tokovima dana*
 d. *za potrebe ustanove* ‘for the needs of the institution’ / **za potrebu ustanove*
- replacement of the lexical nucleus by a term with a related meaning (see 35a–b) – if impossible, this suggests a relative fixedness of the sequence (see 35c–d). Note however that in many cases, this variation is blocked by the absence in the lexicon of suitable synonyms for the lexical nucleus:
- (35) a. *na sred* ‘in the middle of’ / *na sredini* ‘in the middle of’
 b. *u okolini* ‘in the vicinity of’ / *u blizini* ‘near’
 c. *iz oblasti* ‘in the area of’ / **iz polja*
 d. *na polju* ‘in the field of’ / **na oblasti*
- modification of the lexical nucleus by adjectives if the nucleus is nominal, or by adverbials in other cases – CAs can be expected not to allow any modification of the nucleus, unlike syntactically free constructions. There obviously exist both CAs that block any modification, such as *za razliku od* ‘unlike’ in (36a), and those where it is possible to a greater or lesser extent, as in the case of *u skladu s(a)* ‘in accordance with’ and *na Adj strani* ‘on the Adj side of’ (36b–c) (see § 4.2.3 for further discussion):
- (36) a. **za izvesnu / određenu / malu razliku od* ‘in certain / given / little difference from’
 b. *u relativnom / određenom / izvesnom skladu s(a)* ‘in relative / given / certain accordance with’
 c. *na istoj / levoj / desnoj / prednjoj / suprotnoj strani* ‘on the same / left / right / front / back side of’

- omission of the final relator and complement – the sequences *u skladu sa* ‘in accordance with’ (39b) and *po ceni od* ‘at the cost of’ (40b) do not license such an omission, which makes their status of CAs as fixed constructions more relevant than that of *istočno od* ‘to the east of, east of’ (37) and *na početku* ‘at the beginning of’ (38) that do not so strongly require the presence of a complement (see 37b and 38b):
- (37) a. *toranj je istočno od grada* ‘the tower is to the east of the town’⁷
 b. *toranj je istočno* ‘the tower is to the east’
- (38) a. *pojavió se na početku* filma ‘he appeared at the beginning of the movie’
 b. *pojavió se na početku* ‘he appeared at the beginning’
- (39) a. *te odredbe su u skladu sa zakonom* ‘these provisions are in accordance with the law’
 b. **te odredbe su u skladu* ‘these provisions are in accordance’
- (40) a. *kupljeno po ceni od 10 dinara* ‘bought at the cost of 10 dinars’
 b. **kupljeno po ceni* ‘bought at the cost of’
- replacement of the final relator and complement by possessives – this diagnostic opposes *u vreme* ‘in the time/days of’ and *na štetu* ‘at the expense of’ to *u oblasti* ‘in the field of’ and *preko puta* ‘opposite, over, across’, suggesting that the latter two sequences are more fixed than the former two:
- (41) a. *u vreme mojih roditelja* ‘in the time/days of my parents’
 b. *u njihovo vreme* ‘in their time/days’
- (42) a. *na štetu studenata* ‘at the expense of students’
 b. *na njihovu štetu* ‘at their expense’
- (43) a. *u oblasti ekonomije* ‘in the field of economics’
 b. **u njenoj oblasti* ‘in its field’
- (44) a. *preko puta banke* ‘opposite the bank’
 b. **preko njenog puta*

7 Very short examples such as this one are not glossed.

- fronting – only relatively independent constituents can be moved to the initial or final position of the sentence; this should not be possible in the case of CAs, as shown in (45), and as opposed to (46) where *problem s gubitkom kose* ‘problem with loss of hair’ is a free syntagmatic construction allowing dislocation of *s gubitkom kose*:

- (45) a. *crkva je na dva koraka od glavnog trga* ‘the church is a few steps away from the main square’
 b. **od čega je crkva na dva koraka?* ‘lit. from what the church is a few steps away?’

- (46) a. *on ima problem s gubitkom kose* ‘he has a problem with loss of hair’
 b. *s čim ima problem?* ‘lit. with what does he have a problem?’

- insertion of different types of modal adverbials – according to some scholars, CAs as single units should not license any insertion of external elements. The reality of linguistic data from various languages seems to contradict this claim, because it happens that even very fixed CAs allow the addition of modal adverbials. This double behavior of CAs is also attested in Serbian, although it remains relatively marginal in corpora:

- (47) a. *počev na primer od tog izuzetnog ostvarenja* ‘beginning for example from this exceptional realization’
 b. *u odnosu bar na druge dve stavke, put i brzina* ‘in relation to at least two other items, the road and the speed’ (<https://www.b92.net>)

- (48) a. **u na primer okolini* ‘lit. in for example vicinity of’ (cf. *na primer u okolini* ‘for example in the vicinity of’ / *u okolini na primer* ‘in the vicinity of for example’)
 b. **u zamenu svakako za* ‘in exchange of course for’

- coordination – only free, functionally identical constituents can be coordinated, so that coordinating candidate sequences with simple prepositions is a way to evaluate their categorical status, as in (49). Note, however, that the impossibility of coordinating two elements does not necessarily mean that they are not functionally equivalent – several semantic features must also be considered:

- (49) a. *vodič će biti u blizini ili tačnije ispred crkve* ‘the guide will be in the vicinity or more precisely in front of the church’

- b. *na čelu i iza kolone su bili naoružani ljudi* ‘at the head of and behind the column were armed people’

In addition to these syntactic and morphological diagnostics, there is also a series of semantic criteria that are invoked in defining complex adpositions. The most important of them are:

- semantic equivalence with a simple adposition, as in (50). Note, however, that this test is of very limited scope because of the discrepancy between the relatively limited set of SAs and the large number of CAs, which uncontroversially involves a much larger range of meanings conveyed by the latter:

- (50) a. *naš razgovor u vezi sa advokatom* ‘our discussion related to the lawyer’
 b. *naš razgovor o advokatu* ‘our discussion about the lawyer’

- semantic bleaching – if the sequence does not express the lexical concept that its nucleus entails, but is rather used for linking together two other contentful elements, then it could be considered as a CA. The Serbian examples (51) and (52) are thus completely acceptable even though one cannot talk about the “verge of the catastrophe” (**rub katastrofe*) or about the “heart of a canyon” (**srce kanjona*). The loss of semantic content is usually considered as a typical characteristic of more advanced degrees of grammaticalization, and it is observed in several CAs in Serbian.

- (51) *Argentina je pre toga bila*
 Argentina.NOM be.AUX.PRS.3sg before that.GEN be.PST.PTCP.f.sg
na rubu katastrofe (<http://www.politika.co.rs>)
 on verge catastrophe.GEN
 ‘Argentina was before that on the verge of catastrophe’

- (52) *obilazak mnogih atraktivnih lokaliteta*
 visit.NOM many.m.pl.GEN attractive.m.pl.GEN place.pl.GEN
u srcu kanjona Tare. (<http://www.mojodmor.rs/>)
 in heart canyon.GEN Tara.GEN
 ‘a visit of many attractive locations in the heart of Tara canyon’

- context extension is another feature closely related to semantic bleaching – it corresponds to the possibility of combining a given expression with NPs that could not be selected in their referential usages. This is the case, for example,

of the sequence *na pola puta od* ‘halfway to’ in (53), or *na rubu* ‘on the verge of’ and *u srcu* ‘in the heart of’ in (51) and (52), which are all employed in a much larger set of contexts than in their primary spatial meanings:

- (53) *Tako da ovu moju literarnu zajebranciju pišem*
 so that this.ACC my.f.ACC literary.ACC knot.ACC write.PRS.1sg
sada na pola puta od ručka do večere. (<http://nervozni.blog.rs>)
 now on half way from lunch.GEN until dinner.GEN
 ‘so I’m working on my literary knot halfway between lunch and dinner’

Concerning Serbian more particularly, Popović (1966) suggests some other cues for the identification of CAs:

- the maintenance of a form even though variations of some parts of the sequence are largely attested in other contexts – for instance, the expression *s obzirom na* ‘given’ never varies, as was to be expected; the variant **sa obzirom na* is not used, whereas in this type of morphophonological context, the variation *s/sa* is largely attested (cf. *sa obavezom da* ‘with the obligation to’, *sa objašnjenjem* ‘with the explanation’, and so on)
- the tone shift from the nucleus onto the proclitic, that is onto P_i, even in geographical areas where this phonological phenomenon is disappearing or missing. This phenomenon is observed in the case of the expressions: *u ime* > *ù ime* ‘on behalf of’, *u srèd* > *ù sred* ‘in the middle of’
- phonological erosion as seen in the case of *s obzirom na* > *obzirom na* ‘given’, *u vezi sa* > *u vezi* ‘in relation to’, and so on
- the increase in frequency of occurrence generally follows the previously mentioned possible evolutions, as demonstrated in research in the field of grammaticalization
- the existence in a foreign language (in contact with Serbian) of an equivalent whose adpositional status is well established. If, in addition, the borrowing from a certain language is proven, this can be considered as evidence of the adpositional status of a given expression in Serbian.

There are two points to be noted about these criteria. First, most of them are naturally, and not surprisingly, invoked within the larger scope of research on grammaticalization phenomena, which is consistent with the fact that the formation of adpositions, both simple and complex, is a specific instance of grammaticalization. Second, a growing body of studies dealing with issues of the delineation of the class of CAs has clearly shown that, for different reasons, none of these criteria are necessary or sufficient for a valid evaluation of the status of

CAs (see, among others, Borillo 1997, Adler 2001, Melis 2003, Hoffmann 2005, Fagard and De Mulder 2007, Stosic and Fagard 2019). An efficient method of identification should jointly consider a set of morphological, syntactic, phonological, semantic, and evolutionary features, as well as the frequency of use of CAs in large corpora. For these reasons, in this study, the aforementioned diagnostics were primarily used as possible indicators of category membership, and not as decisive criteria (see § 4.1 below for further methodological details).

4 Complex adpositions in Serbian

In this section, I describe in detail the system of CAs found in Serbian. In addition to simple prepositions, Serbian has a great number of expressions that are functionally equivalent to simple prepositions, and that are considered by Serbian grammarians and linguists as complex prepositions, or prepositional locutions. The rise of CAs in Serbian occurred in the 18th and 19th centuries during the Enlightenment, and according to Popović (1966: 195–196), it should be related to “the development of modern sciences, administration and journalism”. The need to develop complex relators can be explained, according to the author, by the fact that speakers’ experience had “abruptly widened”, and relationships between notions had begun to be captured and expressed in a more in-depth and more accurate manner. Since the available devices in the language system of the time were not sufficient for these new communication needs, the creation of CAs was largely facilitated (see also Jelínek 1964, Melíšková 2008, Hill 2012, Gavrilenko 2013, for similar statements about the evolution of Czech and Russian complex prepositions). Another factor is stressed by Popović, namely language contact, i.e. the evident impact of foreign languages on the Serbian system within the more general process of the accelerated intellectualization of Serbian culture in the 18th and 19th centuries under the influence of Enlightenment writers and philosophers. This societal evolution went hand in hand with the development of written discourse, which further supported the language system change in question. These external factors are commonly considered in Serbian linguistics as the main driving forces of the expansion of CAs.

4.1 Indexing complex adpositions in Serbian: Methodological considerations

One of the main aims of this chapter is to provide as exhaustive a list of complex prepositions in Serbian as possible. However, indexing complex adpositions in

languages is not a trivial task for both theoretical and practical reasons, as is clearly suggested by the papers in this volume and many other studies on the topic (cf. Borillo 1997, 2002, Melis 2003, Stosic & Fagard 2019 for French, Hoffmann 2005 for English, Cifuentes Honrubia 2003 for Spanish, Lima 2014, 2019a, 2019b, for Portuguese, and so on). Caught between their structuring role at the syntactic level and their substantial lexical load, complex adpositions are deemed to constitute a sub-class with ill-defined, fuzzy boundaries. More specifically, they are known to be distributed along a cline that goes from prototypical items, close to simple adpositions, to borderline items, close to free structures. This is why the sub-class of complex adpositions should be considered as a continuum between sequences that are formally entirely frozen and semantically opaque and those that are syntactically fairly free and semantically significantly transparent (see Melis 2003, Hoffmann 2005, Stosic & Fagard 2019). This fact makes it extremely difficult to establish a reliable set of criteria to build the inventory of complex adpositions in a given language.

Given these theoretical difficulties and the lack of an exhaustive inventory of complex adpositions in Serbian, their indexing was conducted based, first of all, on the definition and criteria given in Section 3.3 above, next, on the relevant literature on the topic (cf. Lalević 1955, Ivić 1957–58, 1983, Popović 1966, Švačko 1994, Šarić 1999, Rišner 2001, Klajn 2002, Bučar 2009, Dragičević 2012, Krijezi 2012, Šekrst 2011, 2013, Glušac and Rišner 2016), as well as on different Serbian grammars and lexicographic resources (cf. Stevanović 1991, Stanojčić and Popović 2011, Mrazović 2009, RMS). Finally, some items were introduced due to the comparison with French where the inventory of complex adpositions is better established (cf. Borillo 1997, 2002, Melis 2003, Stosic and Fagard 2019, Stosic forthcoming a). The identification of translational equivalents in Serbian was performed by using the French-Serbian parallel corpus ParCoLab (cf. Miletic, Stosic, and Marjanovic 2017) and two bilingual dictionaries (Putanec 1995, Jovanović 2014). The list of complex adpositions established in this way made it possible to uncover the main patterns of formation of complex adpositions in Serbian (see § 4.2.1 below). These schemas were subsequently used to detect other potential candidates for inclusion in the sub-class by exploring a tagged version of a 477-million-word monolingual Serbian corpus srWaC (v. 1.2), available in Sketch Engine.⁸ Further, a manual verification of candidate sequences obtained by automatic extraction from srWaC was performed. More precisely, only sequences

⁸ The Serbian web corpus (srWaC) is a Serbian corpus made up of texts collected from the Internet by Tomaž Erjavec and Nikola Ljubešić. The corpus was created in January 2014 (see Ljubešić & Klubička 2014).

that fit the retained definition and whose frequency is equal to or higher than 0.1 pmw (i.e. per million words) were included in the list. This means that all candidate expressions were evaluated for frequency of occurrence in the same corpus (see Section 4.3 for more details). The resulting inventory thus includes, in accordance with the idea of a continuum, both prototypical complex adpositions and borderline items whose membership in the sub-class is questionable. Consequently, my working corpus of complex adpositions in Serbian contains almost 400 items, which will be described in the following sections.

4.2 The morphosyntax of complex adpositions in Serbian

This sub-section aims to both bring to the fore and to describe the main construction schemas of complex adpositions in Serbian, by reporting their morphological and syntactic features.

4.2.1 Patterns of formation

Complex adpositions in Serbian are formed following a series of morpho-syntactic patterns. They can be grouped according to the nature of their lexical nucleus into three broad types: CAa with nominal, verbal and adverbial nuclei. Each group contains several more specific patterns of formation:

N. Complex adpositions with a **nominal nucleus**

[P ₁ N]	(e.g. <i>na kraju</i> ‘at the end of’, <i>u toku</i> ‘during’)
[P ₁ N P ₂]	(e.g. <i>u odnosu na</i> ‘with respect to’, <i>u vezi sa</i> ‘in relation to’)
[P ₁ Adj N]	(e.g. <i>s leve strane</i> ‘on the left of’)
[P ₁ Adj N P ₂]	(e.g. <i>na suprotnoj strani od</i> ‘on the opposite side of’)
[P ₁ Adj <i>i</i> Adj N] ⁹	(e.g. <i>s jedne i druge strane</i> ‘on each side of’)
[P ₁ N P ₂ N]	(e.g. <i>s kraja na kraj</i> ‘from end to end’)
[P ₁ N <i>i</i> N]	(e.g. <i>na milost i nemilost</i> ‘at the mercy of’)
[P ₁ Adj <i>i</i> N P ₂ Adj]	(e.g. <i>s jednog kraja na drugi</i> ‘from one to the other side of’)
[N P ₂]	(e.g. <i>nalik na</i> ‘like’, <i>obzirom na</i> ‘given’)
[P ₁ Adv N P ₂]	(e.g. <i>na pola puta od</i> ‘halfway to’)

⁹ The element *i* is a coordinating conjunction, which corresponds to *and* in English.

V. Complex adpositions with a **verbal nucleus**

- [V P₂] (e.g. *počev od* ‘from, since’, *sudeći po* ‘to judge from’)
 [V P N] (e.g. *imajući u vidu* ‘bearing in mind’)
 [Pron V] (e.g. *što se tiče* ‘as for, concerning’)
 [Neg V P₂] (e.g. *ne gledajući na* ‘regardless of’)

Adv. Complex adpositions with an **adverbial nucleus**

- [Adv P₂] (e.g. *zavisno od* ‘depending on’, *zapadno od* ‘to the west of’)
 [Adv Adv] (e.g. *obrnuto srazmerno* ‘inversely proportional to’)
 [P₁ Adv P₂] (e.g. *za manje od* ‘in less than’)

Judging on the basis of the number of patterns each group displays, nouns are the main source for constructing complex adpositions in Serbian, as in many other languages. Consequently, and not surprisingly, complex adpositions with a nominal nucleus are by far the most numerous, as suggested by Table 2:

Table 2: Nature of lexical nucleus of CAs in Serbian.

Nature of nucleus	Number of CAs	%
Nominal	361	90%
Adverbial	26	6.5%
Verbal	14	3.5%
Total	401	

It is worth noting that such a massive use of nouns in forming CAs in Serbian represents a considerable shift within the language system. Originally, i.e., in late Old Slavonic (11th century) and early Old Serbian, prepositions were exclusively formed from adverbs (cf. Meillet 1902, Belić 1998). In Old Serbian, there started to emerge, first, prepositions resulting from the conversion of nouns (e.g. *kon* ‘end’ > *kod* ‘at, near, *međa* ‘limit’ > *među* ‘among’, etc), second, by the 15th century, those combining two prepositions, and finally from the 17th century on, prepositions formed of a preposition and a noun (e.g. *pokraj* ‘next to’, *usred* ‘in the middle of’).¹⁰

¹⁰ I would like to thank Slobodan Pavlović (University of Belgrade) for these historical insights into the evolution of prepositions in Serbian.

The most productive pattern of formation of complex adpositions in Serbian is made up of two elements, namely an initial preposition and a noun [P1 N], introducing in 98% of cases a NP in genitive case:

- (54) *Ovako je izgledao Beograd*
 thus be.AUX.PRS.3sg look like.PST.PTCP.m.sg Beograd.NOM
za vreme Karađorđeve vlasti.
 during Karađorđe's.f.GEN rule.GEN
 'that is what Belgrade looked like under the rule of Karađorđe'
- (55) *Predavanje je organizovano*
 lecture.NOM be.AUX.PRS.3sg organize.PAS.PTCP.n.sg
u okviru kursa
 in the framework of course.GEN
„Međunarodni odnosi i spoljna politika”
 international.m.pl.NOM relation.pl.NOM and foreign.GEN policy.NOM
 'The lecture was organized within the course "International Relations and Foreign Policy"'

This pattern is followed by a very similar one, namely [P1 N P2]:

- (56) *Nezaposlenost je znatno manja u odnosu na*
 unemployment.NOM be.PRS.3sg significantly lesser.f.NOM with respect to
prošlu godinu.
 last.f.ACC year.ACC
 'Unemployment is significantly lower compared to last year'
- (57) *Na zahtev države članice, mogu se usvojiti*
 on request state.GEN member.GEN can.PRS.3pl themselves adopt.INF
mere u skladu sa postupkom utvrđenim u
 measures.ACC in accordance with procedure.INS established.INS in
članu 4. (ParCoLab, Evroteka)
 article 4.LOC
 'Measures may be adopted at the request of a Member State in accordance with the procedure laid down in Article 4.' (ParCoLab, JRC)

Nouns are mainly in the singular, and present a wide variety of meanings due to the fact that they come from many different semantic domains.

Table 3 shows the most important indexed schemas of formation of CAs identified in Serbian, of which there are 19 altogether (see the Appendix for a detailed list of patterns), and their representativeness in the corpus.

Table 3: Main patterns of CAs in Modern Serbian (on the basis of grammars, dictionaries and existing literature on CAs; Numbers indicate the number of items found for each pattern, not their frequency).

Pattern	Number of CPs	Examples of CAs
[P ₁ N]	268	<i>iz razloga</i> ‘due to, because of’, <i>iz ugla</i> ‘from the vantage point of’, <i>na čelu</i> ‘at the head of’, <i>na dnu</i> ‘at the bottom of’, <i>na domak</i> ‘within the range of’, <i>na kraju</i> ‘at the end of’, <i>od strane</i> ‘on behalf of’
[P ₁ N P ₂]	60	<i>na čelu sa</i> ‘headed by’, <i>po uzoru na</i> ‘following, based on the model of’, <i>u odnosu na</i> ‘with respect to’, <i>u roku od</i> ‘within, in’, <i>u vezi sa</i> ‘in relation to’, <i>za razliku od</i> ‘unlike’
[Adv P ₂]	24	<i>daleko od</i> ‘far away from’, <i>desno od</i> ‘to the right of’, <i>vezano za</i> ‘related to’, <i>više od</i> ‘more than’, <i>zavisno od</i> ‘depending on’
[P ₁ Adj N]	18	<i>na drugom kraju</i> ‘at the other end of’, <i>na suprotnoj strani</i> ‘on the opposite side of’, <i>s druge strane</i> ‘on the other side of’, <i>s prednje strane</i> ‘on the front side of’, <i>sa leve strane</i> ‘on the left-hand side of’
[V P ₂]	9	<i>oslanjajući se na</i> ‘leaning upon’, <i>počev od</i> ‘from, since’, <i>polazeći od</i> ‘starting from’, <i>sudeći po</i> ‘to judge from’
[N P ₂]	5	<i>obzirom na</i> ‘given’, <i>nalik na</i> ‘like’, <i>korak do</i> ‘one step away from’
other	17	<i>što se tiče</i> ‘as for, concerning’, <i>na dva koraka od</i> ‘two steps away from’
Total	401	

Table 3 clearly suggests that there are four or five really productive patterns; all the others are rather occasional formations resulting from the freezing of different syntactic constituents. The case of the fourth pattern of formation, namely [P₁ Adj N], is very specific because of the presence of adjectives, which often causes a certain degree of variability. For instance, when occurring in this pattern, the noun *strana* ‘side’ forms more than 15 items. Indeed, it would probably be more accurate to say that there is a single CA *s(a) Adj strane* rather than as many CAs as possible combinations with various adjectives (e.g. *s(a) desne strane*, *s(a) leve strane*, *s(a) prednje strane*, *s(a) zadnje strane*, and so on). However, in this paper, I chose the latter option in order to document as fully as possible the system of CAs in Serbian. The recurrence of most of these constructions in corpora further supports this choice, even though a more sophisticated and reliable method must be developed in order to establish the exact status of these constructions (for an attempt in this direction concerning CAs in French, see Stosic and Fagard 2019).

It is worth noting that 17 out of 19 patterns contain at least one simple preposition, and 6 of them include two prepositions. The next subsection provides more details on the prepositions involved in the construction of CAs.

4.2.2 Initial and final prepositions

Twelve out of the nineteen identified patterns of formation of CAs in Serbian include a preposition in the initial position (labeled P_1 , in Table 3), and twelve of them display a preposition in the final position (labeled P_2 , in Table 3) (see the Appendix for the entire list of patterns of formation). It is worth noting, however, that the most productive patterns include an initial preposition, so that 89% of indexed CAs in Serbian contain a P_1 . The initial preposition is most typically *u* ‘in, into’ and *na* ‘on, onto’, which covers more than half of the constructions, and less frequently *pod* ‘under’, *sa* ‘with’, *po* ‘around, on’, *iz* ‘from, of’, *za* ‘for’ and *bez* ‘without’, rarely other adpositions (see Table 4a). As for the final preposition, it appears in only 26% of indexed CAs. This very low score can be accounted for by

Tables 4a & 4b: Patterns of CAs in Serbian: (a) initial and (b) final adposition.

a. Prep ₁	Productivity (number of CAs)	b. Prep ₂	Productivity (number of CAs)
<i>u</i> ‘in, into, at’	131	<i>od</i> ‘from, of’	40
<i>na</i> ‘on, onto’	81	<i>sa</i> ‘with, of’	24
<i>pod</i> ‘under’	42	<i>na</i> ‘on, onto’	20
<i>sa</i> ‘with, of’	24	<i>za</i> ‘for, behind’	12
<i>po</i> ‘around, on’	24	<i>prema</i> ‘towards’	3
<i>iz</i> ‘from, out of’	15	<i>iz</i> ‘from, out of’	2
<i>za</i> ‘for, behind’	11	<i>do</i> ‘up to, untill’	2
<i>bez</i> ‘without’	10	<i>ka</i> ‘towards’, <i>posle</i>	
<i>pri</i> ‘at’	5	‘after’ <i>o</i> ‘about’,	1
<i>uz</i> ‘up, close to, with’	5	<i>po</i> ‘around, on’	
<i>van</i> ‘outside’	2	Total	107
<i>izvan</i> ‘outside’, <i>o</i> ‘about’, <i>od</i> ‘from, of’, <i>preko</i> ‘over, across, by’, <i>prema</i> ‘towards’, <i>usled</i> ‘subsequent to’	1		
Total	356		

the availability of case marking, which – as a systemic feature – can serve to integrate a lower nominal constituent in a larger syntactic structure. This observation is consistent with the fact that in languages without case marking, the presence of P_2 in CAs is much more widespread. In French, for instance, where there is no case marking, up to 93% of CAs display a final preposition (see Stosic forthcoming).¹¹ Returning to Serbian, the most frequent P_2 is the preposition *od* ‘from, of’, followed by less frequent but recurrent ones such as *sa* ‘with, of’, *na* ‘on, onto’ and *za* ‘for’ (see Table 4b).

Tables 4a and 4b display a very small intersection (e.g. *na* and *za*), showing that P_1 and P_2 have clearly different semantic and functional profiles. The same holds for other languages described in this volume.

4.2.3 Preposition-case constructions

Another, more traditional, way of analyzing the corpus of CAs compiled for the present study would be to look directly at preposition-case constructions, whose identification requires considering both the pattern of formation and the prepositions and cases involved in the schema. Popović (1966: 209) argued that there are three main types of preposition-case constructions, which match what is called here patterns of formation (or construction schemas):

- 1) Case + Case (\Rightarrow [N], e.g. *putem* ‘by’, *pomoću* ‘by means of’, *krajem* ‘at the end of’, *tokom* ‘during’)
- 2) Prep + Case + Case (\Rightarrow [P_1 N], e.g. *u toku* ‘during’, *pod kontrolom* ‘under the control of’)
- 3) Prep + Case + Prep + Case (\Rightarrow [P_1 N P_2], e.g. *u odnosu na* ‘with respect to’, *iz straha od* ‘for fear of’)

Popović’s first type is considered here as forming simple non-canonical prepositions resulting from the conversion of nouns into prepositions (see § 3.2. above). The second type corresponds to the most frequent pattern of formation, namely [P_1 N], and it involves 28 preposition-case constructions. The most frequent ones according to Popović (1966: 209) are – translated in my terminology –: [$u + N_{\text{Loc}} + \text{GEN}$], [*pod* + $N_{\text{INS}} + \text{GEN}$], [$u + N_{\text{ACC}} + \text{GEN}$], [$na + N_{\text{Loc}} + \text{GEN}$] and [$na + N_{\text{Loc}} + \text{GEN}$]. Table 5a showing the 7 most frequent constructions partly validates Popović’s

¹¹ Note that the use or not of the case marking in a given language does not impact the widespread use of P_1 in the structure of CAs. That is at least what the comparison between Serbian and French adpositional systems suggests: the two languages display about 88% of CAs with P_1 .

analysis of the second type; the score of all other preposition-case constructions, 21 altogether, is equal to or less than 6. As for the third type, Popović argues that the only significant preposition-case construction is [$u + N_{Loc} + sa$ INST]; according to my data, it does indeed rank first out of 22 constructions of this type indexed. The next most frequent preposition-case construction of the third type, namely [$u + N_{Loc} + od/na$ GEN], can also be considered as significant, even though its frequency is relatively low compared to the first type (see Table 5b). The score of all other preposition-case constructions of the third type never exceeds 4.

Tables 5a & 5b: The most frequent preposition-case constructions with CAs in Serbian: (a) [P₁ N] type and (b) [P₁ N P₂] type.

a. [P ₁ N] type preposition- case constructions	Productivity (number of CAs)	b. [P ₁ N P ₂] type preposition- case constructions	Productivity (number of CAs)
[$u + N_{Loc} + GEN$]	77 (e.g. 58)	[$u + N_{Loc} + sa/za + INST$]	18 (e.g. 65)
[$pod + N_{INST} + GEN$]	42 (e.g. 59)	[$u + N_{Loc} + od/na + GEN$]	11 (e.g. 66)
[$na + N_{Loc} + GEN$]	40 (e.g. 60)		
[$na + N_{Acc} + GEN$]	25 (e.g. 61)		
[$po + N_{Loc} + GEN$]	20 (e.g. 62)		
[$iz + N_{GEN} + GEN$]	13 (e.g. 63)		
[$u + N_{Acc} + GEN$]	13 (e.g. 64)		

- (58) *u okviru* ‘in the framework of’, *u toku* ‘during’, *u cilju* ‘with the aim of’
- (59) *pod uticajem* ‘under the influence of’, *pod pritiskom* ‘under the pressure of’
- (60) *na kraju* ‘at the end of’, *na nivou* ‘at/to the level of’, *na putu* ‘on the road to’
- (61) *na temu* ‘on the topic of’, *na način* ‘in the manner of’, *na račun* ‘at the expense of’
- (62) *po pitanju* ‘in terms of’, *po osnovu* ‘on the basis of’, *po završetku* ‘after’
- (63) *iz oblasti* ‘in the area of’, *iz razloga* ‘due to’, *iz vremena* ‘in the days of’
- (64) *u vreme* ‘in/at the time of’, *u ime* ‘on behalf of’, *u korist* ‘in favor of’

- (65) *u društvu sa* ‘in the company of’, *u poređenju sa* ‘compared to’, *u poteri za* ‘in search of’
- (66) *u odnosu na* ‘in relation to, with respect to’, *u roku od* ‘within, within a period of’, *u zavisnosti od* ‘depending on’

As for the remaining patterns of formation, there are not any preposition-case constructions that are really significant compared to those of [P₁ N] and [P₁ N P₂] types.

It is worth noting the existence of a series of CAs with the same lexical nucleus but belonging to different preposition-case constructions (cf. Popović 1966: 210). Generally speaking, the variation can occur either at the level of the pattern of formation – meaning that one can find in synchrony a more extensive construction and its reduced form –, or within the same pattern of formation. Whatever the type of variation observed, such processes lead to the formation of semantically more or less close pairs. More precisely, the variation can come from:

- i. the final preposition, as in *u vezi sa / u vezi* ‘in relation to’ (cf. 67a vs b) and *u društvu sa / u društvu* ‘in the company of’ (cf. 68a vs b), where the P₂ disappears. Consequently, the original [u + N_{Loc} + sa + INS] preposition-case construction (cf. 67a and 68a) shifts to the most productive [u + N_{Loc} + GEN] construction (cf. 67b and 68b). This also involves a shift from the [P₁ N P₂] to the [P₁ N] pattern of formation. The same shift is observed in the following doublets: *po uzoru na / po uzoru* ‘following, based on the model of’, *na visini od* ‘to the height of, at the altitude of’ / *na visini* ‘level with’, *u visini od* ‘of up to’ / *u visini* ‘level with’, *na izlazu iz* ‘at the exit of, upon leaving’ / *na izlazu* ‘at/on the output of’, *u osi sa / u osi* ‘in line with’, *bez potrebe za / bez potrebe* ‘without the need for’, *u ravni sa* ‘at/to the level of’ / *u ravni* ‘level with’, *u skladu sa / u skladu* ‘in accordance with’. In some rare cases, the final preposition does not disappear but alternates with another simple preposition as in the pairs *na putu ka* ‘on the path towards’ / *na putu za* ‘on the path to’, *računajući na* ‘relying on’ / *računajući od* ‘starting from, as from’;

- (67) a. *Najčešće postavljena pitanja u vezi sa*
 the most often asked.f.pl.NOM question.pl.NOM in relation to
primenom Opšte uredbe o zaštiti
 application.INS general.GEN regulation.GEN about protection.LOC
podataka. (<https://www.poverenik.rs>)
 data.pl.GEN
 ‘The most frequently asked questions in relation to the application of
 the General Data Protection Regulation (GDPR)’

- b. *Mišljenja* **u vezi** *primene* *propisa*
 opinion.pl.NOM in relation (to) application.GEN regulations.GEN
u oblasti zaštite konkurencije. (<http://www.kzk.gov.rs/>)
 in the field of protection.GEN competition.GEN
 ‘Opinions about the application of regulations in the field of the
 protection of competition’
- (68) a. *General Lazarević* **u društvu sa** *medijima* (www.pikanal.rs)
 general. Lazarević. in the company media.pl.INS
 NOM NOM of
 ‘General Lazarević accompanied by media’
- b. *Jelena Gavrilović* **u društvu** *svog dečka*
 Jelena.NOM Gavrilović.NOM in the company of her.GEN boyfriend.GEN
uživa u Grčkoj. (<https://nadlanu.com>)
 enjoy.PST.3sg in Greece.LOC
 ‘Jelena Gavrilović together with her boyfriend is enjoying herself in
 Greece’
- ii. the initial preposition, that can either disappear as in *s obzirom na / obzirom na* ‘given’ (cf. 69a vs b) and *uz pomoć / pomoću* ‘with the help of, by means of’ (cf. 72a vs c) – leading to the change in the pattern of formation for example in the first case from [P₁ N P₂] to [N P₂], or alternate with another simple preposition as in the pairs: *iz straha od* ‘due to the fear of’ / *u strahu od* ‘for fear of’ (cf. 70a vs b), *na uglu* ‘at the corner of’ / *iz ugla* ‘from the vantage point of’ (cf. 71a vs b), and *uz pomoć / s pomoću* ‘with the aid/help of’ (cf. 72a vs b), where the pattern of formation is not affected at all (see Šarić 1999: 87–93 for an in-depth analysis of these items). The preposition-case construction is necessarily transformed in both situations. For instance, while *iz straha od* and *u strahu od* are formed using the same pattern, namely [P₁ N P₂], they come under two different preposition-case constructions, respectively [*iz* + N_{Gen} + *od* + GEN] and [*u* + N_{Loc} + *od* + GEN]. The alternation of the initial preposition is much wider than the alternation of the final preposition as shown by the following pairs: *na manje od* ‘less than’ / *za manje od* ‘for under, for less than’, or *iz aspekta / s aspekta* ‘from the perspective of’, *sa ciljem / u cilju* ‘with the aim of’, *iz doba* ‘from the time of’ / *u doba* ‘at the time of’, *na dnu* ‘on the bottom of’ / *u dnu* ‘at the bottom of, in the back of’ / *pri dnu* ‘near the bottom of’, *na ime* ‘in the name of’ / *u ime* ‘on behalf of’, *pod nadzor / uz nadzor* ‘under the supervision of’, *iz nedostatka* ‘due to the lack of’ / *u nedostatku* ‘for lack of’, *na nivou / u nivou* ‘at/to the level of’, *u oblasti*

/ *iz oblasti* ‘in the field of’, *po osnovu* ‘on the basis of’ / *na osnovu* ‘based on’, *na početku* / *u početku* ‘at the beginning of’, *na podnožju* / *u podnožju* ‘at the foot of’, *na području* / *u području* ‘in the area of’, *na primeru* ‘(based) on the example of’ / *po primeru* ‘following the example of, on the model of’, *u prisustvu* ‘in the presence of’ / *bez prisustva* ‘without the presence of’, *na putu* ‘on the road to’ / *preko puta* ‘opposite, over, across’, *na račun* ‘at the expense of’ / *za račun* ‘for the account of’, *prema računici* / *po računici* ‘according to the calculations of’, *u slučaju* / *za slučaj* ‘in (the) case of’, *na sred* ‘in the middle of’ / *po sred* ‘by the middle of’ / *u sred* ‘in the middle of’ / *iz sred* ‘from the middle of’. In some cases, the alternation of the initial preposition triggers a change in the form of the lexical nucleus as well, as seen in *iz straha od* ‘due to the fear of’ / *u strahu od* ‘for fear of’ (cf. 70a vs b) where the noun *strah* ‘fear’ is used respectively in the Genitive and the Locative case, or in *na uglu* ‘at the corner of’ / *iz ugla* ‘from the vantage point of’ (cf. 71a vs b) where the noun *ugao* ‘corner’ appears respectively in the Locative and the Genitive;

- (69) a. **S obzirom na** *veliko interesovanje za stanove*
 with regard to big.n.ACC interest.ACC for apartments.ACC
u projektu Beograd na vodi, važno je
 in project.LOC Belgrade.NOM on water.LOC importantly be.PRS.3sg
što pre završiti proces
 as soon as possible finish.INF process.ACC
eksprijacije... (www.nspm.rs)
 expropriation.GEN
 ‘Given the high interest in apartments within the Belgrade Waterfront project, it is important to finish the process of expropriation as soon as possible’
- b. **Obzirom na** *veliko interesovanje umesto jednog*
 regard to big.n.ACC interest.ACC instead of one.m.GEN
podelićemo 3 prstena. (https://insta01pho.com)
 distribute.FUT.1pl 3 ring.pl.GEN
 ‘Given the high interest, instead of one, we shall distribute 3 rings’
- (70) a. *Čovek je spreman na sve iz straha od*
 man.NOM be.PRS.3sg ready.m.NOM on all.ACC from fear of
smrti.
 death.GEN
 ‘Humans are ready for anything due to the fear of death’

- b. *Postoje ljudi koji pribegnu religiji*
 exist.PRS.3pl man.pl.NOM who.NOM resort.PRS.3pl religion.DAT
u strahu od smrti.
 in fear of death.GEN
 ‘There are people who resort to religion for fear of death’
- (71) a. *Dodite u bar na uglu Francuske i*
 come.IMPER.2pl in(to) bar.ACC on corner French.f.GEN and
Dušanove ulice.
 Dušan’s.f.GEN street.GEN
 ‘Come into the bar at the corner of French and Dušan Streets’
- b. *Pustošenje zlatiborske šume iz ugla*
 devastation.NOM Zlatibor’s forest.GEN from corner
novinara b92. (<http://zlatiborskasuma.blogspot.com/>)
 journalist.GEN of b92
 ‘Devastation of the forest of Zlatibor from the b92 journalist’s vantage point’
- (72) a. *Uz pomoć aplikacije trebalo bi da*
 with help application.GEN shall.PAST.PART.n.sg be.COND.3sg that
bude moguće zakazati pregled kod izabranog
 be.PRS.3sg possible schedule.INF consultation.ACC at chosen.GEN
lekara. (<https://www.bbc.com/serbian/lat/srbija-44395059>)
 doctor.GEN
 ‘With the help of this application, it should be possible to arrange a consultation with a chosen doctor’
- b. *Ako želite nadograditi softver uređaja*
 if want.PRS.2pl update.INF software.ACC device.GEN
s pomoću aplikacije Garmin Connect, morate imati
 with help application.GEN Garmin Connect need.PRS.2pl have.INF
Garmin Connect račun. (<https://www8.garmin.com>)
 Garmin Connect account.ACC
 ‘If you want to update the software of your device with the help of the Garmin Connect application, you must have a Garmin Connect account’
- c. *Pomoću aplikacije MiBT možete da resetujete*
 with help application.GEN MiBT can.PRS.2pl that restart.PRS.2pl
ruter pomoću svog mobilnog
 router.ACC with help itself.GEN mobile.m.GEN

telefona. (<https://www.wbsdigital.com/blog>)

phone.GEN

'By means of the MiBT application, you can reboot your router using your mobile phone'

- iii. the lexical nucleus, which occurs the most frequently when some simple prepositions displaying double case government, such as *u* (ACC/LOC), *na* (ACC/LOC), or *pod* (ACC/INS) appear in the initial position of the construction. One can thus find many pairs with the preposition *u* governing the same noun either in the Accusative or the Locative case (e.g. *u znak* 'as a sign of' / *u znaku* 'in the form of', *u nadležnost* 'into the jurisdiction of' / *u nadležnosti* 'in the jurisdiction of', *u ruke* 'into the hands of' / *u rukama* 'in the hand(s) of', *u poteru za* 'into pursuit of' / *u poteri za* 'in pursuit of'), with the preposition *na* displaying the same governing alternation (e.g. *na domak* / *na domaku* 'within (the) range/reach of', *na kraj* / *na kraju* 'at the end of', *na marginu* / *na margini* 'at the margin of'), or with *pod* governing the same noun either in the Accusative or the Instrumental case (e.g. *pod okrilje* / *pod okriljem* 'under cover of', *pod zaštitu* / *pod zaštitom* 'under the protection of'). The main function of this alternation is to distinguish between atelic (Locative and Instrumental) and telic (Accusative) readings of the PPs in question. The case marker on the noun forming the CA is directly dependent on the telic/atelic nature of the governing predicate (see 73a as opposed to 73b):

- (73) a. *Stigli smo na kraj ulice.*
 arrive.PAST.PART.m.pl be.AUX.PRS.1pl on end.ACC street.GEN
 'We reached (arrived at) the end of the street'
- b. *Čekali smo na kraju ulice.*
 wait for.PAST.PART.m.pl be.AUX.PRS.1pl on end.LOC street.GEN
 'We were waiting at the end of the street'

In some cases, the alternation observed in the nominal nucleus is due to the variation in number as in the pairs: *na margini* 'at the margin of' / *na marginama* 'at/on the margins of', *u okviru* 'in the framework/scope of' / *u okvirima* 'within the scopes of', *pod uslovom* 'under condition of' / *pod uslovima* 'under conditions of':

- (74) a. *Proizvođači se nadaju rodu u okviru višegodišnjeg proseka.* (<http://www.kikinda.org.rs>)
 producer.pl.NOM themselves hope.PRS.3pl crop.DAT in scope
 multi-year.m.GEN average.GEN
 'Producers are hoping for a crop in the range of the multi-year average'

- b. *U Srbiji se očekuju temperature*
 in Serbia.LOC themselves expect.PRS.3pl temperature.pl.NOM
u okvirima višegodišnjeg proseka. (<http://meteoplaneta.rs/>)
 in scopes multi-year.GEN average.GEN
 ‘In Serbia one expects temperatures within the range of the multi-year average’

- iv. the modifier of the lexical nucleus, when such an element appears in the construction as in the case of the pattern [P₁ Adj N]. Serbian thus forms a series of CAs on the basis of the noun *strana* ‘side’, which display a high degree of syntagmatic variability due to the possibility of combining the nucleus with a wide range of adjectives: *na levoj / desnoj / prednjoj / suprotnoj strani* ‘on / at the left / right / front / opposite side of’, *s(a) druge / one / desne / leve / zadnje strane* ‘on the other / this / right / left / back side of’, and so on. CAs formed on the basis of the nouns *deo* ‘part’ and *kraj* ‘end’ also allow several adjectives, which enables a certain variation in the construction schema: *u prednjem / zadnjem / drugom delu* ‘in the forward / backward / other part of’. As already mentioned, it is a debatable matter whether there are as many CAs as possible combinations, or just a single one (e.g. *na Adj strani*) (cf. above). Moreover, given the degree of variation displayed in the latter examples, their class membership is highly dubious (see § 4.2.1 above).

The two (or three) forms can be either semantically very close to each other (cf. 67a–b, or 69a–b), or relatively far from each other (cf. 71a vs b). If two forms of a pair are similar in meaning, their distribution is identical, they are generally concurrent and reflect two different stages of the process of lexicalization and/or grammaticalization (cf. Lehmann 2002b, Hoffmann 2005, Brinton & Traugott 2005, Fagard and De Mulder 2007 for fuller discussion on the latter issue). In many such cases, the shorter form is the more recent one, which testifies to the general tendency of languages to reduce complex forms to simpler ones (cf. *s obzirom na / obzirom na* ‘given’, *u vezi sa / u vezi* ‘in relation to’, *u ravni sa* ‘at/to the level of’ / *u ravni* ‘level with’, and so on).¹² If there is a significant difference between the two forms of a given pair, it could be due to several linguistic and non-linguistic factors. In some cases, one of these forms, due to its higher degree of lexicalization/grammaticalization, matches the category of prepositions better

¹² It is worth noting that some of these shorter forms are prohibited by the standard, which does not mean that they are prohibited by the system. For instance, in srWaC, one can find 29,437 occurrences of the short form *u vezi* as opposed to 45,391 occurrences of *u vezi sa*.

than the other, while in some others, both forms are equally good candidates for the class.

4.3 Frequency of use: Towards a quantitative approach to CAs in Serbian

In this section, I present a summary of the most important results of a quantitative corpus-based study of sequences that I consider here as complex adpositions. The main idea behind the recourse to a quantitative approach was that, due to their structuring role in the sentence, most of the candidate expressions that were inventoried as CAs in Serbian should occur in language use with a significant if not (very) high frequency. If so, this would support the claim that they are durably stored in the speakers' lexical memory as whole indivisible units,¹³ as simple adpositions are, and are not built by syntax during on-line language production (cf. Hoffmann 2005). Note however that, in line with Hoffmann's (2005) study on CAs in English, I consider that frequency cannot be taken as a decisive criterion for differentiating between sequences belonging to the class of adpositions and those that do not deserve this status (see also Bybee 2010, Bybee and Hopper 2001). As shown by Hoffmann (2005), some low frequency expressions are undoubtedly CAs, whereas, on the other hand, there are many sequences formed following the most productive CA construction schemas (e.g. P N (P)) with a very high frequency but whose class membership is out of the question.¹⁴ The frequency of use is thus taken here as (simply) offering an additional clue for analyzing the status of candidate sequences as CAs or not, but it is obviously interrelated with many other morphological, syntactic, semantic and evolutionary parameters that must be taken into account in such an undertaking (see Stosic and Fagard 2019 for an attempt to develop a multivariable approach to CAs in French).

¹³ I shall not tackle the issue of whether these units are grammatical or lexical in nature (for a detailed discussion of this problem, see among others Lehmann 2002b, Hoffmann 2005, Fagard and De Mulder 2007).

¹⁴ For instance, in the srWaC corpus, there are 11,481 occurrences (20.70 pmw) of the free construction *za borbu protiv* 'for the fight against', and 7,436 occurrences of the free expression *na konferenciji za* 'at the conference for' (13.41 pmw), which is a very high frequency for [P₁ N P₂] items: they are respectively the 14th and the 20th most frequent expressions with this structure. In contrast, in the same corpus, there are only 841 (1.76 pmw) and 367 (0.77 pmw) occurrences respectively of the expressions *u strahu od* 'for fear of' and *u društvu sa* 'in the company of' which qualify as CAs.

For the purposes of this study, a pilot quantitative exploration of srWaC, a 477-million-word Serbian corpus available in Sketch Engine, was conducted¹⁵ aiming to establish the relative frequency of occurrence for both the simple and complex adpositions of my corpus. It is interesting to measure the frequency of simple adpositions because it leads to a better understanding of the way complex adpositions are positioned with respect to them, and consequently of how they function in the language system.

The first interesting observation emerges by comparing the average frequencies of the two sub-groups of adpositions. The average frequency of simple adpositions ranges up to 737 occurrences pmw, whereas that of complex adpositions is about 14 occurrences pmw, i.e. 52-fold lower. Next, the frequency of simple adpositions in Serbian is far from homogeneous, which is consistent with what has been found for several other languages (see Fagard forthcoming, Fagard et al. this volume, for Modern French). Accordingly, on the one hand, there are in Serbian a few simple adpositions that are very frequent in use: *u* ‘in, into, at’, *na* ‘on, onto, at’ and *za* ‘for’. Their frequency ranges from 11‰ (i.e. per thousand words) to 27‰ in srWac (respectively 27‰, 15‰ and 11‰), which clearly suggests that they are highly grammaticalized, and that they typically have a functional role in discourse production. On the other hand, the great majority of the remaining simple adpositions display a significantly lower frequency, less than 1‰ (i.e. less than a thousand occurrences per million words). It is worth noting that the frequencies of only seven other simple prepositions range from 1‰ to 10‰, as shown in the second column of Table 6.¹⁶

Another finding worth mentioning here has to do with the relationship between the frequency of simple adpositions and their morphological type. More precisely, different behaviors are displayed depending on whether they belong to canonical, converted or compound prepositions (see § 3.2), as suggested by Graph 2. It clearly shows first that canonical simple prepositions are the most frequent ones, second that converted prepositions are generally less frequent than the latter but more frequent than compounds, and third that compound prepositions display the lowest frequencies. This contrast thus strongly supports the validity of this tri-partitioning of simple prepositions in Serbian.

15 I would like to thank Benjamin Fagard for his valuable help in establishing the relative frequencies for CAs from my corpus.

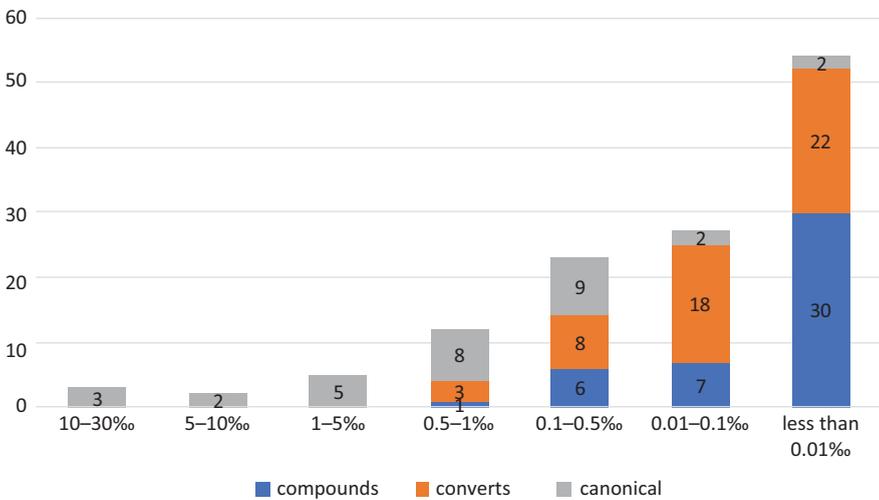
16 Six items were excluded from the quantitative analysis because of their polyfunctionality: the expressions *mesto* ‘place’ / ‘instead of’, *put* ‘road’ / ‘toward’ and *vrh* ‘top’ / ‘atop’ can function as either nouns or prepositions, *koristeći se* ‘using, by means of’ and *plašeci se* ‘for fear of’ as impersonal verb forms and prepositions, and *više* ‘more’ / ‘above’ as adverb and preposition. Given their high frequency of occurrence in the corpus and the lack of reliable NLP tools for automatic POS tagging, isolating prepositional uses from other uses was not possible.

Table 6: Relative frequency of simple and complex adpositions in Serbian (srWaC, 20th c., roughly 477 million words). Examples are ordered from more to less frequent.

Frequency range (%)	Simple adpositions	Complex adpositions
between 10 and 30	(3) <i>u</i> 'in, into, at', <i>na</i> 'on, onto, at', <i>za</i> 'for'	
between 5 and 10	(2) <i>od</i> 'of, from', <i>sa</i> 'with, of, out'	
between 1 and 5	(5) <i>o</i> 'about', <i>iz</i> 'from', <i>do</i> 'until', <i>po</i> 'around', <i>zbog</i> 'because of'	
between 0.5 and 1	(12) <i>oko</i> 'around', <i>bez</i> 'without', <i>kod</i> 'at', <i>pre</i> 'before', <i>prema</i> 'towards', <i>uz</i> 'up, close to, with', <i>posle</i> 'after', <i>preko</i> 'over, across, by', <i>pod</i> 'under', <i>protiv</i> 'against', <i>kroz</i> 'through', <i>između</i> 'between'	
between 0.1 and 0.5	(23) <i>nakon</i> 'after', <i>tokom</i> 'during', <i>pred</i> 'in front of', <i>pored</i> 'next to', <i>među</i> 'among', <i>pri</i> 'at', <i>ka</i> 'towards', <i>prilikom</i> 'on the occasion of', <i>osim</i> 'except', <i>nad</i> 'above, overhead', <i>iza</i> 'behind', <i>kraj</i> 'next to', <i>van</i> 'outside', <i>niz</i> 'down'	(12) <i>više od</i> 'more than', <i>u okviru</i> 'in the framework/scope of', <i>od strane</i> 'on behalf of, from', <i>u odnosu na</i> 'with respect to', <i>na kraju</i> 'at the end of', <i>na osnovu</i> 'on the basis of', <i>u skladu s(a)</i> 'in accordance with', <i>u toku</i> 'during'
between 0.01 and 0.1	(27) <i>početkom</i> 'at the beginning of', <i>krajem</i> 'at/by the end of', <i>blizu</i> 'near, close to', <i>slično</i> 'similarly to', <i>pomoću</i> 'by means of', <i>usled</i> 'subsequent to', <i>sem</i> 'except', <i>izvan</i> 'outside', <i>uoči</i> 'on the eve of, on the brink of', <i>zarad</i> 'for the sake of, for', <i>duž</i> 'along', <i>mimo</i> 'by, next to', <i>naspram</i> 'opposite', <i>nasuprot</i> 'opposite'	(97) <i>u vezi s(a)</i> 'in relation to', <i>u slučaju</i> 'in (the) case of', <i>u vreme</i> 'at the time of', <i>za vreme</i> 'during', <i>bez obzira na</i> 'without regard to', <i>u cilju</i> 'with the aim of', <i>uz pomoć</i> 'with the help/aid of', <i>u vezi</i> 'in relation to', <i>u periodu</i> 'at the time of', <i>iz oblasti</i> 'in the field of', <i>na nivou</i> 'at/to the level of', <i>u pogledu</i> 'in respect of, about', <i>na putu</i> 'on the road to', <i>u centru</i> 'in/at the center of', <i>u smislu</i> 'in terms of, within the meaning of'

Table 6 (continued)

Frequency range (‰)	Simple adpositions	Complex adpositions
less than 0.01	(54) <i>polovinom</i> 'in the middle of', <i>spram</i> 'opposite', <i>protivno</i> 'contrary to', <i>povrh</i> 'above, in addition to', <i>nasred</i> 'in the middle of', <i>pokraj</i> 'next to', <i>naočigled</i> 'in front of', <i>posred</i> 'by the middle of'	(291) <i>u čast</i> 'in honour of', <i>u znaku</i> 'in the form of', <i>u sred</i> 'in the middle of', <i>u visini od</i> 'of up to', <i>u svetlu</i> 'in the light of', <i>iz ugla</i> 'from the vantage point of', <i>u zamenu za</i> 'in exchange for', <i>o trošku</i> 'at the expense of', <i>pod nadzorom</i> 'under the eye of, under the supervision of', <i>za ljubav</i> 'for (the) love of, for', <i>na dohvat</i> 'within range of', <i>na rubu</i> 'on the verge of', <i>u ritmu</i> 'at the pace of', <i>na uštrb</i> 'to the detriment of', <i>u nivou</i> 'at/to the level of', <i>u dnu</i> 'at the bottom of, in the back of'

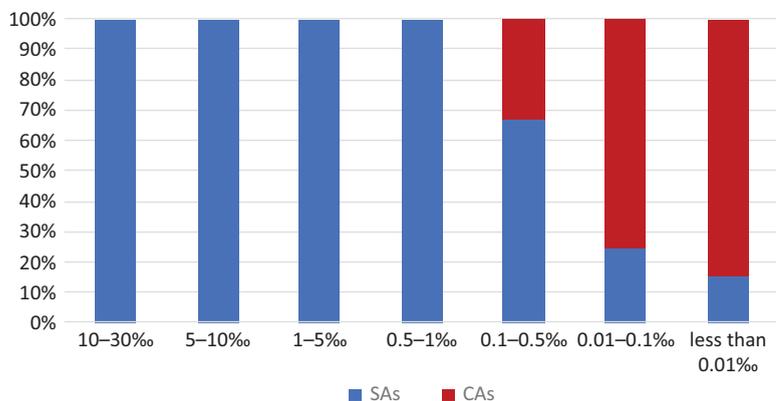


Graph 2: Distribution of SAs in Serbian according to their morphological type and frequency of use in srWac (ptw).

These differences in the behavior of the three types of SAs are further supported by their average frequency of occurrence, which ranges from 2890.87 occurrences pmw for canonical SAs to 92.26 occurrences pmw for converts, and to 40.72 occurrences pmw for compound CAs. This cline in frequency is to be seen in connection

with the more or less functional (vs lexical) nature of the adpositions in question: canonical simple prepositions thus obviously display a highly functional-like behavior, which highlights their fundamentally structuring role in the language system.

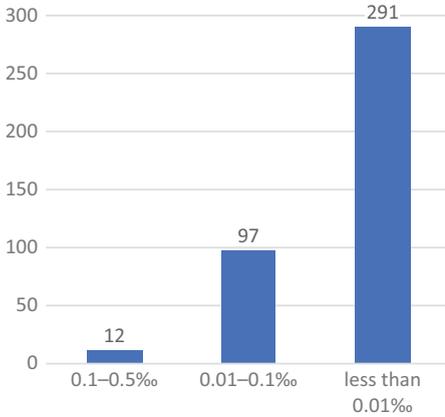
As for comparing the relative frequencies of simple and complex prepositions, I already mentioned that, as expected, the latter have a lower frequency of use in corpora due to their substantial lexical load. Both Table 6 and Graph 3 show that the first four high-frequency slots (ranging from 0.5‰ to 30‰) are exclusively occupied by simple prepositions. The level ranging from 0.1‰ to 0.5‰ is the first to be split, but it is nevertheless largely dominated by simple prepositions. The last two bars, with the lowest rates of frequencies (up to 0.1‰), are predominantly filled by CAs. Table 6 provides examples of CAs for each level of frequency, ordered from the more frequent ones to the less frequent ones.



Graph 3: Division of frequency scale between SAs and CAs in Serbian (ptw).

Compared to SAs, the variation in the frequency levels of CAs is more restricted. My pilot corpus study clearly suggests however that their frequency is much lower than that of SAs, as shown by Graph 4. The largest number of CAs (about 73% of them) thus occur less than once in a thousand words, as opposed to SAs, only 40% of which display such low frequencies.

The following table shows the top 20 CAs in Serbian ranked by their frequency of occurrence per million words in srWaC. This result is necessarily corpus dependent, but as the srWaC corpus was compiled from web pages, it is assumed to cover a large variety of lexical domains, so that this finding can be considered as relevant. It thus provides a first insight into the frequency rate that CAs can reach in Serbian.



Graph 4: Frequency levels of CAs in Serbian (ptw).

Table 7: Top 20 CAs in Serbian (based on their relative frequency in srWaC, 20th c., roughly 477 million words).

Complex preposition	Number of occurrences (pmw)
<i>više od</i> 'more than'	299.5
<i>u okviru</i> 'in the framework/scope of'	198.7
<i>od strane</i> 'on behalf of, by'	197.9
<i>u odnosu na</i> 'with respect to'	192.2
<i>na kraju</i> 'at the end of'	183.3
<i>na osnovu</i> 'on the basis of'	176.0
<i>u skladu s(a)</i> 'in accordance with'	161.3
<i>u toku</i> 'during'	138.0
<i>zajedno sa</i> 'together with'	125.6
<i>u vidu</i> 'in the form of'	113.5
<i>u oblasti</i> 'in the field of'	102.2
<i>u vezi s(a)</i> 'in relation to'	95.2
<i>u slučaju</i> 'in case of'	88.3
<i>u vreme</i> 'at the time of'	78.7
<i>za vreme</i> 'during'	74.9
<i>bez obzira na</i> 'without regard to'	69.5
<i>u cilju</i> 'with the aim of'	66.0
<i>uz pomoć</i> 'with the aid of'	64.1
<i>u vezi</i> 'in relation to'	61.7
<i>u periodu</i> 'at the time of'	59.6

Three patterns of formation are present in this top 20 CAs list: the schema [P₁ N] forms 14 of them, 4 of them follow the pattern [P₁ N P₂], and 2 of them are based on the schema [Adv P₂]. Considering the cumulative frequency for each pattern of formation makes it possible to assess their respective extent in language production. One can expect that the more CAs a given pattern has produced, the more recurrent it will be in the corpus. Overall, that is what results from a quantitative exploration of CAs in srWaC, as seen in Table 8. The most productive pattern, namely [P₁ N] (see Table 5a), thus appears as the most widespread, which is expected given that there are about 290 CAs formed on the basis of this construction schema, some of which are very frequent. The correlation holds true for the first five patterns of formation, but not for the others that are much less productive, which makes them almost directly dependent on the frequency rate of the few CAs that they form. This is what is observed about the sixth pattern [Pron V] that forms only one CA, *što se tiče* ‘as for, concerning’, but whose frequency of occurrence in the corpus is significantly high (30.52 pmw).

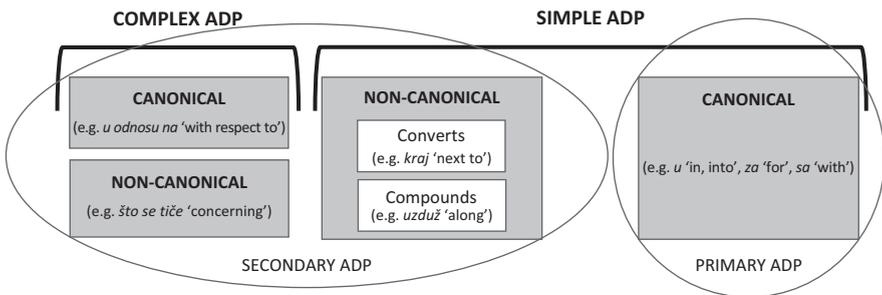
Table 8: Cumulative frequency of the 10 main patterns of formation of CAs in Serbian in srWaC (pmw).

Pattern of formation	Cumulative frequency (pmw)
[P ₁ N]	3916.22
[P ₁ N P ₂]	871.36
[Adv P ₂]	573.49
[P ₁ Adj N]	64.82
[V P ₂]	41.60
[Pron V]	30.52
[N P ₂]	22.50
[V P N]	21.33
[P ₁ Adv P ₂]	8.32
[P ₁ N i N]	1.70

This pilot quantitative analysis of CAs in Serbian thus clearly supports the claim that canonical CAs are those formed by an initial preposition and noun, introducing a NP in the Genitive case. Typically, the initial preposition corresponds to *u* ‘in, into’ or *na* ‘on, onto, at’. Consequently, canonical CAs in Serbian are mainly morphologically complex but syntactically and semantically indivisible units performing a double relation marking, ensured by a preposition in initial position and by cases, or by prepositions and cases, on the right. All CAs in Serbian contain a lexical nucleus, which in a large majority of cases is a noun.

5 Conclusion

This chapter offers an in-depth exploration of the system of CAs in Serbian, as representing the Slavic sub-group of languages. First of all, it shows that complex adpositions, which are prepositions, can be considered as well-established in Serbian – about 400 items were collected for the purposes of this study. Their number is fairly impressive given that it is generally claimed in Serbian linguistics that their emergence is to be related to the overall development and intellectualization of Serbian culture – including sciences, philosophy, administration and journalism –, in the 18th and 19th centuries during the Enlightenment. Second, as in several other European languages, CAs in Serbian are functional equivalents of simple adpositions, which means that they display the same distributional properties as simple prepositions. While playing a very important role in the grammatical system, especially in (nominal) relation marking, as simple prepositions do, CAs provide the lexical grounds for expressing a much wider range of semantic relations between the constituents that they link together than simple prepositions and cases could do, making it possible to encode semantic relations between constituents in a more precise and finer-grained way. As such, they are all to be considered as secondary prepositions. Schema 1 summarizes the Serbian prepositional system as a whole:



Schema 1: General architecture of the Serbian adpositional system.

Third, CAs in Present-day Serbian are predominantly of nominal origin, meaning that their main pattern of formation involves a nominal nucleus preceded by a simple preposition in the initial position [P₁ N] (e.g. *u okviru* ‘in the framework/scope of’). It is also shown that there are about 20 different patterns of formation of CAs, of which only one is highly productive, and three others that can be considered as relatively productive. Whatever their schema of construction, all CAs maintain privileged relationships with case marking on lower NP constituents, with the

Genitive being the most frequently selected case by CAs in Serbian. Fourth, it is also quite clear from the pilot quantitative analysis that the class of CAs is to be thought of as a continuum between sequences that are formally coalesced and semantically rather opaque and those that exhibit a certain degree of transparency, without being assembled at the syntactic level in the very process of on-line language production. Consequently, while displaying variable degrees of fossilization, all the 400 expressions providing the empirical basis for this study are considered as units stored in the speakers' memory as whole entities, deserving the status of CAs.

Needless to say, this study calls for several extensions. From a synchronic point of view, several semantic issues should be addressed, including i) the semantics of the lexical nucleus with a particular emphasis on the semantic types of nouns involved in the formation of CAs, as well as ii) the range of meanings conveyed by CAs. Next, the boundaries of the class should be further investigated by searching for a set of relevant criteria likely to improve the idea of a continuum for CAs. Such a set of criteria should also be applied in a systematic way to a whole range of candidate sequences with a view to assessing their class membership more accurately. Additionally, in order to give a more accurate picture of CAs in Slavic languages, the study must take data from other languages of the subgroup into account. From a diachronic point of view, there is substantial work to be done in order to describe the emergence and evolution of CAs in Serbian, and more generally in Slavic languages. While the need to accompany societal evolution was probably the mechanism that triggered the emergence of CAs in Serbian, the real factors of their rapid rise are to be found in the language system itself. Nevertheless, giving an accurate image of such mechanisms requires addressing a whole range of principles of language change as described in various theories of grammaticalization, lexicalization, and constructionalization.

Abbreviations

1	first person	C	complex
2	second person	COND	conditional
3	third person	DAT	dative
A	adposition	f	feminine
ACC	accusative	FUT	future
ADJ	adjective	GEN	genitive
ADV	adverb(ial)	IMP	imperative
ART	article	INF	infinitive
AUX	auxiliary	INS	instrumental

LOC	locative	PASS	passive
m	masculine	pl	plural
n	neuter	PRS	present
N	noun	PST	past
NEG	negation, negative	PTCP	participle
NOM	nominative	sg	singular
P	preposition	SUP	superlative

References

- Adler, Silvia 2001. Les locutions prépositives : questions de méthodologie et de définition, *Travaux de linguistique* 42–43(1), 157–170.
- Belić, Aleksandar [1959] 1998. *Opšta lingvistika. O jezičkoj prirodi i jezičkom razvitku* [General linguistics. On language nature and language development]. In *Izabrana dela Aleksandra Belića* [Selected writings of Aleksandar Belić], t. I–II. Beograd: Zavod za udžbenike i nastavna sredstva.
- Blagus Bartolec, Goranka & Ivana Matas Ivanković 2013. *Ispred, iza, u i oko* – gdje sve mogu doći prijedlozi [*Ispred, iza, u i oko* – where prepositions can appear]. *Rasprave: Časopis Instituta za hrvatski jezik i jezikoslovlje* 39/2, 361–376.
- Borillo, Andrée 1997. Aide à l'identification des prépositions complexes de temps et de lieu, *Faits de langue* 9, 173–184.
- Borillo, Andrée 2002. Il y a prépositions et prépositions, *Travaux de Linguistique* 42–43, 141–155.
- Brinton, Lauren & Elizabeth Closs Traugott 2005, *Lexicalization and language change*. Cambridge: Cambridge University Press.
- Bučar, Mirjana 2009. Gramatikalizacija i sekundarni prijedlozi glagolskoga i imeničkoga porijekla [Grammaticalization and secondary prepositions of verbal and nominal origin]. *Jezikoslovlje* 10(2): 183–216.
- Bybee, Joan L. & Paul Hopper (eds.). 2001. *Frequency and the emergence of linguistic structure*. Amsterdam & Philadelphia: John Benjamins.
- Bybee, Joan 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Čermák, František 2007. Grammatical idioms. *Linguistica Pragensia* 2, 75–90.
- Cifuentes Honrubia, José Luís 2003. *Locuciones prepositivas – sobre la gramaticalización preposicional en español*. Alicante: Universidad de Alicante.
- Comrie, Bernard and Garrett Corbett (eds) 1993. *The Slavonic languages*. London and New York: Routledge.
- Comrie, Bernard & Barbara Grimes 2003. Slavic Languages. In Frawley, William (ed.), *International Encyclopedia of Linguistics*. Oxford: Oxford University Press. (www.oxfordreference.com)
- Dragičević, Rajna 2012. O imenicama u službi predloga [On nouns in prepositional function]. *Južnoslovenski filolog* LXVIII, 91–111.
- Duškova, Libuše 1994. *Mluvnice současné angličtiny na pozadí češtiny* [A grammar of contemporary English against the background of Czech]. Prague: Academia Praha.
- Fagard, Benjamin forthcoming. Prépositions et locutions prépositives. In Marchello-Nizia, Christiane, Bernard Combettes, Sophie Prévost & Tobias Scheer (eds.), *Grande grammaire historique du français*. Berlin & New York: De Gruyter.

- Fagard, Benjamin & Walter De Mulder 2007. La formation des prépositions complexes: grammaticalisation ou lexicalisation?, *Langue française* 156, 9–29.
- Fleischer, Jürg 2002. Preposition Stranding in German Dialects. In Sjeff Barbiers, Leonie Cornips & Susanne van der Kleij (eds.), *Syntactic Microvariation*, 116–151. Amsterdam: Meertens Instituut.
- Gavrilenko, Viktorija 2013. Semantiko-sintaksičeskie osobennosti proizvodnyx predlogov ‘v forme’ i ‘v vide’ [Semantic and syntactic features of derivative prepositions ‘v forme’ and ‘v vide’], *Filologičeskie nauki* 3 (21/1), 37–40.
- Glušac, Maja & Vlasta Rišner 2016. O uporabi i značenju sekundarnih prijedloga nastalih gramatikalizacijom imenica [On the use and meaning of secondary prepositions formed by grammaticalization of nouns]. *Rasprave* 42(2), 409–442.
- Heine, Bernd (1989). Adpositions in African languages. *Linguistique africaine* 2, 77–127.
- Herrity, Peter 2000. *Slovene: A Comprehensive Grammar*. London & New York: Routledge.
- Hill, Steven 2012. *The N-Factor and Russian Prepositions. Their Development in 11th–20th Century Texts*. Berlin: De Gruyter.
- Hoffmann, Sebastian. 2005. *Grammaticalization and English Complex Prepositions. A corpus-based analysis*. London & New York: Routledge.
- Hopper, Paul & Elizabeth Closs Traugott 2003. *Grammaticalization*. Cambridge: Cambridge University Press.
- Hrubý, Věnceslav 1909. *Predložky slovanské* [Slavic prepositions]. Prague: V. Hrubý.
- Ivić, Milka 1957–1958. Sistem predložkih konstrukcija u srpskohrvatskom jeziku [System of prepositional constructions in Serbo-Croatian]. *Južnoslovenski filolog* XXII: 1–4, 141–166.
- Ivić, Milka 1983. *Srpskohrvatski padežni sistem za označavanje prostornih odnosa* [The system of Serbocroatian cases denoting spatial relations]. In *Lingvistički ogledi*. Beograd: Prosveta.
- Jelínek, Milan 1964. Výrazy předložkové povahy v dnešní spisovné češtině [The nature of prepositional expressions in contemporary standard Czech]. *Sborník prací Filozofické fakulty brněnské univerzity*, A 12, 117–130.
- Klajn, Ivan 2002. *Tvorba reči u savremenom srpskom jeziku*, t. 1–t. 2 [Word formation in contemporary Serbian]. Beograd: Zavod za udžbenike i nastavna sredstva.
- Krijezi, Merima 2012. *Značenje predloga u albanskom i srpskom jeziku* [The meaning of prepositions in Albanian and Serbian]. PhD dissertation. Beograd: Filološki fakultet Univerziteta u Beogradu.
- Kurzon, Dennis & Silvia Adler (eds). 2008. *Adpositions. Pragmatic, semantic and syntactic perspectives*. Amsterdam & Philadelphia: J. Benjamins.
- Lalević, Miodrag 1955. Iz sintakse reči, oblika i rečenica [On the syntax of words, forms and sentences]. *Nastava i vaspitanje* 7, 393–402.
- Lehmann, Christian 1991. Grammaticalization and related changes in contemporary German. In Elizabeth Traugott & Bernd Heine (eds.), *Approaches to grammaticalization*, II: 493–535. Amsterdam & Philadelphia: J. Benjamins.
- Lehmann, Christian 2002a. *Thoughts on Grammaticalization*. Erfurt: Seminar der Sprachwissenschaft. Second Edition.
- Lehmann, Christian 2002b. New Reflections on Grammaticalization and Lexicalization. In Ilse Wiescher & Gabriele Diewald (eds.), *New Reflections on Grammaticalization*, 1–18. Amsterdam: John Benjamins.
- Lehmann, Christian 2019. Complex spatial prepositions from Latin to Castilian. *Revue romane* 54(1), 93–125.

- Leeman, Danielle 2007. De la préposition à la locution prépositionnelle. *Modèles linguistiques* 55, 7–15.
- Lima, José P. 2014. Grammaticalization of complex prepositions in European Portuguese. In José Pinto de Lima, *Studies on Grammaticalization and Lexicalization*, 163–183. Munich: Lincom Europa.
- Lima, José P. 2019a. On grammaticalized complex prepositions in Portuguese: Deployment, shift, redundancy, complementation. *Revue Romane* 54(1), 126–140.
- Lima, José P. 2019b. Preposições complexas em Português. In Ernestina Carrilho, Ana Martins, Sandra Pereira, João P. Silvestre (orgs.). *Estudos linguísticos e filológicos oferecidos a Ivo Castro*, 1181–1211. Lisboa: Centro de Linguística da Universidade de Lisboa.
- Ljubešić, Nikola & Filip Klubička 2014. {bs,hr,sr}WaC – Web corpora of Bosnian, Croatian and Serbian. Proceedings of the 9th Web as Corpus Workshop (WaC-9), ACL, 2014.
- McGranahan, William 1971. V in Phrasal and Complex Prepositions in Scientific Russian. *The Slavic and East European Journal* 15 (3), 332–338. (www.jstor.org/stable/306828)
- Meillet, Antoine 1902. *Études sur l'étymologie et le vocabulaire du vieux slave*. Paris: Librairie Emile Bouillon.
- Melis, Ludo 2003. *La préposition en français*. Paris: Ophrys.
- Melíšková, Petra 2008. *Secondary prepositions in English and Czech contemporary journalistic texts*. Unpublished Bachelor's thesis, University of Prague, Prague.
- Miletic, Aleksandra, Dejan Stosic & Saša Marjanović 2017. ParCoLab: A Parallel Corpus for Serbian, French and English. In Kamil Ekštejn & Václav Matoušek (eds.), *Text, Speech, and Dialogue. TSD 2017*, 156–164. Lecture Notes in Computer Science book series, vol 10415. Springer, Cham.
- Mrazović, Pavica 2009. *Gramatika srpskog jezika za strance* [Serbian grammar for foreigners]. Sremski Karlovci: Izdavačka knjižarnica Zorana Stojanovića.
- Popović, Ljubomir 1966. Predložki izrazi u savremenom srpskohrvatskom jeziku [Phrasal prepositions in contemporary Serbian], *Naš jezik*, XV, vol. 3–4, 195–220.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik 1985. *A Comprehensive Grammar of the English Language*, London: Longman.
- Rišner, Vlasta 2001. *Prijedlozi u djelima slavonskih književnika od 18. do 20. stoljeća* [Prepositions in the works of Slavonian writers from the 18th to the 20th century]. PhD dissertation. Zagreb: Filozofski fakultet u Zagrebu.
- Šarić, Ljiljana 1999. Gramatikalizacija i sekundarni prijedlozi (*pomoću, s pomoću, uz pomoć*) [Grammaticalization and secondary prepositions (*pomoću, s pomoću, uz pomoć*)]. *Riječ: časopis za filologiju* 5(1), 81–96.
- Šekrst, Kristina 2011. *Baltoslavenski adpozicijski sintaktički odnosi* [Balto-Slavic adpositional syntactic relationships]. Undergraduate dissertation. Zagreb: Sveučilište u Zagrebu, Filozofski fakultet.
- Šekrst, Kristina 2013. *Prijedlozi kao riječi uza sve leksičke kategorije s posebnim naglaskom na udvajanje prijedloga u južnoslovenskim jezicima*. [Prepositions as a loose lexical category, with special emphasis on prepositional doubling in South-Slavic languages]
- Seppänen, Aimo, Rhonwen Bowen & Joe Trotta 1994. On the so-called complex prepositions. *Studia Anglica Posnaniensia* 29: 3–29.
- Šiganova, Galina Aleksandrovna 2001. *Sistema leksičkih i frazeoloških predlogov v sovremennom rusckom jazyke* [The system of lexical and phraseological prepositions in contemporary Russian]. Unpublished PhD dissertation, University of Tobol'sk.

- Stevanović, Mihailo 1991. *Savremeni srpskohrvatski jezik*, II: *Sintaksa* [Contemporary Serbian, II: Syntax]. Beograd: Naučna knjiga.
- Stanojčić, Živojin & Ljubomir Popović 2011. *Gramatika srpskog jezika za gimnazije i srednje škole* [Serbian grammar for grammar and professional schools]. Beograd: Zavod za udžbenike.
- Stosic, Dejan forthcoming a. Les prépositions complexes en français. *Carnets de grammaire*.
- Stosic, Dejan forthcoming b. Complex prepositions. In *Encyclopedia of Slavic Languages and Linguistics*. Brill.
- Stosic, Dejan & Benjamin Fagard 2019. Les prépositions complexes en français: Pour une méthode d'identification multicritère. *Revue romane* 54(1), 8–38.
- Švačko, Vanja 1994. Prijedložni izrazi kao gramatička i leksikografska jedinica [Prepositional expressions as grammatical and lexicographical units], *Filologija* 22–23, 307–311.
- Zwarts, Joost 1997. Complex prepositions and P-Stranding in Dutch. *Linguistics* 35, 1091–1112.

Corpora

ParCoLab: <http://parcolab.univ-tlse2.fr>

SrWaC: via <https://www.sketchengine.eu>

Dictionaries

Jovanović, Slobodan 2014. *Francusko-srpski rečnik*, Beograd: Prosveta.

Putanec, Valentin 1995. *Francusko-hrvatski rječnik*, Zagreb: Školska knjiga.

RMS. *Rečnik Matice Srpske*. Beograd: Prosveta.

Appendix

Patterns of CAs in Modern Serbian (on the basis of grammars, dictionaries and existing literature on CAs; the list cannot be exhaustive, but is representative. Numbers indicate the number of items found for each pattern).

Pattern	Number of CPs	Examples of CAs
[P ₁ N]	268	<i>iz razloga</i> ‘due to, because of’, <i>iz ugla</i> ‘from the vantage point of’, <i>na čelu</i> ‘at the head of’, <i>na dnu</i> ‘at the bottom of’, <i>na domak</i> ‘within the range of’, <i>na kraju</i> ‘at the end of’, <i>na nivou</i> ‘at/to the level of’, <i>na osnovu</i> ‘based on’, <i>na račun</i> ‘at the expense of’, <i>na vrhu</i> ‘at the top of’, <i>o trošku</i> ‘at the expense of’, <i>od strane</i> ‘on behalf of’, <i>po nalogu</i> ‘by direction of’, <i>pod dejstvom</i> ‘under the influence of’, <i>preko puta</i> ‘opposite, over, across’, <i>u čast</i> ‘in honor of’, <i>u cilju</i> ‘with the aim of’, <i>u duhu</i> ‘in the spirit of’, <i>u interesu</i> ‘in the interest of’, <i>u nastavku</i> ‘in the continuity of’, <i>u okviru</i> ‘in the framework of, in the scope of’, <i>u periodu</i> ‘at the time of’, <i>u ritmu</i> ‘at the pace of’, <i>u sferi</i> ‘in the sphere of’, <i>u susret</i> ‘towards’, <i>u toku</i> ‘during’, <i>u znak</i> ‘in sign of’, <i>uz pomoć</i> ‘with the help of’, <i>van dometa</i> ‘outside the scope of’, <i>za ljubav</i> ‘for (the) love of, for’, <i>za vreme</i> ‘during’
[P ₁ N P ₂]	60	<i>bez obzira na</i> ‘without regard to’, <i>iz obzira prema</i> ‘having regard to’, <i>iz straha od</i> ‘for fear of’, <i>na čelu sa</i> ‘headed by’, <i>na putu ka</i> ‘on the path towards’, <i>po uzoru na</i> ‘following, based on the model of’, <i>pri pomisli na</i> ‘at the thoughts of’, <i>s obzirom na</i> ‘given’, <i>u iznosu od</i> ‘to the value of’, <i>u odnosu na</i> ‘with respect to’, <i>u potrazi za</i> ‘in search of’, <i>u razmaku od</i> ‘within an interval of’, <i>u roku od</i> ‘within, in’, <i>u skladu sa</i> ‘in accordance with’, <i>u strahu od</i> ‘for fear of’, <i>u vezi sa</i> ‘in relation to’, <i>u visini od</i> ‘of up to’, <i>u zamenu za</i> ‘in exchange for’, <i>u zavisnosti od</i> ‘depending on’, <i>za razliku od</i> ‘unlike’

Pattern	Number of CPs	Examples of CAs
[Adv P ₂]	24	<i>daleko od</i> 'far away from', <i>desno od</i> 'to the right of', <i>južno od</i> 'južno', <i>levo od</i> 'to the left of', <i>nedaleko od</i> 'not far away from', <i>nevezano za</i> 'regardless of', <i>srazmerno sa</i> 'in proportion to', <i>suprotno od</i> 'opposite to', <i>uporedo sa</i> 'parallel to', <i>vezano za</i> 'related to', <i>više od</i> 'more than', <i>zavisno od</i> 'depending on'
[P ₁ Adj N]	18	<i>na drugom kraju</i> 'at the other end of', <i>na levoj strani</i> 'on the left-hand side of', <i>na prednjoj strani</i> 'on/at the front side of', <i>na suprotnoj strani</i> 'on the opposite side', <i>s druge strane</i> 'on the other side of', <i>s prednje strane</i> 'on the front side of', <i>sa leve strane</i> 'on the left-hand side of', <i>sa obe strane</i> 'on each side of', <i>u zadnjem delu</i> 'on the back of'
[V P ₂]	9	<i>misleći na</i> 'thinking of', <i>oslanjajući se na</i> 'leaning upon', <i>počev od</i> 'from, since', <i>polazeći od</i> 'starting from', <i>sudeći po</i> 'to judge from'
[N P ₂]	5	<i>obzirom na</i> 'given', <i>nalik na</i> 'like', <i>korak do</i> 'one step away from'
[P ₁ Adj N P ₂]	2	<i>na suprotnoj strani od</i> 'on the opposite side of'
[V P N]	2	<i>imajući u vidu</i> 'having regard to'
[Neg V P ₂]	2	<i>ne gledajući na</i> 'regardless of'
[P ₁ Adv P ₂]	2	<i>za manje od</i> 'in/for less than'
[P ₁ N i N]	1	<i>na milost i nemilost</i> 'at the mercy of'
[P ₁ N P ₂ N]	1	<i>s kraja na kraj</i> 'from end to end'
[Pron V]	1	<i>što se tiče</i> 'as for, concerning'

Pattern	Number of CPs	Examples of CAs
[P ₁ Adj i Adj N]	1	<i>s jedne i druge strane</i> ‘on each side of’
[Num N P ₂]	1	<i>dva koraka od</i> ‘two steps away from’
[P ₁ Num N P ₂]	1	<i>na dva koraka od</i> ‘two steps away from’
[P ₁ Adv N P ₂]	1	<i>na pola puta od</i> ‘halway to’
[P ₁ Adj N P ₂ Adj]	1	<i>s jednog kraja na drugi</i> ‘from one to the other side of’
[Adv Adv]	1	<i>obrnuto proporcionalno</i> ‘inversely proportional to’
Total	401	

Kevin J. Rottet

4 Complex prepositions in Breton

Abstract: In this chapter we report on the use of simple and complex adpositions in Breton, a Brythonic Celtic language of western France. Prepositions represent a significant part of what is distinctive about the Celtic languages. Many prepositions trigger initial consonant mutations (*Breizh* ‘Brittany’ but *da Vreizh* ‘to Brittany’); simple prepositions are inflected for pronominal complements (*evit ma breur* ‘for my brother’ but *evitañ* ‘for him’). Complex prepositions come in several patterns, including one in which pronominal complements take the form of possessive infixes (*a-raok* ‘before’ but *en ho raok* ‘before you’, cf. *ho* ‘your’) and another in which locative prepositions can be given dynamic interpretations by adding or changing a prefix or an element perceived as one (*war* ‘on’ versus *diwar* ‘from on’, or *etre* ‘between’ versus *a-dre* ‘from between’). A corpus study based on thirty novels, autobiographies and collections of folktales from the 19th to the 21st century reveals more variation in complementation patterns and attested combinations than is noted in the descriptive and pedagogical literature. The corpus also reveals robust cases of ablative-locative transfer in which ablative prepositions can come to be interpreted as static locatives over time, requiring another element to be added to restore the ablative sense, in a cycle that may then be repeated.

Keywords: Celtic, Brythonic, Breton, inflected prepositions, possessive infixes, ablative-locative transfer

1 Introduction

Prepositions represent a significant part of what is distinctive about the Celtic languages. Not only do they have unusual typological properties such as being inflected or “conjugated” for pronominal objects, they are also extensively used in the expression of verbal aspect and various kinds of idiom such as the expression of possession. This case study of Celtic prepositions is based on Breton, a highly endangered Brythonic Celtic language of western France. Arguably less standardized than the other languages in the family, the Breton vernaculars may reveal developments largely unfettered by deliberate intervention.

Kevin J. Rottet, Department of French & Italian, Indiana University, Bloomington, IN, U.S.A.
krottet@indiana.edu

<https://doi.org/10.1515/9783110686647-005>

Prepositions are quite numerous in Breton. The great majority of them are Complex Prepositions (CPs) (often called *locutions prépositives* in French, the language of most of the descriptive studies of Breton). Various aspects of Breton prepositional usage are rather more variable than is generally acknowledged in descriptive grammars and pedagogical works, particularly concerning complementation patterns and permutations to distinguish static location and movement. The portrayal of Breton prepositions that we will lay out here is based primarily on the study of a written corpus consisting of some thirty novels, autobiographies and collections of folktales dating from the late 19th century up to the early 21st century. Comparisons will also be made with the descriptive grammatical literature.

This article is structured as follows. In section 2 we will give a brief overview of some typological properties of the Celtic languages. In section 3 we lay out the basic facts about simple prepositions, some of the details of which are essential for understanding CPs. In section 4 we discuss CPs in Breton, describing the different complementation patterns and attested variation. Finally, in section 5 we treat permutations which distinguish location from movement, including a discussion of ablative-locative transfer, and section 6 presents some concluding remarks.

2 The Celtic languages: A brief overview of typological features

The modern Celtic languages fall into two branches, called Goidelic or Q Celtic (Irish, Scottish Gaelic and Manx) and Brythonic or P Celtic (Welsh, Breton and Cornish). Similarities are much greater within each branch than between the two branches, though the Goidelic languages are closer to each other than the Brythonic languages are; mutual intelligibility across the two branches is impossible. Within Brythonic, Breton and Cornish are closer to each other than either is to Welsh. The Cornish language officially died out in the late eighteenth or early nineteenth century while Manx lingered on until as recently as the 1970s, and both languages today are subject to revival movements. The contemporary variety known as Revived Cornish is based not on Late Cornish but on Middle Cornish, which was less anglicised and in fact better documented; part of the elaboration of Revived Cornish has involved looking to Breton and Welsh to fill lexical and paradigmatic gaps.¹

¹ For these reasons, in this article we will primarily be concerned with the two continuously spoken Brythonic languages, Breton and Welsh.

Breton is spoken in several divergent dialects traditionally associated with the four historical dioceses of Lower Brittany known as Cornouaille, Léon, Trégor and Vannes, giving the dialect names *kerneveg*, *leoneg*, *tregerieg* and *gwenedeg*. Gwenedeg is by far the most divergent, with the result that it is often treated separately from the other three dialects, which are referred to collectively as KLT from the first letters of their Breton names. The standard written norm is based heavily on *leoneg*, though traditional native speakers rarely know the standard. In the course of the twentieth century, no fewer than three different orthographies were proposed. By the end of the century more than 80% of publishing used the *peurunvan* ('completely unified') orthography associated with the University of Rennes, while the remainder used the "university" orthography (*skolveurieg*) associated with the University of Brest.²

The Celtic languages share a number of typological properties. The basic word order is VSO, except in Breton (and Cornish) that have a V2 constraint in main clauses. Celtic languages have definite but not indefinite articles (although here, too, Breton is exceptional, having grammaticalized reflexes of its numeral 'one' as an indefinite article), and the traditional number system is partially vigesimal, e.g. Breton *pevar-ugent* 'eighty (lit. four-twenty)'. They all have a genitive construction in which two nouns can be juxtaposed in the order possessed-possessor. The possessed noun (*kambr* 'room' in (1)) is interpreted as definite and cannot be marked with any overt determiner:

- (1) kambr ma c'hoar
 room my sister
 'my sister's room'

² In this article works are cited in their published spellings, so the reader should expect to encounter orthographic as well as dialect variation. Some of the phenomena described are subject to dialect variation, though generally not in ways that affect the claims being made. For instance, in Leoneg, the prepositions usually spelled *ouzh* 'at, against' and *diouzh* 'from, from-at' in the standard written language (pronounced [uz ~ us] and [djuz ~ djus]) generally take the forms *o'ch* [ox] and *dioc'h* [djoχ]; the preposition *etre* 'between' is usually reinforced to *kenetre*; the inflected forms of *war* 'on' use a stem *warnez-* rather than standard *warn-* (e.g. *warnezañ* for *warnañ* 'on him'); and the 1sg inflectional ending takes the form *-en* rather than standard *-in* (e.g. *ganen* 'with me' for *ganin*). There are also dialect prepositions that are not part of the standard language; Leclerc notes *merak* 'if not for' and *hepout da* 'unbeknownst to' (pp. 112–113) as Tregor dialect forms. These and other details of dialect variation will not be mentioned unless relevant to the discussion.

The pattern is recursive, so one genitive phrase can be put inside another (2):

- (2) dor kambr ma c'hoar
 door room my sister
 'the door of my sister's room'

In the Goidelic languages, the possessor is marked using a residual genitive case, whereas in Brythonic, morphological case has been entirely lost. The primary nonfinite form of Celtic verbs, the so-called verbal noun, has certain noun-like properties, including the fact that pronominal direct objects are homophonous with possessives. Perhaps the most well-known feature of the Celtic languages is initial consonant mutation (ICM), in which the initial consonant of lexical words undergoes specific changes in various morphosyntactic contexts such as after certain possessives, numbers, articles and simple prepositions. ICMs also play a role in marking grammatical agreements such as gender.

Prepositions present several striking typological features in the Celtic languages. Simple prepositions are inflected for pronominal complements (e.g. Breton *gant ma zad* 'with my father' but *gantañ* 'with him'), and these languages lack a lexical verb of possession, instead using prepositional possessive constructions. The two Breton constructions illustrated below differ in that (3), using *gant* 'with', expresses temporary possession, while (4), using a clitic pronoun with dative interpretation (here *en*, understood as 'to him') and a copula, is for more permanent possession:

- (3) Ne oa bag ebet gant ma zad. (Bijer 140)
 NEG was_{3SG-IMP} boat at-all with my father
 'My father had no boat at all (with him on that occasion).'
- (4) Va zad en deus ur c'hi ivez, met brasoc'h eo. (Ur vouezh 90)
 my father 3SGM is a dog also but bigger is
 'My father has a dog also, but it's bigger.'

All of the Celtic languages deploy verbal constructions consisting of a copula, a preposition and a verbal noun to mark aspect, as in the progressive in Scottish Gaelic (5) and Breton (6):

- (5) Tha am balach ga cheannach.
 is the boy at_{3SGM} buying_{VN}
 'The boy is buying it.' (Macaulay 2008: 183)

- (6) “Ha hennzh eo an den emaoch
 and that-one is the person are_{ZPL}
ouzh hor c’has d’ e vete?” (Hob 141)
 at our send_{VN} to his *bete*
 ‘Is that the person you are taking us to now?’

Having given a brief overview of some Celtic typological features, we turn now to an in-depth examination of prepositions in Breton, beginning with a brief overview of simple prepositions.

3 Simple prepositions in Breton

Breton has just under twenty simple prepositions³; these can straightforwardly take a lexical but not a pronominal complement. The list of simple prepositions includes *a* ‘of, from’, *da* ‘to’, *dre* ‘through’, *dreist* ‘over’, *e(n)* ‘in’, *eget* ‘than’, *etre* ‘between’, *eus* ‘of, from’, *evel* ‘like, as’, *evit* ‘for; than’, *gant* ‘with; by’, *goude* ‘after’, *hebiou* ‘past’,⁴ *hep* ‘without’, *hervez* ‘according to’, *kent* ‘before’, *ouzh* ‘at, against’, *rak* ‘before’ and *war* ‘on’. To illustrate different uses, we give examples

³ In the interest of making frequency comparisons with other languages, we calculated frequencies for a selection of Breton prepositions based on a 200,000-word subcorpus of extracts from nine texts by nine different authors or translators (Ar Gow 1978, Carroll 1995, Drezen 1977, Dupuy 2005, Duval 1993, Gerven 2002, Huon 1980, Riou 1995, Tolkien 1995). This selection of texts covers several genres (autobiography, original Breton fiction, translation) as well as writers from each of the four traditional Breton dialect areas. The results for several prepositions are as follows (the number of occurrences is out of just over 200,000 words): *goude* ‘after’ (267), *diwar* ‘from-on’ (201), *dindan* ‘under’ (154), *war-lerc’h* ‘after’ (146), *digant* ‘from-with’ (112), *e-kichen* ‘beside’ (107), *abalamour da* ‘because of’ (85), *hervez* ‘according to’ (76), *a-dreñv* ‘behind’ (65), *a-drek* ‘behind’ (36), *e-tal* ‘beside, near’ (26), *e-keñver* ‘across from; with regard to’ (27), *a-enep* ‘against’ (9), *edan* ‘under’ (7), *da geñver* ‘for, on the occasion of (an event)’ (8), *e-skoaz* ‘compared with’ (4). The figures given include variants as well as permutations with internal possessives, where applicable; thus, *war-lerc’h* includes *war ma lerc’h* ‘after me’, etc. The only really conspicuous facts about the distribution of these prepositions across the texts are: (1) *edan* ‘under’, a Gwenedeg dialect variant of *dindan*, occurs only in the translation of the Hobbit – and even that text uses the standard variant almost twice as often; (2) The less frequent variant for ‘behind’, *a-drek*, occurs in four of the nine texts while the standard variant *a-dreñv* occurs in all texts except *Ur vouezh er vorenn*, the only one to consistently use *a-drek*. This text is not heavily dialectal but its author was from the Tregor region. The exclusive use of *a-drek* may simply reflect his personal preference.

⁴ In the oldest pattern, *hebiou* is an adverb which takes the AuxPrep *da* ‘to’ before a complement. However, in the modern language *da* may be absent and *hebiou* may function as a preposition (Hémon 1984: 108).

of *evit* ‘for’ showing an NP complement (7), a verbal noun complement (8), and its use coupled with *ma* to form the conjunction ‘so that, in order that’ in (9)⁵:

- (7) Penaos ho trugarekain, Armel, *evit* ho madelez ? (Mezvier 417)
 how 2PL thank_{FUT.1SG} Armel for your goodness
 ‘How will I thank you, Armel, for your goodness?’
- (8) Roet ’m eus de’i⁶ alc’hweziou ma oto
 given_{PAP} to-me is to_{3SG-F} keys my auto
evit mont davit ur sac’had pechez (WTL 182)
 for go_{VN} after a sackful peaches
 ‘I gave her the keys to my car so she could go (*lit.* for to go) fetch a sack of peaches’
- (9) *evit ma* vezimp ar c’henta ’r gwella unanet (Mezvier 417)
 for that be_{FUT.1PL} the first the best united_{PAP}
 ‘so that we can be wed as soon as possible’

Two of the etymologically simple prepositions, *etre* ‘between’ and *hebiou* ‘past’,⁷ have in some dialects been reanalyzed bimorphemically as *e-tre* and *e-biou* by analogy with CPs that consist of *e* ‘in’ plus a relational noun (e.g. *e-kichen* ‘beside’, *e-tal* ‘in front of’, cf. *kichen* ‘side’, *tal* ‘brow’). These reanalyses reveal that CPs in Breton can involve relational nouns that never existed as autonomous nouns, a point to which we will return.⁸

⁵ A number of prepositions can be used as conjunctions by joining them with *ma* ‘that’, e.g. *evit* ‘for’, *evit ma* ‘so that, in order that’. Other common conjunctions following this model include *goude ma* ‘after’, *a-raok ma* ‘before’, *dre ma* ‘since, because’ (cf. *dre* ‘through’), and *hervez ma* ‘as, in the same way that’ (cf. *hervez* ‘according to’).

⁶ The nonstandard spelling *de’i* ‘to her’ for what is usually spelled *dezhi* reflects the most common pronunciation of this form. One similarly finds *de’o* for *dezho* ‘to them’ and *de’añ* for *dezhañ* ‘to him’.

⁷ *Etre* and its Irish cognate *idir* are cognate with Latin *inter* (Deshayes 2003, Jackson 1967: 794–95). Even though the Indo-European word is etymologically polymorphemic, there is no basis on which to claim this of its modern descendants including French *entre* or the Celtic words *etre* / *idir*. *Hebiou* ‘past’ (Welsh *heibio*) is from an adverbialized form of the preposition *hep* ‘without’ (Jones 1913: 401–02, 04).

⁸ The opposite tendency is also found, namely the case where an etymologically polymorphemic preposition (i.e. a CP) comes to be analyzed as morphologically simple by modern speakers. The preposition *davit* ‘after, in order to fetch’ (as seen in example [8]) is such a case. Kerrain (2010: 60) indicates that it is a contraction of *da-evit*, and *davet* with similar form and semantics is a contraction of *da + betek*. It seems unlikely that modern speakers recognize the internal composition of these.

A few simple prepositions trigger initial consonant mutations (ICMs), in the immediately following word, of the type known as lenition.⁹ Such prepositions include *da* ‘to’, *war* ‘on’, *diwar* ‘from, from-on’, *a* ‘from’, and *dre* ‘through’. Thus, we get *Montroulez* ‘Morlaix’ but *a Vontroulez* ‘of, from Morlaix’, *kousket* ‘to sleep’ but (*mont*) *da gousket* ‘(to go) to sleep’, *gras* ‘grace’ but *dre c’hras* ‘by grace’, and *bigi* ‘boats’ but *war vigi* ‘on boats’.

With very few exceptions (to be discussed below), prepositions cannot take independent pronominal complements directly.¹⁰ Rather, most simple prepositions are inflected or ‘conjugated’, taking endings that agree in person, number and gender (the latter for the 3sg only). There are two sets of endings, those in *-o* and those in *-i*:

	<u>-o endings</u>	<u>-i endings</u>
1sg	-on	-in
2sg	-out	-it
3sg-m	-añ	-añ
3sg-f	-i	-i
3sg-gen ¹¹	-or	-or
1pl	-omp	-imp (or -eomp)
2pl	-oc’h	-eoc’h
3pl ¹²	-e / -o	-e / -o

With the *-o* endings, the stress pattern is penultimate, as is usual in the KLT dialects, but with the *-i* endings stress is on the suffix in the first and second persons singular and plural (Ternes 2008: 426). The *-i* endings are used only with *da* ‘to’,

⁹ In lenition, initial [p], [t], [k] become [b], [d], [g], respectively; initial [b], [d], [g] become [v], [z], [h] or [y], and [m] becomes [v]. These changes are all reflected in the current orthographies. Some additional consonants undergo lenition in some dialects, but these cases are not consistently represented in spelling; for instance [ʃ] becomes [z] and [f] becomes [v]. The latter are often called “new lenition”.

¹⁰ Desbordes (1983: 75) even states explicitly that the sequence preposition + personal pronoun is impossible in Breton. The Celtic languages all make use of inflected prepositions but differ in the extent to which uninflected prepositions governing independent pronouns are also found. Unlike Breton, Welsh has a number of simple prepositions that take pronominal complements directly.

¹¹ The ending labeled *3sg-gen* in the table above is a “generic” form, used for a non-specific human subject. Some Breton grammars give this form (e.g. Favé 2003: 7) though many others do not. It seems to be very infrequent in actual usage.

¹² The 3pl ending (in both paradigms) is *-e* in most areas, but *-o* in the Tregor dialect.

gant ‘with’, *digant* ‘from-with’, *ouzh* ‘at’ and *diouzh* ‘from-at’, although dialect variation can move a preposition from one paradigm into another. The paradigms are illustrated below with the prepositions *hervez* ‘according to’ (-o endings) and *ouzh* ‘at, towards’ (-i endings):

	<u>hervez</u> ‘according to’	<u>ouzh</u> ‘at, towards’
1sg	hervezon ‘according to me’	ouzhin ‘at me’
2sg	hervezout ‘according to you’	ouzhit ‘at you’
3sg-m	hervezañ ‘according to him’	outañ ‘at him’
3sg-f	hervezi ‘according to her’	outi ‘at her’
1pl	hervezomp ‘according to us’	ouzhimp ‘at us’
2pl	hervezoc’h ‘according to you’	ouzhoc’h ‘at you’
3pl	herveze / hervezo ‘according to them’	oute / outo ‘at them’

The first and second person endings historically come from the inflectional suffixes of the verb *bezañ* ‘to be’: the -o endings are drawn from the present indicative (e.g. *on* ‘I am’, *out* ‘you are’ etc.), while the -i endings come from the former present subjunctive turned future (e.g. *bezin* ‘I will be’, etc.) (Favereau 1997: 408). Dialect variation in some of the endings generally involves substituting other verb endings such as the informal 2sg -*ez* [əs] which competes with -*out* [ut] and -*it* [it] (Kervella 336), thus giving forms like *ganez* instead of *ganit* in some dialects (cf. *komzez* ‘you speak’).^{13,14} Wmffre (1998: 29), in his description of the Breton of a centrally located commune (Plounévezel) in the middle of Breton-speaking Brittany, in fact gives -*ez* as the regular 2sg ending in that dialect.

The personal endings are attached to a bound form of the simple preposition. In a few cases, the bound form is the same as the simple preposition, as for *hervez* seen above, but more often the bound form is a modified stem ending in an epenthetic consonant (Press 1986: 116 ff.) which is variably *n* (e.g. *war* / *warn-* ‘on’; *e* ‘in’ / *enn-*), *z* (e.g. *etre* / *etrez-* ‘between’; *dre* / *drez-* ‘through’), or *d/t* (e.g. *hep* / *hepd-* ‘without’, *e-giz* / *egist-* ‘like’):

- (10) ur c’heloù hag en deus graet efed fall **warnout**. (WTL 44)
 a news-item PRT 3SGM is made_{PAP} effect bad on_{2SG}
 ‘a piece of news that had a bad effect on you.’

¹³ Forms like *ganez* are rigorously avoided in the written language (Davalan 2001: 250).

¹⁴ Similarly, in Welsh the 3pl preposition ending is -*yn(t)*, which is directly borrowed from the present tense verb paradigm (Jones 1913: 398), replacing the earlier endings that are cognate with those used in Breton.

- (11) Brenda, [...] ne c'hellan ket bevañ **hepdout**. (Ur vouezh 132)
 Brenda NEG can_{1SG} not live_{VN} without_{2SG}
 'Brenda, I cannot live without you.'

A regular alternation has a stem-final *d* in first and second persons and a *t* in the third person. This group includes *evel* 'like', in which the consonant is epenthetic (cf. *evel dout* 'like you', *evel ti* 'like her'), as well as *eget* 'than', *evit* 'for', *nemet* 'but, only' and *davet* 'towards' (e.g. *davedon* 'to me', *davetañ* 'to him') where the final consonant is part of the stem.¹⁵

A few prepositions simply have an irregular paradigm with stem variation. Thus, *gant* 'with' has the stem *gan-* in first and second persons, and *gant-* in 3sg and 3pl, while *ouzh* 'at, against' has the stems *ouzh-* and *out-* for these respectively.

With conjugated verbs (as in (12)) as well as with inflected prepositions (as in (13)), the personal inflection may optionally be reinforced by an overt pronoun, for emphasis¹⁶:

- (12) Penôs e rin **-me** evit dizrei ac'han? (PG 90)
 how PRT do_{1SG-FUT} 1SG for turn-back_{VN} from-here
 'How will I get out of here?'

- (13) C'houi a dle gouzout gwelloc'h **egedon-me...** (Milin 606)
 you PRT should_{3SG} know_{VN} better than_{1SG} 1SG
 'You should know better than I do...'

Some simple prepositions lack inflected forms. In some instances this is semantically justified; for instance, *abaoe* 'since' is a temporal preposition, so there is no reason for it to take personal complements, but in other cases, the gap is not semantically motivated. The prepositions *a* and *eus*, both meaning 'of, from', share a single inflectional paradigm, which is historically the paradigm of *a*; *eus* has no paradigm of its own. In other cases, a preposition that cannot be directly inflected is joined with an inflectable auxiliary preposition (an AuxPrep) so that it can take a complement. Thus, for instance, the prepositions *etrezek* 'towards'

¹⁵ The [d], considered underlying, changes to a [t] via provection, because the third person pronouns (from which the third person inflections historically derive) were originally [h]-initial, causing devoicing of the stem-final consonant. (Wmffre 1998: 29)

¹⁶ Note that pronouns do not have different case forms in Breton, so *me* is simply a 1SG pronoun that can occur in any syntactic position where a personal pronoun is called for. Prenominal possessive adjectives do have a different shape though, in this case *ma* or *va* (depending on dialect) for 1SG.

and *hebiou* ‘past’ do not have inflected forms and require an AuxPrep in order to take a pronominal complement, *e(n)* ‘in’ in the case of *etrezek*, whereas *hebiou* takes *da* ‘to’. In (14) and (16) we see how these work with a lexical complement, whereas (15) and (17) show the AuxPrep used with a pronominal complement:

- (14) a-dreuz lanneier, **etrezeg** ar c’hoad. (PG 198)
 across moors towards the woods
 ‘across moors, towards the woods’
- (15) Ober a ra daou gamm **etrezek ennon.** (WTL 68)
 make_{VN} PRT does two step towards in_{1SG}
 ‘He takes two steps towards me.’
- (16) Tremen a eure **e-biou** va zi. (Jezegou 19)
 pass_{VN} PRT did_{3SG-PRET} past my house
 ‘He went past my house.’
- (17) pa dremenas **hebiou dezhi** ur c’honikl gwenn,
 when passed_{3SG-PRET} past to_{3SGF} a rabbit white
 roz e lagad. (Alis 12)
 pink his eyes
 ‘when a white rabbit with pink eyes ran past her’

There are very few occasions in Breton where an uninflected preposition can take an independent pronoun as a complement. This possibility only arises when the complement of the preposition consists of two conjoined elements, i.e. either two pronouns, or a pronoun and a noun phrase (in either order). In such cases, the preposition may simply be repeated before each conjunct, and, of course, inflected when the immediately following conjunct is pronominal:

- (18) Dilun a zeu e vo cheñchamant
 Monday PRT comes PRT be_{3SG-FUT} change
evidoc’h hag evit ho kenderv. (Bijer 262)
 for_{2PL} and for your cousin
 ‘This coming Monday there will be a change for you and for your cousin.’

But if the preposition is not repeated, there are two patterns. It may be inflected when the first complement is pronominal, regardless of whether the second complement is lexical (19) or pronominal (20):

- (19) Bloavezhioù a erru, en devez a-boan a-walc’h
 years PRT arrived_{PAP} 3SGM is_{HAB} barely enough
evitañ hag e diegezh. (Duval 68)
 for_{3SGM} and his family
 ‘Years came in which he barely had enough for him and his family.’

- (20) o klask, moarvat, kompezañ an traoù **etrezañ ha me...** (Deizlevr 61)
 PRT try_{VN} perhaps smooth_{VN} the things between_{3SGM} and me
 ‘seeking, perhaps, to smooth things out between him and me...’

Or the preposition may remain uninflected, thus taking one or more independent pronouns as its complement:

- (21) Meur a wech e oa bet ar gaoz -mañ
 many a time PRT was been_{PAP} the conversation -this
etre eñ ha Jefinig. (WTL 35)
 between him and Jefinig
 ‘Many a time there had been this conversation between him and Jefinig.’

- (22) Naig ne oar ket c’hoazh komz brezhoneg
 Naig NEG know_{3SG} not yet speak_{VN} Breton
evel te ha me... (WTL 174)
 like you and me
 ‘Naig still doesn’t know how to speak Breton like you and me...’

Some pedagogical materials for Breton teach the latter as the correct pattern and criticize making the preposition agree with only one conjunct. For instance, in his online grammar column¹⁷ the novelist and grammarian Yann Gervenn argues that a conjunction should join two like things, e.g. two nouns or two adverbs, and, he argues, this principle is violated in *etrezañ ha me* ‘between him and me’, since *me* is a pronoun while *etrezañ* is an inflected preposition.¹⁸ But other materials teach both

¹⁷ Accessed June 6, 2018 at: <http://www.reseau-canope.fr/tes/yanngerven/index.php?pennad=B024>

¹⁸ There is much variation in what Celtic inflected (or “conjugated”) prepositions are called, and it is not unusual to consider them a kind of pronoun. For Breton, Hardie (1948: 127 ff.) called

patterns without expressing a preference for either one (e.g. Vallée 1902: 111; Chalm 2008: 194-95). Indeed, some writers vary between the two patterns in their writing:

- (23) **etre ni ha Breizhveuriz** (Taldir 180)
 between us and Britons
 ‘between us and Britons’
- (24) Ne oa ket a zarempred **etrezomp hag i...** (Taldir 51)
 NEG was_{IMP} not of communication between_{1PL} and them
 ‘There was no communication between us and them...’

At any rate, the pattern of leaving the preposition uninflected with conjoined complements is well attested, but it stands out as an isolated case in which prepositions can take independent pronominal complements in Breton.

Some simple prepositions tend to be replaced with phonetically more robust competitors. For instance, *a* ‘from’ is frequently reinforced with *eus* or simply replaced by it altogether:

- (25) em eus klasket gouzout en aner **eus a** belec’h
 1SG is sought_{PAP} know_{VN} in vain from from where
 e teu ar poz -se. (EST 64)
 PRT come_{3SG} the lyrics -that
 ‘I tried to vain to learn where those lyrics come from.’

Similarly, *e* ‘in’ (*en* before a vowel) is often replaced by *e-barzh*, and *rak* ‘before, in front of’ is generally ousted by the prefixed *dirak* today, except in certain collocations such as with *aon* ‘fear’.

Finally we note that, although preposition stranding is not traditionally allowed in the Celtic languages, inflected prepositions can and do regularly occur in clause-final position:

- (26) ar savadur m’ emañ e ranndi **ennañ.** (WTL 21)
 the building REL is his apartment in_{3SGM}
 ‘the building that his apartment is in (it).’

them *preposition-pronouns* and Vallée (1902) called them *pronomes composés* ‘compound pronouns’. In the Goidelic languages they can be called *prepositional pronouns* at least in Scottish Gaelic (Gillies 1993: 182) and Manx (Broderick 1993: 262).

This reveals that an inflected preposition is not *just* a preposition.

Having completed our survey of the main properties of simple prepositions, we are now ready to turn the focus to complex prepositions (CPs).

4 Complex prepositions (CPs) in Breton

CPs in Breton are numerous, with no two grammars giving identical lists. Kerrain (2010: 60) claimed that Breton has around 300 prepositions; by our count, all but around nineteen of these are CPs. The most extensive coverage of Breton CPs by far is that of Gros (1982: 93–241) in his study of *éléments de stylistique trégorroise* ('stylistic elements in the dialect of Tregor').

The problem of enumerating CPs is complex, first, because it is not always clear whether a given sequence should be viewed as a lexicalized CP or just a free sequence. Since lexicalization is a gradual process, a number of strings are located on a continuum between free and lexicalized combination. For example, the adverb *sko* 'touching, in contact' can be joined with *ouzh* 'at':

- (27) a-zioc'h al lojeiz bihan **sko ouzh** an tour -meur. (Beyer 178)
 above the flat small touching at the tower -great
 'above the small flat, right up against the great tower'

This usage raises the possibility that *sko ouzh* is a CP meaning 'touching, in contact with', and it is sometimes so classified.

Secondly, the line of demarcation between preposition and adverb is blurred in that many prepositions can be used as adverbs (or vice versa) with no change in form, or their forms may differ for some speakers and writers but not others. For instance, 'behind' as a preposition is usually *a-dreñv* (varying with *a-drek*). The adverbial equivalent is most often *war-dreñv*, but *a-dreñv* and the reinforced form *en-a-dreñv* also occur adverbially. The latter can then be used to derive a new preposition, by adding the AuxPrep *da*, giving *en a-dreñv da*:

- (28) Kraou ar moc'h a zo **en a-dreñv d'** ar c'harri. (Bijer 182)
 sty the pigs PRT is in behind to the garage
 'The pigsty is behind the garage.'

Such facts complicate any attempt to count the number of prepositions.¹⁹

¹⁹ Here it is not our purpose to offer a formal count or provide any kind of exhaustive list.

For the analysis of CPs, we will be operating with the conceptual background provided by Fagard et al. (this volume). According to these authors, a simple preposition serves to do the following three things:

- a. introduce a constituent
- b. establish a relation between this constituent and a higher constituent, or the sentence as a whole
- c. contribute to determining the semantic nature of this relationship

In their model, complex adpositions do the same three things but are morphologically complex. In the case of a *canonical* complex adposition, each of the three functions is carried out by a different element, giving the shape P_1 (Det) N P_2 . In this template, P_1 and P_2 are both simple prepositions which correspond to functions (a) and (b) respectively, while N is the “nucleus”, the lexical element which provides (part of) the semantic content of the CP, thus corresponding to function (c). In the most usual case, N is a noun, which may be preceded by a determiner, hence the parenthetical (Det) of the formula. The canonical formula can be given succinctly as PNP, that is Preposition-Noun-Preposition. Consider the following French example:

- (29) à côté de
 at side of
 ‘beside, next to’

Here, the simple preposition *à* ‘to, at’ corresponds to P_1 , while *de* ‘of, from’ is the P_2 and the noun *côté* ‘side’ is the semantic nucleus. Fagard, Stosic, and de Lima (this volume) note that in some languages the function of P_2 may be carried out by a case marker. This can be illustrated with a Romanian example:

- (30) în spatele tatălui meu
 in back-the father-the_{GEN} my
 ‘behind my father’

Here, although we have no overt P_2 , the complement *tatălui meu* is in the genitive case, thus functioning as the equivalent of, say, *de mon père* ‘of my father’ in French.

The majority of CPs in Breton (as in other languages) involve a nominal element, but a small number do not involve a noun either etymologically or synchronically. The majority of these are based on a verb form, as in the following:

- *abaoe* ‘since’ (< *a ba oe* ‘from when [it] was’)
- *panevet* ‘if not for’ (< *pa ne vez* ‘when it is not’)
- *da-heul* ‘following’ (< *da heul* ‘to follow’)

- *da-gaout* ‘towards’ (< *da gaout* ~ *gavout* ‘to find’)
- *daoust da* ‘despite’ (< *da c’houzout* ‘to know’)
- *gouez da* ‘according to’ (‘known to’)

In addition, a few etymologically complex prepositions involved foreign source material that probably ceased long ago to be analyzable to Breton speakers, including the following:

- *estreget* ‘besides, other than, but’ (< *estr eget*). The element *estr* is a reflex of the Latin borrowing *extra*, but it has no independent existence outside of this preposition (Deshayes 2003).
- *abalamour da* ‘because of’ (< French *par l’amour de* ‘out of love for’). This is often pronounced [blām] in spoken Breton, where the etymology is completely obscured.

Two of the above, *panevet* and *estreget*, have inflected forms:

- (31) Er c’houlz vad oun arru; **panevedon**, e oa maro
 in-the time good am arrived if-not-for_{1SG} PRT was dead
 breman Janig. (PG 206)
 now Janig
 ‘I arrived at the right time; if not for me, Janig would be dead now.’

All of the other CPs in the corpus have a noun, whether synchronically autonomous or not (about which more later), as their semantic nucleus. The most common patterns take one of the following two forms in which P_1 and P_2 are both simple prepositions:

AuxPrep Pattern (PNP): $[P_1 + \text{NOUN} + P_2]_{\text{PREP}}$
 Genitive Construction Pattern (PN): $[P_1 + \text{NOUN}]_{\text{PREP}}$

It should be noted that in colloquial speech and dialect, P_1 can sometimes be deleted, though it may resurface in vowel-initial forms of a paradigm.²⁰

Each of these two Breton patterns will now be discussed as to its internal structure and complementation patterns.

²⁰ In a long grammatical preface to his dictionary of the Tregor dialect, Le Dû (2012: 59–71) gives paradigms for a number of prepositions. Although he represents *e-kichen* ‘beside, near’ as *kichenn*, without the initial P_1 *e-*, the latter appears in vowel-initial forms, e.g. 3sg *n i gichenn* ‘near him’, *n i hichenn* ‘near her’ (cf. standard *en e gichen* and *en he c’hichen* respectively; the P_1 *e* ‘in’ becomes *en* before a vowel, reduced to [n] in Le Dû’s Tregerieg forms).

4.1 The Auxiliary Preposition (AuxPrep) Pattern (PNP)

The AuxPrep Pattern gives a CP having the canonical structure PNP, where the second P is what we are calling an auxiliary preposition or AuxPrep. The AuxPrep is generally one of the following four (each is followed by a list of some CPs which take it as their auxiliary):

- a) *da* ‘to’: *abalamour* ‘because of’, *a-dal* ‘across from’, *a-dreñv* ‘behind’, *a-dreuz* ‘across’, *a-enep* ‘against’, *a-us* ‘above’, *daoust* ‘despite’, *e-harz* ‘at the limit of’, *en-dro* ‘around’, *e-tal* ‘facing; next to’, *en abeg* ‘because of’, *e-biou* (*hebiou*) ‘past’, *en despet* ‘in spite of’, *en tu-mañ* ‘on this side of’, *en tu-se* ‘on that side of’, *en tuhont* ‘beyond’, *en tu all* ‘on the other side of’, *tost* ‘near, close to’, *hep gouzout* ‘unbeknownst to’
- b) *a* or *eus* ‘of, from’: *er-maez* ‘out of’, *war-c’hed* ‘in expectation of’
- c) *e(n)* ‘in’: *etrezek* ‘towards’, *etramek* ‘towards’, *war-du* ‘towards’, *e-barzh* ‘in’
- d) *gant* ‘with’: *a-stumm* ‘of the same shape as’, *a-du* ‘on the same side as’, *a-unan* ‘in unison with’, *a-gevret* ‘together with’, *asamblez* ‘together with’.

The AuxPrep pattern can be illustrated with *en abeg da* ‘because of’ (‘in reason to’) and *en-dro da* ‘around’ (‘in turn to’):

- (32) Mantret e tlee bezañ eveldomp-ni **en abeg d’**
 heartbroken PRT must be_{VN} for_{1PL-1PL} in reason to
 ar gwalleur c’hoarvezet diouzh ar beure; (ATP 76)
 the misfortune happened_{PAP} from-at the morning
 ‘He must have been heartbroken for us *because of* the misfortune that
 happened in the morning.’
- (33) Savet en deus ar gwaz kolier e chupenn
 raised_{PAP} to-him is the man collar his jacket
en-dro d’ e c’houzoug (WTL 6)
 in-turn to his neck
 ‘The man raised the collar of his jacket *around* his neck.’

Several prepositions in the pattern PNP involve nouns borrowed from French, thus appearing to replicate French CPs, e.g. *a-gaozenn da* ‘because of’ (cf. French *à cause de*), *en desped da* ‘in spite of’ (cf. French *en dépit de*) and *e-fas da* ‘across from’ (cf. French *en face de*).

Prepositions belonging to the PNP pattern can take pronominal complements straightforwardly, since the AuxPrep in every case is one that inflects:

- (34) Teñval-sac'h e oa **e-dro din** (VZB 88)
 dark-sack PRT was_{IMP} in-turn to_{1SG}
 'It was pitch black *all around me*.'
- (35) un daolenn istribilhet **e-fas dezhañ**, ouzh ar voger. (A-BMG 101)
 a painting hung_{PAP} in-face to_{3SGM} against the wall
 'a painting hanging *across from him*, on the wall.'
- (36) sanket en dienez hep spi ebet
 plunged_{PAP} in-the misery without hope in-world
 da zont **er -maez anezhi**. (Deizlevr 64)
 to come_{VN} to-the field of_{3SGF}
 'plunged into misery without any hope of getting out of it.'

The Breton AuxPrep pattern thus corresponds quite directly with the canonical CP described by Fagard et al. (this volume).

We turn now to the second and more complex pattern in which the relational noun and the complement of the CP instantiate the genitive construction.

4.2 The genitive construction pattern (PN)

In a highly productive pan-Celtic pattern, the relational noun of the CP forms a genitive construction with the complement that follows:

- (37) **e- kreiz** ar c'hoad (SL 21)
 in middle the wood
 'in the middle of the wood.'
- (38) Paotred, ha graet he deus Bro-C'hall he dever
 boys INT done_{PAP} 3SGF is France her duty
e- keñver ar Vretoned? (IVG 143)
 in regard the Bretons
 'Boys, has France done her duty *towards (with regard to) the Bretons?*'

- (39) Chomet eo digor-bras an Telegramme **e-kichen** e asied; (WTL 10)
 remained_{PAP} is open-big the Telegramme in-side his plate
 ‘The copy of *Le Télégramme* was wide open *beside* his plate.’
- (40) Bet em eus ul lizher **a-berzh** Yann ar Gall. (Deizlevr 24)
 been_{PAP} 1SG is a letter from-behalf Yann Ar Gall
 ‘I received a letter *from (on the part of)* Yann Ar Gall.’
- (41) ur c’hwezenn sklaset o tiverañ a-hed
 a sweat iced_{PAP} PRT drip_{VN} of-length
 va c’hein (ATP 28)
 my back
 ‘An icy sweat dripping along my back.’
- (42) Stag e oa gwirionez ma bugaleaj **war-bouez**
 attached PRT WAS_{IMP} truth my childhood on-weight
 ur fiseleenn vresk (WTL 179)
 a thread fragile
 ‘The truth of my childhood was attached *by means of* a fragile thread...’

Let us look more closely at the example of *e-kreiz* ‘in the center of’ in (37). At one level we seem to have the structure *e-kreiz* + *ar c’hoad* (i.e. $[[P_1 + \text{NOUN}]_{\text{CP}} + [\text{NP}]_{\text{PP}}]$) where *e-kreiz* ‘in the center of’ is a CP. At another level we have the structure *e* + *kreiz ar c’hoad* (i.e. $[[P_1] + [\text{NOUN} + \text{NP}]_{\text{GEN}}]_{\text{PP}}$), where the relational noun *kreiz* plus its noun complement together form an instance of the genitive construction (*kreiz ar c’hoad* ‘the center of the wood’); this is preceded by the simple preposition (P_1) *e* ‘in’.

The validity of the second analysis is clear historically. First of all, in the Goidelic languages which still have remnants of a morphological case system, the noun complement is overtly marked for the genitive case. In the following Scottish Gaelic examples, we see the nominative forms for ‘the cat’ and ‘the boy’ in (43), whereas in (44), after the CPs *ri taobh* ‘beside’ and *às aonais* ‘without’ we get the genitive forms of these nouns:

- (43) a. an cat
 the cat
 b. am balach
 the boy

- (44) a. ri taobh a' chait
 by side the_{GEN} cat_{GEN}
 b. às aonais a' bhalaich
 without the_{GEN} boy_{GEN}

But even in the Brythonic languages which no longer have morphological case, a pronominal complement of a CP takes the form of a possessive clitic. This shows that the relationship of the relational noun to its complement is construed as possessive; compare *e-kreiz ar c'hoad* 'in the middle of the wood' with *em c'hreiz* 'inside me' (lit. 'in my center'):

- (45) Tridal a ra va c'halon **em c'hreiz**
 pound_{VN} PRT does my heart in-my center
 ouzh ho klevet, va mab... (EK 23)²¹
 at 2PL hear_{VN} my son
 'My heart pounds *inside me* upon hearing you, my son...'

Thirdly, the syntactic behavior of the relational noun and its complement with respect to definiteness marking reveals that the two nouns instantiate the genitive construction. Recall that the possessed noun in the genitive construction (*kreiz*, in this example) is inherently definite, and cannot be overtly marked with any determiner. Thus, the latter is ungrammatical:

- (46) *er c'hreiz ar c'hoad
 in-the center the wood

But when *e-kreiz* has no expressed complement, *kreiz* requires the definite article in order to be interpreted as definite, thus giving *er c'hreiz* 'in the middle'²²:

- (47) a-zioc'h un daolig -varbr gwenn,
 above a table_{DIM} -marble white
 gant ur gibellig **er c'hreiz** (SL 119)
 with a basin in-the center
 'above a small white marble table with a basin *in the center*'

²¹ In the 1sg and 2sg there is a special enclitic form of the possessive, *-m* for 1sg and *-z* for 2sg which is normally selected after a vowel, although some variation is attested.

²² The definite article has three allomorphs in Breton (*ar*, *an*, *al*), depending on the initial sound of the following word. Each of these contracts with the preposition *e* 'in', giving *er*, *en*, and *el*, and any ICMs that normally apply after articles also apply after these contractions. In this case, we see that the initial <k> of a masculine singular noun obligatorily changes to <c'h> after the definite article.

As mentioned above, when the complement of a PN preposition is a personal pronoun, the latter takes the form of a possessive clitic, giving the word order P + POSS + N. In other words, the possessive is inserted between the two parts of a CP, whence the occasional name “fractional preposition” (e.g. Wmffre 1998). Those possessives which regularly trigger an ICM on their complement have the same effect on the relational noun in a CP (cf. the nouns *tro* and *keñver*):

(48) Galvet e voe ar medisin **war ma zro** (EST 194)
 called_{PAP} PRT was_{PRET} the doctor on my turn
 ‘The doctor was called to tend to me (lit. *around me*).’

(49) Pemp ki kounnaret a lampe hag a zailhe **war**
 five dog enraged_{PAP} PRT jumped_{3SG-IMP} and PRT leapt_{IMP} on
e dro... (PG 194)
 his turn
 ‘Five enraged dogs were jumping and leaping *around him*...’

(50) Muioc’h eget mat eo bet hennezh **en da geñver...** (SL 146)
 more than good is been_{PAP} this-one_M in your regard
 ‘He has been more than good *to you*.’

(51) Piv oac’h- c’hwi **en he c’heñver?** Hec’h amouros? (A-BGM 140)
 Who were_{2PL-IMP} you in her regard her lover
 ‘Who were you *to her*? Her lover?’

This results in paradigms like the following:

	war-dro ‘around’	e-keñver ‘in regard to, for’
1sg	war ma zro ‘around me’	em c’heñver ‘for me’
2sg	war da dro ‘around you’	ez keñver (~en da geñver) ‘for you’
3sg-m	war e dro ‘around him’	en e geñver ‘for him’
3sg-f	war he zro ‘around her’	en he c’heñver ‘for her’
1pl	war hon tro ‘around us’	en hor c’heñver ‘for us’
2pl	war ho tro ‘around you (pl.)’	en ho keñver ‘for you (pl.)’
3pl	war o zro ‘around them’	en o c’heñver ‘for them’

Not all CPs of the PN type take a possessive clitic. In some cases, an AuxPrep is used before a pronominal complement. Some CPs take an AuxPrep before either a lexical or a pronominal complement, as with *en-dro da* ‘around’:

(52) en-dro d’ an ti
in-turn to the house
‘around the house’

(53) en-dro dezhañ
in-turn to_{3SGM}
‘around it’

Others take the AuxPrep only before a pronominal complement:

(54) e-barzh an ti
in-part the house
‘in the house’

(55) e-barzh ennañ
in-part in_{3SGM}
‘inside it’

Now we will make a few more observations about the different parts of the CP.

4.3 The introductory preposition or P₁

We mentioned earlier that the preposition *a* ‘of, from’ is defective in not having its own paradigm of inflected forms. Similarly, when it is the simple preposition (P₁) that precedes the noun, a different preposition, variously *e(n)* ‘in’, *war* ‘on’ or *eus* ‘from’, replaces it before a pronominal clitic:

(56) ha n’ on ket den da dec’het **a-raok** va skeud. (EK 22)
and NEG am not person to flee_{VN} before my shadow
‘and I am not one to flee *from* (lit. ‘before’) my shadow’

(57) Mes eun all a dlee digoueout **en hon rôk**. (PG 159)
but one other PRT should arrive_{VN} in our fore
‘But another was to arrive *before* us.’

- (58) den ne ouie mann **a-zivout** ar Pembroke. (ATP 19)
 person NEG knew_{IMP} nothing of-subject the Pembroke
 ‘no one knew anything *about* the Pembroke’
- (59) Gwelout a reomp [...] e oa faziet ar barner
 see_{VN} PRT do_{1PL} [...] PRT was_{3SG-IMP} mistaken_{PAP} the judge
war da zivout... (IVG 251)
 on your subject
 ‘We see [...] that the judge was mistaken *about you*.’
- (60) degaset e voe din **a-berzh** Porfirio Diaz ur montr arc’hant (MZB 99)
 sent_{PAP} PRT was_{PRET} to_{1SG} on-behalf Porfirio Diaz a watch silver
 ‘a silver watch was sent to me *from* (lit. ‘on behalf of’) Porfirio Diaz’
- (61) Ha lavarit **eus va ferzh** d’ an deskarded all... (MZB 86)
 and speak_{2PL} from my behalf to the pupils other
 ‘and speak on my behalf to the other pupils’

4.4 The relational noun

The relational noun (N) in a CP of the PN type functions as the possessee in a genitive construction. This is so whether or not the N actually exists autonomously in the language (or ever did). Breton CPs can be arranged along a continuum with respect to the status of the N element:

- a) those where the N still exists autonomously and has more or less the same meaning as in the CP (e.g. *e-kreiz* ‘in the center of’, cf. *kreiz* ‘center’)
- b) those where the N still exists autonomously but has a non-literal meaning in the CP (e.g. *e-maez* ‘out’, cf. *maez* ‘countryside’)
- c) those where the N element must have historically been an autonomous N but which now no longer exists outside of the CP (e.g. *a-enep* ‘against’ cf. Welsh *wyneb* ‘face’; *war-lerc’h* ‘after’, cf. Welsh *llwrw* ‘track, trail’)
- d) those where the N element was not originally a noun at all, but has emerged from the reanalysis of a monomorphemic preposition as a PN (e.g. *etre* ‘between’ and *hebiou* ‘past’, analyzed by some speakers now as *e-tre* and *e-biou*).

In some complex adpositions, the relational noun can still function in more or less noun-like ways. For instance, it may be reduplicated (62), take a diminutive suffix (63), or even both of those at the same time (64):

- (62) er foñs emaint, **e-kichen-kichen** evel div durzhunellig... (WTL 114)
 in-the back are_{3PL} in side-side like two turtledove
 ‘they are at the back, *right beside one another* like two turtledoves’
- (63) da sankañ adarre en e Intrépide, gant Paotr-Teo **en e gichenig**. (SL 55)
 to sink_{VN} again in his Intrépide with Boy-Fat in his side_{DIM}
 ‘to sink back into his Intrépide, with Fatboy *right at his side*.’
- (64) gant un nozelenn gouevr velen lufrus
 with a knob copper yellow shiny
er c’hreizig -kreiz anezhi. (Hob 11)
 in-the center_{DIM} -center of_{3SGF}
 ‘with a shiny yellow brass knob *in the exact middle*.’

Also, the CP may be “remotivated” by replacing a weak P1 with a more robust one with fuller semantic content. The simple preposition *e* ‘in’ that occurs as the P1 in many CPs is no longer productive in many dialects, where reflexes of the CP *e-barzh* [ebars], often reduced to [ba], are generally used instead (65):

- (65) **Ba’ ma c’hichen** emañ Karine o sellet deus ar gartenn (WTL 211)
 in my side is Karine PRT look_{VN} at the map
 ‘Karine is *beside me* looking at the map.’

These facts about the relational noun in some CPs may suggest that they have a lesser degree of grammaticalization than is the case, for instance, with the relational nouns *côté* and *side* in the French and English translation equivalents *à côté de* and *beside*.

4.5 CPs and language variation

When examining the behavior of CPs in the Breton textual corpus, one of the things that stands out is the fairly high level of variation as to which complementation patterns a given preposition takes, sometimes even in the same text. To be sure, the descriptive literature does indicate that variation in complement structure is possible with some prepositions. For instance, Wmffre (1998: 30) notes a “tendency for the auxiliary inflection through *de* (i.e. *da---*kjr) to replace some of

the inflected prepositions” in the Central Breton dialect of Plounévezel that he is describing, giving the example of **vel'dit** varying with **vel'tas** ‘as you (sg.)’ (i.e. *evel dit* and *eveltez*, the latter a nonstandard form of *evelout* ‘like you’). Such comments are more often found in dialect monographs than in grammars of the standard language.

There is also considerable variation as to whether a given CP takes an AuxPrep before a lexical complement or not. Here we see both *a-dreuz* and *a-dreuz da* ‘across’ in the same text:

- (66) Neuze e welis eun den o tont
 then PRT saw_{1SG-PRET} a person PROG come_{VN}
a-dreuz d' al lanneg (PG 159)
 across to the moor
 ‘Then I saw a man coming across the moor.’

- (67) ar vandenn o vont, **a-dreuz** lanneier, *etrezeg* ar c'hoad. (PG 198)
 the band PROG come_{VN} across moors towards the wood
 ‘...the group coming, across moors, towards the woods.’

Similarly, the same phrase, ‘behind her back’, is attested in the corpus both with the AuxPrep *da* ‘to’ and without it:

- (68) Stagañ a reas he daouarn **a-dreñv d'** he c'hein. (A-BGM 113)
 attach_{VN} PRT did_{PRET} her hands behind to her back
 ‘He tied her hands behind her back.’
- (69) hag e tiskouez gant he groñj an estaj kentañ **a-dreñv** he c'hein. (WTL 37)
 and PRT show with her chin the floor first behind her back
 ‘and she motioned with her chin to the first floor behind her back.’

Along the same lines, there is a fair amount of variation with pronominal complements; some CPs are attested variously with an AuxPrep, and with a possessive clitic:

- (70) ul loen iskis ha spontus o tont **war -zu**
 an animal strange and scary PRT come_{VN} on -side
ennon. (EST 128)
 in_{1SG}
 ‘a strange and frightening animal coming towards me.’

- (71) Pa zeuis tre, e troas **war va zu**
 when came_{PRET.1SG} through PRT turned_{3SG-PRET} on my side
 e zaoulagad diverv (ATP 76)
 his two-eye lifeless
 ‘When I came in, he turned his lifeless eyes *towards me*.’

In a few cases there is even a third possibility. Some CPs come to function like simplex prepositions, taking personal inflections directly. This is attested, for instance, with *a-enep* ‘against’, which as a result is found in all three of the following complementation patterns:

- Pattern 1: en da enep
 in your against
 Pattern 2: a-enep dit
 against to_{2SG}
 Pattern 3: enepdout
 against_{2SG}

Examples from the corpus showing this variation are given below:

- (72) Droug a zo ennan **en e eneb** e- unan; (PG 147)
 bad PRT is in_{3SGM} in his against his self
 ‘He is angry at himself (lit. *against himself*).’
- (73) hag evit sevel klemm **a-enep dezhañ** ivez. (An avel 47)
 and for raise_{VN} complaint in-against to_{3SGM} also
 ‘and also to lodge a complaint against him.’
- (74) ar Republik a zifenn reiñ golo d’ ar veleien
 the Republic PRT forbids give_{VN} cover to the priests
 pennoù -fall a ya **enepti** (EK II, 131)
 heads -bad PRT go against_{3SGF}
 ‘The Republic forbids anyone to provide cover to insurgent priests who go against it.’

Of the three patterns attested with *a-enep* ‘against’, Kervella (1995: 337) and Hemon (1984: 110) mention only the first two. Hemon notes that only the possessive clitic pattern (Pattern 1) was found with *a-enep* in Middle Breton; the AuxPrep pattern (Pattern 2) emerged in the modern language. Pattern 3, in which the preposition is directly inflected as though it were a simplex preposition, is clearly the

newest; it is the least well attested (among the grammars, only Favereau [1997: 419] alludes to it), and it clearly presupposes a complete cycle of grammaticalization from morphologically complex to simplex preposition.

In summary, what we find is that some CPs are attested in the corpus in only one pattern (e.g. *diwar-benn* ‘about, concerning’, *en-dro da* ‘around’), some in two (e.g. *a-dreñv* ‘behind’), and some in all three (e.g. *a-enep* ‘against’):

CP	Pattern 1: Possessive clitic	Pattern 2: Auxiliary preposition	Pattern 3: Inflection like simple preposition
<i>diwar-benn</i> ‘about’	<i>diwar o fenn</i> from-on their head ‘about them’	*	*
<i>en-dro da</i> ‘around’	*	<i>en-dro dezho</i> ‘around them’	*
<i>a-dreñv</i> ‘behind’	<i>en o zreñv</i> in their behind ‘behind them’	<i>a-dreñv dezho</i> of behind to _{3pl} ‘behind them’	*
<i>a-enep</i> ‘against’	<i>en ma enep</i> in my “face” ‘against me’	<i>enep din</i> “face” to _{1sg} ‘against me’	<i>enepton</i> “face” _{1sg} ‘against me’

It is quite likely that a dialect survey would expand the range of attested patterns for particular prepositions. For example, although it is unattested in the written corpus used here, there are spoken dialects which have inflected forms for *diwar-benn* ‘about, concerning’, e.g. *diwarbennon* ‘about me’ (Favereau 1997: 423).

5 Permutations to express movement

The corpus reveals that the shape of many Breton CPs that express static location can be altered to express movement, whether *ablative* (movement away from) or *allative* (movement toward). This semantic complexity is virtually never made explicit in the descriptive sources except for Trépos (1966: 145-47), who gives a few relevant examples. Such permutations can be achieved either by adding an autonomous preposition, or by replacing the simple P1 (usually *e* ‘in’ or *a* ‘of, from’) with another that communicates movement more explicitly (generally *eus* ‘from’ or *dre* ‘through’). These patterns are not standardized; the

corpus reveals great variation. For instance, let us consider *a-dreñv* ‘behind’. To create the ablative meaning ‘from behind’ we find five permutations in the corpus: *a-ziadreñv*, *dre a-dreñv*, *eus a-dreñv*, *eus dreñv*, and *diouzh dreñv*.²³ Two of these are shown below:

- (75) Ha dav! da geitañ war ar goied, **eus dreñv** e voger. (SL II, 59)
 and pow to aim_{VN} on the geese from behind his wall
 ‘And pow! Taking aim at the geese from behind his wall.’
- (76) Soñjal a reas an aotrou Broc’h e oant
 think_{VN} PRT do_{3SG-PRET} the mister Badger PRT be_{3PL-IMP}
 target dre a-dreñv... (An avel 249)
 attack_{PAP} through behind
 ‘The Badger thought they were being attacked from behind...’

It turns out that a number of complex Breton prepositions of the type PN or PNP allow a change of P1 in order to alter the basic meaning of the preposition from locative to directional. Replacing the static preposition *e-* ‘in’ with *a-* ‘from’ is common to give an ablative sense. Note as well the lenition triggered by *a* (*diabarzh* becomes *ziabarzh*):

- (77) Mont a ra **e- diabarzh** ar gar... (WTL 5)
 go_{VN} PRT does in- inside the station
 ‘He goes *inside* the train station...’
- (78) --Piou ’zo aze? eme eur vouez garo, **a- ziabarz**. (PG 188)
 who is there said a voice rough from- inside
 ‘‘Who’s there?’’ said a rough voice *from inside*.’

²³ All five permutations use the lenited form *dreñv* rather than *treñv*, which is to be expected in the first three where the word occurs after *a* ‘of, from’ that causes lenition. However, it is not expected after *eus* and *diouzh*, which do not trigger a mutation. But *treñv* belongs to our category (c) where the relational noun used to exist autonomously (cf. Welsh *tref* ‘home; town’) but no longer does, or only marginally, in Modern Breton. In other words, contemporary speakers may not make any connection with the obsolete noun *treñv* and may have generalized *dreñv* as a formative or pseudo-noun for use in all permutations of this preposition.

Other examples include *e-mesk* and *e-touez*, both meaning ‘among, in the midst of’, which give *a-vesk* and *a-douez* ‘from among’. In addition, this pattern is often extended to *(h)ebiou* ‘past’ and *etre* ‘between’ which, though etymologically simple in Breton, have been analyzed (at least in some areas) as *e-biou* and *e-tre*, by analogy with CPs like *e-mesk* and *e-touez*. Thus, ‘in someone’s hands’, literally ‘between someone’s hands’, is *etre daouarn* ‘in the hands of’, while ‘from someone’s hands’ is literally ‘from-between’ (*a-dre* or even *eus a-dre*) someone’s hands:

- (79) peur e kouezhas al levr **eus a-dre** e zauarn. (A-BGM 161)
 when PRT fell_{3SG-PRET} the book from from-between his hands
 ‘when the book fell *from his hands*’

Those CPs where the P1 is already *a-* can replace it with *dre* ‘through’ to introduce the notion of movement (e.g. *a-dreuz* ‘across’ becoming *dre dreuz*, literally ‘through across’):

- (80) ha me ha troc’hañ berr **dre dreuz** ur foenneg (SL II, 168)
 and 1SG and cut_{VN} short through across a hayfield
 ‘and I took a shortcut *across a hay field*’

An alternate formulation, *a-dreuz dre* (literally ‘across through’) occurs in (81):

- (81) “Deomp dre amañ, dre an tu dehou, eme ar
 come_{1PL} through here through the side right said the
 c’here, **a-dreuz dre** ar c’hoad, evit mired
 cobbler across through the wood for keep_{VN}
 na vezimp gwelet...” (EK II, 113)
 that_{NEG} be_{1PL-FUT} see_{PAP}
 ‘Let us come through here, through the right side,’ said the cobbler,
 ‘*across through the wood*, to keep from being seen.’

The shapes of several locative CPs thus emerge as rather fluid in the corpus, since they can be altered fairly freely to express movement.

We turn lastly to the most interesting set of CPs involved in the expression of movement, those that take the prefix *di-*.

Three very common prepositions result from adding the prefix *di*, meaning ‘from’, to a simple preposition, thus giving ablatives or expressions of movement away:

- (82) Seiz a oa stlapet **diwar** o c’hezeg (PG 169)
 seven PRT were_{3SG-IMP} dragged_{PAP} from-on their horses
 ‘Seven of them were dragged *from* (lit. from-on) their horses’
- (83) a glaske tennañ hec’h arm **digant**
 PRT tried_{3SG-IMP} pull_{VN} her weapon from-with
 ar gozhiadez kounnaret. (IVG 256)
 the old-woman angry
 ‘... tried to take the weapon *from* (lit. from-with) the angry old woman.’
- (84) P’ oa savet **diouz** an daol... (Jezegou 66)
 when was_{IMP} risen_{PAP} from-at the table
 ‘When he rose from (lit. from-at) the table’

The unprefixed forms and the *di*-prefixed forms function together as a contrasting pair. For instance, if one is on (*war*) a horse, one can be dragged off of it (*diwar*, literally ‘from-on’). If someone has something in her possession it is said to be ‘with’ (*gant*) her, and if it is taken away, it is taken *digant* (lit. ‘from with’) her. And one who is seated ‘at the table’ (*ouzh an daol*) must eventually rise from-at the table (*diouzh an daol*). *Ouzh* and *diouzh* are also used contrastively to introduce complements for a number of antonymous adjectives and verbs, e.g. *tost ouzh* ‘close to, near’ versus *pell diouzh* ‘far from,’ *stag ouzh* ‘attached to’ versus *distag diouzh* ‘unattached from’, and *heñvel ouzh* ‘similar to’ versus *dishañvel diouzh* ‘different from’.²⁴

All three of these prepositions (*diwar*, *digant*, *diouzh*) end up corresponding to ‘from’ in English, but with nuances that give information about the relative position of the thing in question before the movement happened.²⁵ Of the three,

²⁴ Russell (2011), following Sommerfelt (1957), argues that a strong case can be made that this particular type of complex preposition by prefixation is due to the contact of Brythonic (Brittonic) Celtic with Latin during the Roman occupation of Britain. His claim is based on several criteria, including the fact that this type of complex preposition is absent from the Goidelic languages but attested in all of the Brythonic languages; it was demonstrably present in Latin at the relevant time; there is no particular evidence for it to be a purely internal development in Brythonic, etc. Without taking a firm stand, we agree that the external explanation appears quite reasonable in this case.

²⁵ In Welsh, the cognate is a separate word, *oddi*, that has functions similar to the Breton prefix *di*-. Watkins (1993: 334) notes the “notable degree of precision” associated with the choice of a

diouzh is the most common and it becomes more or less the usual way to say ‘from’ in many contexts²⁶:

- (85) Hag e tennas un dornadig pezhiau **diouzh**
 and PRT draw_{3SG-PRET} a handful_{DIM} coins from-at
 e chakod. (SL 20)
 his pocket
 ‘And he pulled a small handful of coins *from* his pocket.’

The tendency is clearly for the *di*-prefixed forms to lose their ablative meaning and to be interpreted as merely locative. In two cases, the semantic bleaching of *di*- is already complete; the result is no longer perceived as a compound, and the ablative reading is no longer available. Thus, *dindan* (or *didan* in some dialects; historically *di*- + *dan* ‘under’, thus ‘from-under’) is today the normal word for ‘under’, occurring even when no movement away is intended:

- (86) nemet ur preñv mat da vezañ moustret **dindan** dreid. (ATP 96)
 only a worm good to be_{VN} crushed_{PAP} under feet
 ‘only a worm good to be crushed under foot.’

Contemporary Breton speakers do not recognize *di(n)dan* as a historically bimorphemic form; it is monomorphemic today. Similarly, *dirak* ‘before, in front of’ has almost entirely supplanted the unprefixated form *rak* in its spatial uses:

- (87) Chomet eo Pierre Goer a-sav **dirak** dor ar burev. (WTL 56)
 stayed_{PAP} is Pierre Goer on-stand from-before door the office
 ‘Pierre Goer stopped before the office door.’

word for ‘from’, e.g. *oddi ar y silff* ‘from the shelf’ (lit. ‘from on the shelf’), *oddi wrth y bachgen* ‘from the boy’ (lit. ‘from at the boy’), *oddi am fy ngwddf* ‘from my neck’ (lit. ‘from around my neck’). He goes on to add: “Many of these distinctive nuances are being lost in the present fully bilingual state in which Welsh finds itself.”

26 In a number of Breton dialect areas, the prepositions *ouzh* ‘at, against’, *diouzh* ‘from-at, from-against’ and *eus* ‘from’ fall together into a single preposition variously pronounced [dø̃s] or [doh], e.g. the Kerneveg dialect described by Goyat (2012: 260). Such variants are rarely written and not generally taught, though there are exceptions. Davalan (2001) includes the form *deuzh* [dø̃s] in his *Méthode de breton*, though he flags it with the clear warning: “To be carefully avoided in written Breton, as well as in formal speech” (“À soigneusement éviter en breton écrit, ainsi qu’à l’oral dans un style soutenu”) (p. 251). Kervella (1995: 336) mentions the additional dialect forms *deus*, *dimeus*, *demeus*.

The unprefixd *rak* still exists but it is mostly limited to experiencing fear ‘before, in the presence of’ something:

- (88) Ober a ris eveltañ gant aon **rak**
 do_{VN} PRT did_{1SG-PRET} like_{3SGM} with fear before
 kreskiñ e nec’hamant. (ATP 85)
 increase_{VN} his anxiety
 ‘I did the same as him for fear of increasing his anxiety.’

It happens that turning ablative prepositions into locative ones is not uncommon in the languages of the world. The process has been called ablative-locative transfer (e.g. Mackenzie 1978, Luraghi, Nikitina and Zanchi 2017). Luraghi et al. illustrate with the following example from Latin and Romance:

- (89) earlyLatin *in-tus* ‘from inside’ > Classical Latin *intus* ‘inside’ (static expression) > Vulgar Latin *de intus* ‘from inside’ > Old French *denz* ‘inside’ > Old French *de + denz* ‘from inside’ > later Old, Middle and Modern French *dedans* ‘inside’ > *de dedans* ‘from inside’.
 (Luraghi, Nikitina and Zanchi 2017)

In this remarkable example, every time the ablative sense was lost, a new morpheme was added when that sense was needed again; the process was cyclical, with the ablative sense being repeatedly lost and replaced. Similarly, with the formerly ablative Breton prepositions *dindan*, *dirak*, and *diwar-dro*, we see the meanings ‘from under’, ‘from in front of’ and ‘from around’ being reestablished in (90) and (91) by the addition of the preposition *a* ‘of, from’ (which causes lenition to *zirak*, *ziwar-dro* and *zindan*, respectively):

- (90) ne oa ket aet kuit c’hoazh ar bolised **a-zirak** hag
 NEG was_{IMP} not gone_{PAP} away yet the police from-before and
a-ziwar-dro Ti ar Bredoù. (Deizlevr 74)
 from-around House the Bredoù
 ‘... the police had not yet left *from in front of* and *from around* the Bredoù’s house.’

- (91) **A-** **zindan** e gasketenn lêr e tiboufe kizennoù
 from- under his cap leather PRT poked_{3SG-IMP} tufts
 melen a bep tu d’ e dal (ATP 33)
 yellow of each side to his brow
 ‘*From under* his leather cap blond tufts of hair poked out on both sides of his forehead.’

But like the Latin-Romance example cited above, the ablative sense can be bleached once again, continuing the cycle. Thus, in the following example, *a-zirak* simply means ‘before, in front of’. The sense ‘from’ is clearly absent:

- (92) setu ma teuas daou soudard [...] da lakat o zokoù ruz
 behold CONJ come_{3GS-PRET} two soldier [...] to put_{VN} their hat_{PL} red
 flaerius **a-zirak** hor fri, o c’houlen
 fetid from-before our nose PRT ask_{VN}
 arc’hant... (Ar Floc’h 52)
 money
 ‘...and then two soldiers [...] came to put their fetid red hats *before our noses*, asking for money’

The following examples, from a single work, show *a-zindan* meaning simply ‘under’ in (93), such that to achieve the ablative meaning ‘from under’ requires an additional element. We find this in the combination *eus a-zindan* in (94):

- (93) Chom a rejont ur pennad da sellet ouzh an toull
 stay_{VN} PRT did_{3PL-PRET} a moment to look_{VN} at the hole
 m’ oa kouezhet ennañ Paolig hag an daou soudard,
 REL was fallen_{PAP} in_{3SGM} Pauly and the two soldier
 da c’houzout a dont a raje unan bennak anezho
 to know_{VN} PRT come_{VN} PRT do_{3SG-COND} one some of_{3PL}
 war- gorre an dour. Mes aet oant **a-zindan**
 on surface the water but gone_{PAP} were_{3PL-IMP} from-under
 ar skornenn, e feson, rak hini anezho ne
 the ice in appearance for neither of_{3PL} NEG
 deuas er-maez. (EK II, 119)
 came_{3SG-PRET} out

‘They stayed a moment to look at the hole that Pauly and the two soldiers had fallen into, to know whether anyone would come to the surface of the water. But they had gone *under the ice*, it appeared, for not one of them came out.’

- (94) Pa voe tennet ar baotrezig **eus a-zindan** ar
 when was_{3SG-PRET} drawn_{PAP} the girl from from-under the
 gwele, e- lec’h m’ oa kuzhet... (EK II, 107)
 bed in- place REL was hidden_{PAP}
 ‘When the little girl was pulled *from under the bed* where she was hiding...’

Etymologically speaking, in the form *eus a-zindan* in (94), the sense ‘from’ is actually represented three times, in *eus*, *a-* and *zi-* (lenited form of *di-*) revealing the threefold repetition of the cycle of ablative-locative transfer.

For other writers, *a-zindan* and *eus a-zindan* may be more or less synonymous in giving the meaning ‘from under’, *eus* simply serving to reinforce the ablative sense that *a-* can still have:

- (95) da dennan teil **a- zindan** ar zaoud (PG 84)
 to pull_{VN} manure from- under the cattle
 ‘to pull manure *from under the cattle*’
- (96) Souden, **eus a-zindan** eur bod skô, hag hen o welet
 suddenly from from-under a grove elders and he PRT see_{VN}
 o paran warnan daoulagad eur c’haz (PG 136)
 PRT gaze_{VN} on_{3SGM} eyes_{DUAL} a cat
 ‘Suddenly, he saw the eyes of a cat staring at him *from under a grove of elder trees*.’

We will close with the case of *digant* ‘from-with’. One of its most frequent uses is in the collocation *goulenn digant* ‘to ask (something) of (lit. from-with) someone’. Although this expression is often used when what is being asked for is an actual physical object, like alms in (97) – in which case there is reference to a literal movement of the requested object from giver to receiver – the same construction is used by metonymy when what is being asked for is simply a piece of information, as in (98), in which case there is no literal movement at all:

- (97) tri lampon o stouvañ an hent deomp hag
 three ruffian PRT block_{VN} the way to_{1PL} and
 o c’houlenn **diganeomp** aluzenn ur bilhed bennak. (WTL 38)
 PRT ask_{VN} from-with_{1PL} alm a bill some
 ‘three tough guys blocking our way and asking alms of a banknote from us.’
- (98) Goulenn a reas **diganin** ur bern enklaskoù
 ask_{VN} PRT do_{3SG-PRET} from-with_{1SG} a heap inquiries
 diwar -benn ar varzhed (Taldir 271)
 from-on -head the bards
 ‘He asked me a lot of questions about the poets.’

Perhaps this frequent metonymy contributed slightly to the felt need to reinforce *digant* ‘from-with’ (to *a-zigant* ‘from from-with’) in an example like (99), where clearly the literal ablative sense is desired:

- (99) Ma gwreg zo aet kuit **a- ziganin** e-keit ma
 my wife is gone_{PAP} away from from-with_{1SG} while CONJ
 oan e Mururoa. (WTL 9)
 was_{1SG-IMP} in Mururoa
 ‘My wife left me while I was in Mururoa’

The corpus data on the *di*-prefixed forms thus reveals that the semantic bleaching of the etymological ablative sense of this prefix has gone further, at least in some varieties of Breton and in some works, than is recognized in the standard language. The data nicely reflect the cyclical nature of ablative-locative transfer that has been found in other languages such as the development of Romance from Latin.

6 Conclusion

In this essay we have explored the complex prepositions of the Celtic language Breton. Because this language has only a recent standard (developed mostly in the early 20th century) which has had little influence on the traditional spoken dialects, the study of Breton vernaculars can reveal linguistic processes that were able to develop largely unfettered by prescriptivist intervention. This study, based on a corpus of around thirty Breton texts from the late 19th to the early 21st centuries from various parts of Brittany and reflecting various degrees of contact with the standard language, has revealed interesting patterns of variation along the following lines:

- We found significant variation in complementation patterns going beyond what descriptive grammars would lead one to expect. Three complementation patterns are attested: Pattern 1 expresses a pronominal complement as a possessive clitic (*en ma enep* ‘against me’), Pattern 2 makes use of an AuxPrep which is inflected (*a-enep din*), and Pattern 3 treats the etymologically complex preposition as a simplex preposition, allowing it to be inflected directly (*enepton*). Different CPs are attested in one, two, or even all three of these complementation patterns.
- Several CPs with static semantic interpretations allow permutations of their P1 in order to express movement, e.g. *e-touez* and *e-mesk* ‘among’ becoming

a-douez and *a-vesk* ‘from among’ and *dre douez*, *dre vesk*, literally ‘through among’. Two prepositions that are etymologically simplex, *etre* ‘between’ and *(h)ebiou* ‘past’ have become integrated into this pattern, thus leading to the creation of forms like *a-dre* and *dre-biou*.

- We also encountered an interesting set of data bearing on the process of ablative-locative transfer which, if it is widely found in the languages of the world, is still not commonly documented in the linguistic literature. The Breton data provide clear examples of the cyclical nature of this process, with a complex form like *eus a-zindan* ‘from-under’ revealing the application of ablative-locative transfer three consecutive times.

Abbreviations

3SGF	third singular feminine	P	preposition
3SGM	third singular masculine	PAP	past participle
DIM	diminutive	PL	plural
FUT	future	POSS	possessive
HAB	habitual	PRET	preterite
IMP	imperfect	PROG	progressive
INT	interrogative	PRT	particle
N	noun	REL	relative
NEG	negative particle	VN	verbal noun

References

- Ball, Martin J. & James Fife (eds.). 1993. *The Celtic Languages*. London & New York: Routledge.
- Broderick, George. 1993. Manx. In Martin J. Ball & James Fife (eds.), *The Celtic Languages*, 228–286. London & New York: Routledge.
- Broudic, Fañch. 2009. Sociolinguistique historique d’une langue non-romane: le breton.” In Dorothee Aquino-Weber, Sara Cotelli & Andres Kristol (eds.), *Sociolinguistique historique du domaine gallo-roman*, 317–335. Bern: Peter Lang.
- Chalm, Eugène. 2008. *La grammaire bretonne pour tous*. Lannuon: An Alarc’h Embannadurioù.
- Davalan, Nikolaz. 2001. *Brezhoneg : Méthode de breton 2*. Rennes: Skol an Emsav.
- Desbordes, Yann. 1983. *Petite grammaire du breton moderne*. Mouladurioù hor Yezh.
- Deshayes, Albert. 2003. *Dictionnaire étymologique du breton*. Douarnenez: Le Chasse-marée.
- Fagard, Benjamin, José Pinto de Lima, Dejan Stosic & Elena Smirnova. This volume. Introduction: Complex Adpositions and Complex Nominal Relators.
- Fagard, Benjamin, Dejan Stosic & José Pinto de Lima. This volume. Complex adpositions in Romance: Emergence and variation.
- Favereau, Francis. 1997. *Grammaire du breton contemporain*. Morlaix: Skol Vreizh.

- George, Ken. 1993. Cornish. In Martin J. Ball & James Fife (eds.), *The Celtic Languages*, 410–468. London & New York: Routledge.
- Gillies, William. 1993. Scottish Gaelic. Cornish. In Martin J. Ball & James Fife (eds.), *The Celtic Languages*, 145–227. London & New York: Routledge.
- Goyat, Gilles. 2012. *Description morphosyntaxique du parler breton de Plouzévet*. Doctoral thesis, Université Rennes 2.
- Hardie, David F. W. 1948. *A Handbook of Modern Armorican (Breton)*. Cardiff: University of Wales Press.
- Hemon, Roparz. 1984. *A Historical Morphology and Syntax of Breton*. Dublin: The Dublin Institute for Advanced Studies.
- Hemon, Roparz. 1984. *Grammaire bretonne* (9th edn). Brest: Al Liamm. http://vk.com/doc-49400050_255285537?dl=7a48934fe48bf49187
- Hingant, Jean. 1868. *Éléments de la grammaire bretonne*. Tréguier: A. Le Flem.
- Jackson, Kenneth Hurlstone. 1967. *A Historical Phonology of Breton*. The Dublin Institute for Advanced Studies.
- Kergoat, Lukian. 1989. *Yezhadur*. Saint-Brieuc: Ti-Embann ar Skolioù (TES).
- Kerrain, Mark. 2010. *Le guide du bretonnant*. Pornic: An Amzer Embanner.
- Kervella, F. 1995. *Yezhadur bras ar brezhoneg*. Brest: Al Liamm.
- LeClerc, L'Abbé L. 1911. *Grammaire bretonne du dialecte de Tréguier (Deuxième éd. revue et corrigée)*. [First edn. 1908]. Saint-Brieuc: Imprimerie-Librairie de René Prud'homme.
- Le Dù, Jean. 2012. *Le trégorrois à Plougrescant : Dictionnaire breton-français*. Brest: Emgleo Breiz.
- Luraghi, Silvia, Tatiana Nikitina & Chiara Zanchi (eds.). 2017. Chapter 1: Introduction. *Space in diachrony*. Amsterdam & Philadelphia: John Benjamins.
- Macaulay, Donald. 2008. The Scottish Gaelic language. In Donald Macaulay (ed.), *The Celtic Languages*, 137–248. Cambridge: Cambridge University Press.
- MacKenzie, J. Lachlan. 1978. Ablative-Locative Transfers and Their Relevance for the Theory of Case-Grammar. *Journal of Linguistics* 14(2), 129–156
- Morris-Jones, John. 1913. *A Welsh Grammar: Historical and Comparative*. Oxford: Clarendon Press.
- Ploneis, Jean-Marie. 1983. *Le parler de Berrien*. Paris: SELAF.
- Press, Ian. 1986. *A Grammar of Modern Breton*. Berlin: Mouton de Gruyter.
- Russell, Paul. 2011. Latin and British in Roman and Post-Roman Britain: Methodology and morphology. *Transactions of the Philological Society* 109, 138–157.
- Sommerfelt, Alf. 1957. Some notes on the influence of Latin on the Insular Celtic Languages. *Acta Congressus Madvigiani. Proceedings of the Second International Congress of Classical Studies*, vol. 5. Copenhagen: Munksgaard, 157–162.
- Ternes, Elmar. 2008. The Breton language. In Donald Macaulay (ed.), *The Celtic Languages*, 371–452. Cambridge: Cambridge University Press.
- Trépos, Pierre. 1957. *Grammaire bretonne*. Rennes: Imprimerie Simon. (Rééditions 1962, 1980 Rennes: Ouest-France ; 1994, Brest: Brud Nevez.)
- Vallée, François. 1902. *Leçons élémentaires de grammaire bretonne*. Saint-Brieuc: Imprimerie Saint-Guillaume.
- Watkins, T. Arwyn. 1993. Welsh. In Martin J. Ball & James Fife (eds.), *The Celtic Languages*, 289–348. London & New York: Routledge.
- Wmfre, Iwan. 1998. *Central Breton*. München: Lincom Europa.

Breton corpus

- Abeozen, Fañch Eliès. 1980. *Pevar skourr ar Mabinogi*. Preder. **[Abeozen]**
- Al Lay, Fañch. 2003. *Bilzig*. Morlaix: Skol Vreizh. **[Bilzig]**
- Ar Floc'h, Loeiz. 1923. Merc'h ar mezvier. *Buhez Breiz* janvier 1923 no. 25, 415–417. **[Mezvier]**
- Ar Floc'h, Loeiz. 1985. *Va zamm buhez*. Lesneven: Mouladurioù Hor Yezh. **[Ar Floc'h]**
- Ar Gow, Yeun. 1978. *E skeud tour bras Sant Jermen*. Brest: Al Liamm. **[EST]**
- Ar Moal, Erwan. 1925. *Pipi Gonto*. Quimper: Le Goaziou. **[PG]**
- Beyer, Mich. 2010. *Kan ar mein*. Lannuon: Al Alarc'h Embannadurioù. **[Beyer]**
- Bijer, Yann. 2007. *Avel gornôg*. Al Liamm. **[Bijer]**
- Carroll, Lewis. 1995. *Troioù-kaer Alis e Bro ar Marzhoù*. (Breton translation by Herve Kerrain).
Quimper: An Here. **[Alis]**
- Drezen, Youenn. 1972–1973. *Skol-louarn Veig Trebern*. Lodenn gentañ / Eil lodenn. Al Liamm.
[SL I, II]
- Drezen, Youenn. 1977. *Itron Varia Garmez*. Al Liamm. **[IVG]**
- Dupuy, Yann-Fulup. 2005. *Penn ar veaj*. Keit vimp bev. **[Dupuy]**
- Duval, Anjela. 1993. *Tad-kozh Roperz-Huon (1822–1902)*. Hor Yezh. **[Duval]**
- Gerven, Yann. 2003. *War un ton laou*. Morlaix: Skol Vreizh. **[WTL]**
- Gerven, Yann. 2013. *Traoù nevez: Danevelloù*. Morlaix: Skol Vreizh. **[GerventN]**
- Glanddour, Maodez (ed.). 1982. *An Testamant Nevez I : Ar Pevar Aviel*. Al Liamm. **[Bible]**
- Grahame, Kenneth. 2014. *An Avel en Haleg*. (Breton translation by Alan Martel). Lannuon: An
Alarc'h Embannadurioù. **[An avel]**
- Huon, Ronan. 1980. *Ur vouezh er vorenn*. Brest: Al Liamm. **[Ur vouezh]**
- Inizan, Lan. 1977. *Emgann Kergidu*. Lodenn I. Brest: Al Liamm. **[EK I]**
- Inizan, Lan. 1977. *Emgann Kergidu*. Lodenn II. Brest: Al Liamm. **[EK II]**
- Jacq, Yann-Fanch. 2002. *A-berzh mamm-gozh*. Keit vimp bev. **[A-BMG]**
- Jezegou, Christophe. 1923. *E Korn an Oaled*. Quimper: J.-M. Guivarch. **[Jezegou]**
- Milin, Gabriel. 1923. *Ar pesketaer yaouank*. Part 1 published in *Buhez Breiz* June 1923 no. 30
pages 574–577. Part 2 published in *Buhez Breiz* July 1923 no. 31, 606–612. **[Milin]**
- Olier, Youenn. 1986. *Deizlevr Here 1945-Here 1946*. Keneiled Imbourc'h. **[Deizlevr]**
- Priel, Jarl. 1975. *Va zammig buhez*. Brest: Al Liamm. **[VZB]**
- Priel, Jarl. 2002. *An teirgwern Pembroke*. Plougastell-Daoulaz: An Here/Al Liamm. **[ATP]**
- Riou, Yvonne. 1995. *E Bro Yvonne : Eñvorioù ur beizantez a Vro-Leon : Levrenn genta*. Brest:
Emgleo Breiz. (Collected and transcribed by Jean Roparz). **[Yvonne]**
- Rowling, J. K. 2012. *Harry Potter ha Maen ar Furien*. (Breton translation by Mark Kerrain). Pornic:
An Amzer. **[HP]**
- Steinbeck, John. 2013. *A-ziwar Logod ha Tud*. (Breton translation by Kristian Braz). Mouladurioù
Hor Yezh. **[Logod]**
- Taldir-Jaffrennou, François. 1985. *Eñvorennoù*. Lesneven: Mouladurioù Hor Yezh. **[Taldir]**
- Tolkien, J.R.R. 1999. *An Hobbit, pe eno ha distro*. (Breton translation by Alan Dipode).
Argenteuil: Arda. **[Hob]**

Pietro Bortone

5 On complex adpositions in Modern Greek

Abstract: This article reviews Modern Greek word strings that could be considered adpositions, noting some arguments for or against such classification. In Modern Greek, there are many combinations of adverbs and adpositions which are functionally and semantically comparable to simple adpositions; and there are other sequences of words that could be regarded as more elaborate adpositions. We first look at [ADV + P] compounds like *píso apó* ‘behind’ or *anámesa se* ‘among’, and we consider criteria for granting or denying them adpositional status – criteria such as cross-linguistic parallels, unique combinatory meanings, and restricted paradigmatic and syntagmatic flexibility. We then consider, in a similar light, more complex kinds of phrases. These include: the type [[P + Det + N] + NPGEN], e.g. *katá ti diárcia* (+GEN) ‘during’; the type [[P1 + N + P2] + NPACC], e.g. *se sçési me* ‘concerning’; the type [[P + N] + NPGEN], e.g. *ðia mésu* +GEN ‘through’; the type [[P + N] + NPACC], e.g. *me théma* ‘about (i.e. on the topic of)’; and the type [[archaic NPDAT/ACC] + NPGEN], e.g. *lóyo* +GEN ‘because (of)’. In conclusion, we point out diachronic similarities and connections between Greek simple and complex adpositions, showing that even simple adpositions were originally complex in structure.

Keywords: Greek prepositions, compound prepositions, complex prepositions, grammaticalisation

1 Introduction

This article presents Modern Greek word strings that can arguably be classified as complex adpositions, and sets forth some arguments for or against such classification.¹ After noting parallels in the Modern Greek use of cases, simple adpositions, and combinations of adverbs and adpositions (§2.3), we list various types of Greek word strings that may constitute complex adpositions. We start (§3.1) from Greek combinations of adverbs and adpositions, the type represented in English by *out of* or *next to*, and we consider criteria for granting or denying them the

1 Many thanks to Peter Mackridgè for comments on an earlier draft of this article.

Pietro Bortone, University of Oxford, Faculty of Linguistics, Philology, & Phonetics, Oxford, England. pietro.bortone@ling-phil.ox.ac.uk

<https://doi.org/10.1515/9783110686647-006>

status of adpositions (§3.2). We then consider more complex phrases (§4), equivalent to English constructions like *in relation to x*, *by means of x*, *for the duration of x*, *with destination x*, or *via x*, and review reasons for granting or denying them adpositional status. Finally, we note diachronic similarities and links between simple and complex adpositions (§5).

2 Language description

2.1 The Modern Greek language

Modern Greek is usually considered the only living language in the Hellenic branch of Indo-European, although Tsakonian, Cappadocian, Griko, Cypriot Greek, and Pontic (Bortone 2009: 71–73) are claimed by some to be separate languages, rather than Greek dialects. The total number of mother-tongue speakers of Modern Greek is well over twelve million, of which about ten million are in today's Republic of Greece and one million in the Republic of Cyprus. Sizeable communities of Modern Greek speakers are also to be found in other parts of the Balkans, in the US, in Australia, in northern Europe, and in Turkey. Modern Greek is noticeably related to Classical Greek, more than contemporary Romance languages are to Latin. However, the impression of a strong similarity between Modern and Classical Greek is partly due to a conservative spelling (in a unique alphabet) and, above all, by the past influence of *Katharévoussa*, an artificial style of Greek imposed in Greece by law until 1976. This has revived elements of ancient phonology, morphology, syntax, and lexicon. It should also be noted that, since educated speakers can still occasionally draw on the archaising style, the distinction between productive and unproductive usage in Modern Greek is at times nebulous. In what follows, for reasons of space, reference is made only to the standard language.

2.2 Greek – major typological characteristics

Despite a mild drift towards analysis, Greek is, to this day, rather rich in inflections – both in its verbal conjugations and in its nominal morphology. It inflects not only nouns but also their adjectives and definite and indefinite articles – for gender, number, and case. In the modern language, genders are three,

numbers two, and cases, as traditionally classified, four.² The three genders, as in Ancient Greek, are known as masculine, feminine, and neuter; the numbers are the singular and the plural (the obsolescent dual of Ancient Greek has completely vanished); the cases now used productively are called nominative, genitive, accusative, and vocative (the Ancient Greek dative survives only in learned phrases; other Indo-European cases marginally attested in Ancient Greek have long disappeared). Modern Greek is principally SVO, nominative / accusative, and is similar to other European languages with respect to several structural features and, at times, in phraseology. This fact may be interpreted as being, to a degree, also the result of convergence. In particular, Modern Greek is part of the *Balkan Sprachbund*, and exhibits traits absent in Ancient Greek but found in modern neighbouring languages – including in adpositional usage (Bortone 2010: 215–18, 240–45).

2.3 Nominal relation marking and its realisations in Greek

Many linguists accept that case endings, simple adpositions, and multi-word adpositions can all be thought of as marking the same type of nominal relation (cf. e.g. Fillmore 1968: 30) – as being all either simple or complex relators (see Fagard et al., this volume). An overlap in the functions of cases and of adpositions is clear in Modern Greek:

(1) plain case *ta mátja tis yátas*
 the eyes DET.GEN.SG cat.GEN.SG
 ‘the eyes of the cat’

(2) adposition *ta mátia apó tis yátes*
 the eyes from DET.ACC.PL cats.ACC.PL
 ‘the eyes of the cats’

In (1), the plain genitive case marks a partitive/possessive relation between the cat and her eyes; in (2), the same relation is expressed by the preposition *apó* ‘from’ (the genitive plural of many nouns is wanting, and is replaced by a prepositional phrase with *apó*; with the nouns in the singular, however, *apó* is not used as a substitute for the genitive). The preposition *apó* is combined with the

² The ‘case’ status of some of the Greek cases, and their names, are questionable (see Bortone 2002 and 2003).

accusative case because that is the case used in Modern Greek with all prepositions, except for archaic constructions. Unlike the genitive in (1) and (3) below, the accusative in (2) and (4) is therefore a structural case.

More examples of functional equivalence between case and preposition in Modern Greek are easy to come by. For instance, the preposition *s(e)*³ ‘to, at, on, in’:

(3) plain case *ípa* *tu* *Pétru*
 say.PST.IND.1SG DET.GEN.SG Peter.GEN.SG
 ‘I said to Peter’

(4) preposition *ípa* *s ton* *Pétro*
 say.PST.IND.1SG to DET.ACC.SG Peter.ACC.SG
 ‘I said to Peter’

Between (3) and (4) there is total synonymity, though (4) is more common.

Besides *apó* ‘from’ and *s(e)* ‘to, at, on, in’, Modern Greek has a number of other simple adpositions, such as *ja* ‘for’, *me* ‘with’, *metá* ‘after’, *katá* ‘towards’, *pros* ‘towards’, *san* ‘like’, *díhos* ‘without’, *horís* ‘without’, *méhri* ‘as far as’, *(é)os* ‘as far as’, and *pará* ‘contrary to’.

Furthermore, in Modern Greek, some adverbs, when combined with a prepositional phrase, appear to form a compound preposition: [ADV + [P + NP]] has become [[ADV + P] + NP]. Such compound prepositions can be equivalent to simple ones. Compare (5) with (6) and (7) with (8) below:

(5) simple preposition *ítane* *s to* *spíti*
 be.PST.IND.3SG at DET.ACC.SG house.ACC.SG
 ‘(s)he was in the house’

(6) compound preposition *ítane* *mésa s to* *spíti*
 be.PST.IND.3SG inside at DET.ACC.SG house.ACC.SG
 ‘(s)he was in(side) the house’

3 Modern Greek *s* is synchronically considered an elided form, and the full form is now thought to be *se*, which is used as such in other contexts (e.g. *se mia méra* ‘in one day’). However, historically, *s* derived from *is* (< Classical Greek εἰς /e:s/ < /ens/), and the final *-e* of the form *se* is a late addition. The form *s*, in Modern Greek orthography, is spelt together with the following article (e.g. *sto* ‘in the’). Given that this article addresses mostly readers who do not know Modern Greek, in the examples here, *s-* is written on its own so as to show morphemic boundaries; but it should be remembered that this is not done in Greek orthography.

(7) simple preposition

to vivlío íne s to trapézi
 DET.N.SG book be.PRS.IND.3SG at DET.N.SG table
 ‘the book is on the table’

(8) compound preposition

to vivlío íne páno s to trapézi
 DET.N.SG book be.PRS.IND.3SG up at DET.N.SG table
 ‘the book is on the table’

Sentences (5) and (6) above can be wholly synonymous in the sense of describing the very same state of affairs, though (6) is less ambiguous, because it specifies more clearly that the location of the referent is inside the house: (5) could, in principle, also mean ‘at the house’ (outside) or ‘on the house’. Similarly, (7) and (8) can be wholly synonymous, but (7) could also mean ‘in the table’, whereas the compound prepositional phrase in (8) means unequivocally ‘on (top of) the table’.

In the next section, we review several such Modern Greek compound prepositions; further on (in §4), we shall consider more complex strings.

3 The Modern Greek compound adpositions

3.1 Structure of the compound adpositions

These are two-word strings like *píso apó* ‘behind’. They are, in many respects, comparable to English strings like *close to*, *away from*, *inside of*, etc. There are reasons, which we shall discuss further on, for positing that their structure is [[ADV + P] [NP]] and not a mere [ADV [P NP]] sequence – although, of course, the latter is what they developed from, historically.

It must be noted that the second element of such compound prepositions is obligatory (unless the first element is used without a complement, as an adverb):

- (9) * *píso éna ðéndro*
 behind one.n.sg tree
 must be phrased as
píso apó éna ðéndro
 behind from one.n.sg tree
 ‘behind a tree’

- (10) * *páno to kreváti*
 up/on DET.N.SG bed
 must be phrased as
páno s to kreváti
 up/on at DET.N.SG bed
 ‘on the bed’

The second element of a compound preposition can, at most, be replaced by the genitive case, provided that the object is a ‘weak’ clitic pronoun (and, usually, that the referent is animate):

- (11) *páno tu* (12) *píso mas*
 on he.GEN back we.GEN
 ‘on him’ ‘behind us’

Modern Greek compound adpositions consist of adverbs combined with one of the following simple prepositions: *s(e)* ‘at, in, on, to’; *apó* ‘from, off, by’; *me* ‘with’, or *ja* ‘for, about’ (and, if we accept some doubtful cases discussed in §4.2, also with *pros* ‘towards’). For example⁴:

- With *ja*: *andí ja* ‘instead of’
 With *me*: *mazí me* ‘together with’
ísa me ‘as far as’
simfona me ‘according to’
 With *s(e)*: *mésa s(e)* ‘in; inside’
anámesa s(e) ‘between; among’
enándia s(e) ‘against’
 With *apó*: *ékso apó* ‘out of; outside’
ektós apó ‘except’
ísterá apó / épita apó ‘after’
káto apó ‘under’
makriá apó ‘far from’
metá apó ‘after’ (with pronouns and indefinite NPs)
péra apó ‘beyond’
píso apó ‘behind’
prin apó ‘before’

⁴ Some of the compound adpositions here listed are still occasionally replaced by their archaising synonyms, which usually take the genitive instead of the simple preposition, e.g. *kondá s(e)* ‘near’ > *enjís* +GEN; *makriá apó* ‘far from’ > *makrán* +GEN; *anámesa s(e)* ‘between’ > *metaksí* +GEN, etc.

With either *s(e)* *andíkri s(e)* / *andíkri apó* ‘opposite’
 or *apó*: *apénandi apó* / *apénandi s(e)* ‘facing’
đípla apó / *đípla s(e)* ‘beside’
(e)páno s(e) ‘on’ / *(e)páno apó* ‘over; above’
brostá apó / *brostá s(e)* ‘in front of’
plái apó / *plái s(e)* ‘flanking’
jíro s(e) / *jíro apó* ‘around’
trijíro apó / *trijíro s(e)* ‘all around’
olójira s(e) / *olójira apó* ‘all around’
kondá s(e) / rarely *kondá apó* ‘near’
simá s(e) / very rarely *simá apó* ‘near’

It is difficult to draw a neat distinction between the adverbs that form a complex preposition only with *s(e)* or only with *apó* and those that can combine, alternatively, with either. Some adverbs combine very often with one and very rarely with the other.

On the surface, the lack of clarity about this is made worse by the fact that the adverbs listed above as constituting a compound only with one simple adposition do occur also followed by other simple prepositions – but, arguably, they do so without constituting a compound. Observe how – as we noted in the list above – while *ékso apó* and *káto apó* are compounds, *ékso s(e)* and *káto s(e)* are not⁵:

(13) *píje ékso s ti veránda*
 GOPST.3SG out to DET.F.SG veranda
 ‘(s)he went out, on the veranda’

(14) *ton vríkane káto s to ipójio*
 him.ACC find.PST.3PL down/under at/in DET.N.SG cellar
 ‘they found him down(stairs), in the cellar’

In a similar fashion, some adverbs which can combine into a compound adposition with *s(e)*, such as *mésa* ‘inside’, can also be followed by *apó*, but do not seem to form the same kind of unit with *apó* as they do with *s(e)*: *apó* retains its own meaning, because it appears to be selected only when the action described

⁵ See also examples (56), (57), (58) further on.

is ablative (or perlativ: Modern Greek systematically expresses perlativ motions as ablative – cf. Bortone 2010: 233–4). For example:

- (15) *évyala* *ti* *sisceví* *mésa* *apó* *to*
 take.PST.IND.1SG DET.ACC.F.SG appliance inside from DET.ACC.N.SG
kuḗ
 box
 ‘I took the appliance out of [lit. inside from] the box’

This takes us to the issue of how to assess if a string of words makes up a complex adposition. There are multiple, if tentative, ways to distinguish sequences that constitute a compound from others that do not.

3.2 Some possible diagnostics for adpositional status

The idea that multi-word sequences that are semantically and distributionally equivalent to adpositions constitute adpositions was firmly dismissed in the past (e.g. Brøndal 1940). Even monolectic prepositions such as English *without* or Ancient Greek *metaxú* ‘between’ were not accepted as prepositions, presumably because they were identified as polymorphemic – though one may debate also whether, synchronically, they actually are polymorphemic. There has been, however, an increasing recognition that complex adpositions do exist; what remains controversial and uncertain is the status of many specific strings. A number of diagnostic principles can be applied, though these too are open to debate. Let us review a few.

3.2.1 Orthography

One criterion that we can quickly dismiss is orthography; this is obvious to a linguist, but often not to a layperson. Univerbation can be a matter of arbitrary convention and, indeed, spelling can vary. For example, the Modern Greek preposition *isame* ‘as far as’ still appears also as *isa me*. In English, we write *instead* (< *in stead*) as one word, but we can still sunder it with insertions: ‘I can go *in his stead*’; we write also *into* as one word (unlike *out of*), though it comes from *in to*, and it still conveys two sememes: motion + enclosure; we write *in front of* as three words, but *notwithstanding* as one, as we do also with words like *inasmuch* and *insofar*.

3.2.2 Cross-linguistic comparisons

If Modern Greek did not have its compound prepositions (or its learned archaisms), it would have no ability to make the distinction that English makes between *on* and *in*, and would have no equivalents of English words of undisputed adpositional status such as *near* or *under*. One (not particularly strong) argument for the classification of certain Modern Greek word strings as compound prepositions is that they correspond to simple (i.e. monolectic) prepositions in English. Compare the sample sentences with their translations:

- (16) *kriftice píso apó éna ðéndro*
 hide.PST.IND.3SG back from DET.N.SG tree
 ‘(s)he hid behind a tree’
- (17) *méno kondá s to panepistímio*
 reside.PRS.IND.1SG near at/to DET.N.SG university
 ‘I live near the university’
- (18) *vréθikan ékso apó to sholío*
 find.PST.IND.MPASS.3PL outside from DET.N.SG school
 ‘they met up outside the school’
- (19) *ítan ce éna peðí anámesa s ta θímata*
 be.PST.IND.3SG and one child amidst at/in DET.N.PL victims
 ‘there was also a child among the victims’
- (20) *évale ti valítsa káto apó to kreváti*
 put.PST.IND.3SG DET suitcase down from DET.N.SG bed
 ‘(s)he put the suitcase under the bed’
- (21) *ftásane prin apó ména*
 arrive.PST.IND.3PL before from me
 ‘they arrived before me’

Some prudence is needed here, because English cannot not be casually treated as a neutral, universal meta-language. However, several other European languages are aligned with English: the (synchronically) monolectic English *behind* – which, as we saw, in Modern Greek translates as *píso apó* – is matched also, for example, by the monolectic French *derrière*, Swedish *bakom*, Polish *za*, and Dutch *achter*; the monolectic English *under* – corresponding to the Modern Greek polylectic *káto*

apó – is matched by the monolectic French *sous*, Swedish *under*, Polish *pod*, and Romanian *sub*; the monolectic English *among* – corresponding to the Modern Greek polylectic *anámesa s(e)* – is matched by the monolectic French *parmi*, Swedish *bland*, Polish *wśród*, Italian *fra*, Romanian *printre*.⁶ So, at the very least, we can say that the Modern Greek phrases are functionally and semantically akin to adpositions.⁷

Many other Modern Greek compound adpositions have a similar or identical match in English and in Romance languages (though some of those phrases have also monolectic synonyms⁸):

Modern Greek	French	Italian	Spanish	English
<i>dípla s(e)</i>	<i>à/du côté de</i>	<i>a lato di</i>	<i>junto a</i>	<i>next to</i>
<i>méhri s(e)</i>	<i>jusqu'à</i>	<i>fino a</i>	-----	<i>up to, up until</i>
<i>prin apó</i>	<i>antérieurement à</i>	<i>prima di</i>	<i>antes de</i>	<i>prior to, ahead of</i>
<i>makriá apó</i>	<i>loin de</i>	<i>lontano da</i>	<i>lejos de</i>	<i>far from</i>
<i>kondá s(e)</i>	<i>(au)près de</i>	<i>vicino a</i>	<i>cerca de</i>	<i>close to (near to)</i>
<i>ékso apó</i>	<i>en dehors de</i>	<i>fuori da</i>	<i>fuera de</i>	<i>out of</i>
<i>brostá s(e)</i>	<i>(en face de)</i>	<i>davanti a</i>	<i>delante de</i>	<i>in front of</i>
<i>andí ja</i>	<i>au lieu de</i>	<i>invece di</i>	<i>en lugar de</i>	<i>instead of</i>
<i>(e)páno apó</i>	<i>au dessus de</i>	<i>sopra a</i>	<i>encima de</i>	<i>on top of</i>
<i>símfona me</i>	<i>conformément à</i>	<i>in linea con</i>	<i>de acuerdo con</i>	<i>in accordance with</i>
<i>mésa s(e)</i>	<i>à l'intérieur de</i>	<i>dentro a</i>	<i>dentro de</i>	<i>into, inside of</i>
<i>píso apó</i>	<i>en arrière de</i>	<i>dietro a</i>	<i>detrás de</i>	<i>(American in back of)</i>
<i>metá/épita apó</i>	<i>après de</i>	<i>dopo di</i>	<i>después de</i>	-----
<i>káto apó</i>	<i>en dessous de</i>	<i>sotto a</i>	<i>debajo de</i>	-----
<i>péra apó</i>	<i>au delà de</i>	<i>oltre a</i>	<i>(más allá de)</i>	-----
<i>jíro apó</i>	<i>autour de</i>	<i>intorno a</i>	<i>alrededor de</i>	-----

The existence of identical phrases in several languages besides Modern Greek, interesting as it is as evidence of a common pattern, does not in itself prove that the Modern Greek phrases above constitute adpositions. Nonetheless, it is worth noting that linguists and philologists working on those other languages have often recognised such phrases as adpositions.

⁶ Of course, many of these were not monolectic or monomorphemic historically: English *behind* < *bi+hind(an)*; French *derrière* < Latin *de(+ad?)+retro*; Swedish *bakom* < *bak+om*; Romanian *printre* < *pre+între*; English *among* < *on+(ge)mang*; French *parmi* < *par+mi*, etc.

⁷ Conversely, sometimes English uses polylectic adpositions to express what Greek can express with simple adpositions: English *in spite of* = Greek *pará*; English *as far as* = Greek *os*.

⁸ So, e.g. French *antérieurement à* can be synonymous with *avant*, as English *prior to* can be with *before*, etc.

3.2.3 Intra-linguistic comparisons

Several European languages have, synchronically, simple and compound forms that can be used interchangeably. English has largely synonymous pairs such as these:

(22) *outside the house* ≈ *out of the house*

(23) *before your visit* ≈ *prior to your visit*

English has even pairs in which the only difference is the addition of the second element, the first one (the adverb) being the same:

(24) *outside the house* ≈ *outside of the house*

(25) *near the house* ≈ *near to the house*

Similar pairs can also be seen, for example, in Italian:

(26) *dietro quella porta* ≈ *dietro a quella porta* ‘behind that door’

(27) *sotto i tappeti* ≈ *sotto ai tappeti* ‘under the rugs’

It would be somewhat counterintuitive to say that, in each of these pairs, one item is an adposition and the other is not.

For Greek, intralinguistic *diachronic* comparisons can perhaps strengthen the argument that we should recognise certain Modern Greek strings as adpositions: some of the modern compound adpositions are synonymous with simple prepositions of Ancient Greek whose adpositional status is not disputed (though it must be stress that Ancient Greek combined the prepositions, meaningfully, with a case).⁹ Compare the Modern Greek examples below with their Ancient synonyms:

⁹ The overall meaning of the adpositional phrase (e.g. ‘in a room’ vs ‘into a room’) depends also on the semantic contribution of the case. Cf. German ‘in einem Zimmer’ / ‘in ein Zimmer’ or Latin ‘in cubiculo’ / ‘in cubiculum’.

Modern Greek

(28) *ḡipla s to potámi*
by at DET river
'by the river'

(30) *mésa s to potámi*
in at DET river
'in the river'

(32) *páno s to potámi*
on at DET river
'on the river'

(34) *páno apó to potámi*
on from DET river
'above the river'

Ancient Greek

(29) *parà tōi potamōi*
by DET.DAT river.DAT
'by the river'

(31) *en tōi potamōi*
in DET.DAT river.DAT
'in the river'

(33) *epì tōi potamōi*
on DET.DAT river.DAT
'on the river'

(35) *hypèr toũ potamouĩ*
above DET.GEN river.GEN
'above the river'

In the case of some compound adpositions, the equivalent expressions used in Ancient Greek (and in archaising styles of Greek even today), consist of the *same first element* as in Modern Greek, but combined with genitive marking on the noun rather than with a simple adposition:

Modern Greek

(36) *ektós apó ména*
outside from me
'except me'

(38) *prìn apó ton pólemo*
before from DET war
'before the war'

(40) *káto apó ton omfaló*
down from DET navel
'below the navel'

(42) *ékso apó tin póli*
out from DET city
'outside the city'

Ancient Greek (and archaising Modern Greek)

(37) *ektòs emoũ*
outside me.GEN
'except me'

(39) *prìn toũ polémou*
before DET.GEN war.GEN
'before the war'

(41) *kátō toũ omphaloũ*
down DET.GEN navel.GEN
'below the navel'

(43) *éksō tēs póleōs*
out DET.GEN city.GEN
'outside the city'

3.2.4 Combinatory meanings

Another recognised criterion for assessing adpositional status is semantics. When [ADV + P] strings express a combinatory meaning that cannot be read off the meaning of their components, it is usually accepted that such strings constitute a compound adposition (Theophanopoúlou-Kondoú 1994: 236). A plausible argument for the recognition of Modern Greek strings like *píso apó* as adpositions, therefore, is that the elements that make up such compounds do not simply contribute their individual meanings: the compound can have a meaning of its own. For example, despite including the word *apó* ‘from’, *píso apó* does not necessarily mean ‘from behind’: its combinatory meaning need not be ablative or perlocative. Only if it were necessarily ablative or perlocative, this would indicate the distinctive, individual semantic contribution of *apó* as an independent simple adposition. But this is not the case. Although *píso apó* can be used in sentences like

(44) *i yáta vjíce písó apó ton kanapé* ‘the cat came out from behind the sofa’ (ablative)

(45) *i yáta pérase písó apó ton kanapé* ‘the cat passed behind the sofa’ (perlocative)

it is also used in sentences like these:

(46) *i yáta íne písó apó ton kanapé* ‘the cat is behind the sofa’ (locative)

(47) *i yáta píje písó apó ton kanapé* ‘the cat went behind the sofa’ (allative)

And the same applies to *káto apó*:

(48) *i yáta vjíce káto apó ton kanapé* ‘the cat came out from under the sofa’ (ablative)

(49) *i yáta pérase káto apó ton kanapé* ‘the cat passed under the sofa’ (perlocative)

(50) *i yáta íne káto apó ton kanapé* ‘the cat is under the sofa’ (locative)

(51) *i yáta píje káto apó ton kanapé* ‘the cat went (to) under the sofa’ (allative)

Indeed, to express more clearly the ablative or perlocative meaning of (48) and (49), one can also add another *apó* ‘from’ in front of the compound: *apó káto apó* ‘from under’.

Similar considerations apply also to *ékso apó*:

- (52) *vjéno ékso apó to sinemá* ‘I come out of the cinema’ (relative)
- (53) *pernáo ékso apó to sinemá* ‘I pass by / walk past the cinema’ (perlocative)
- (54) *kaθótan ékso apó to sinemá* ‘(s)he was sitting outside the cinema’ (locative)
- (55) *éftase ékso apó to sinemá* ‘(s)he arrived outside the cinema’ (allative)

3.2.5 Comparable non-compounds

Evidence that some [ADV + P] strings are compounds comes also from the fact that the same word strings can behave in two different ways, creating what we could call a minimal pair. For instance, compare two Modern Greek sentences that are lexically identical – except that, in (57), the intonation (or, in writing, a comma) separates the adverb from the simple preposition:

- (56) *cimáte (e)páno s to kreváti tu*
 sleep.PRS.IND.3SG up at DET bed his
 ‘he sleeps on his bed’
- (57) *cimáte (e)páno, s to kreváti tu*
 sleep.PRS.IND.3SG up at DET bed his
 ‘he sleeps upstairs, in his bed’

We want to be able to say that *(e)páno s(e)* in (56) is a compound but that in (57) it is not. Indeed, the sequence *(e)páno s(e)* as used in (56) is almost never broken by another word: very occasionally, a single short lexeme such as *lipón* ‘well’ and *ómos* ‘though’ can be interjected. On the other hand, a non-compound like (57) above can be interrupted by multiple terms. Only (57), and not (56), could become (58):

- (58) *cimáte (e)páno ópos pándote s to krevati tu*
 sleep.PRS.IND.1SG up like always at DET bed his
 ‘he sleeps upstairs, as always, in his bed’

3.2.6 Counterevidence

However, in many cases, the first element of the compound preposition can also be placed *after* the NP, which seems to suggest that it is an independent adverb. So (59) below can be phrased as (60), although (59) is the unmarked and far more common syntax:

(59) *brostá s tin ísoðo*
 (in)front at DET.F.SG entrance
 ‘in front of the entrance’

(60) *s tin ísoðo brostá*
 at DET.F.SG entrance (in)front
 ‘in front of the entrance’

Indeed, the first element of the string can exist independently as an adverb:

(61) *kaθótane brostá*
 sit.PST.IND.3SG front
 ‘(s)he was sitting at the front’

(62) *cimáte (e)páno*
 sleep.PRS.IND.3SG up
 ‘(s)he sleeps upstairs’

More importantly, in [ADV + P] strings that we may have reasons to see as compounds, the simple preposition is not always entirely bleached semantically. This is shown by minimal pairs of compound adpositions in which the same adverb is combined, respectively, with *s(e)* and with *apó* – for example, pairs such as *(e)páno s(e)* and *(e)páno apó*. In such pairs, both combinations might be classified, on the basis of the criteria that we have reviewed above, as compound adpositions. However, one can sometimes detect a subtle semantic difference between the members of the pair; and it can be argued that such difference can, ultimately, be explained by reference to the semantics of *s(e)* and of *apó*: with *s(e)*, there can be a stronger sense of inclusion, or contact, or interaction than with *apó*. In other words, the simple adposition still contributes a bit of its own independent meaning. Consider the following examples:

(e)páno s(e) ‘on’ / *(e)páno apó* ‘over’
 (63) *to paltó páno s to kreváti* ‘the coat on the bed’ (contact)

- (64) *to fotistikó páno apó to kreváti* ‘the light-fitting above the bed’ (distance / no contact)

jíro s(e) / jíro apó ‘around’ (English would translate both with the same adposition)

- (65) *perpatáo jíro s tin periocí* ‘I walk around (in) the area’ (contact and engagement)
- (66) *i trochá jíro apó ton planíti* ‘the orbit around the planet’ (no contact or engagement)

trijíro s(e) / trijíro apó ‘all around’ (English would translate both with the same adposition)

- (67) *kítaksa trijíro s to domátjo* ‘I looked around the room’ (inside; contact and engagement)
- (68) *ðéndra trijíro apó to spíti* ‘trees around the house’ (outside; no contact or engagement)

andíkri s(e) / andíkri apó ‘opposite’ (English would translate both with the same adposition)

- (69) *andíkri s to spíti tu* ‘opposite his house’ (emphasising proximity and visual contact)
- (70) *andíkri apó tin aktí* ‘opposite (=off) the coast’ (emphasising separation)

brostá s(e) / brostá apó ‘in front of’ (English would translate both with the same adposition)

- (71) *státhice brostá s tin tileórasi* ‘(s)he stood in front of the tv’ (to watch it – engagement)
- (72) *státhice brostá apó tin tileórasi* ‘(s)he stood in front of the tv’ (thus preventing others from watching it)

3.3 Combinations normally discounted

If we do accept that certain adverbs, when followed by certain simple adpositions, form compound adpositions, we have also another problem. Which adverbs should we (or should we not) recognise as forming complex adpositions? Several [ADV + P] strings are not only syntactically but also semantically very similar to

recognised adpositions. This perhaps suggests that we should classify these too as compound adpositions. For example:

hória ‘apart’ > *hória apó* ‘away from’
protítera ‘earlier’ > *protítera apó* ‘earlier than’
apómera ‘at some distance’ > *apómera apó* ‘apart from’
kséhora ‘separately’ > *kséhora apó* ‘separately from’

But where do we draw the line? Consider some more candidates:

ðeksiá ‘right-hand side’ > *ðeksiá apó* ‘to the right of’
aristerá ‘left-hand side’ > *aristerá apó* ‘to the left of’
 (73) *ðeksiá / aristerá apó to párho* ‘to the right/left of the park’

vória ‘north(wards)’ > *vória apó* ‘to the north of’
nótia ‘south (wards)’ > *nótia apó* ‘to the south of’
anatoliká ‘east(wards)’ > *anatoliká apó* ‘to the east of’
ðitiká ‘west(wards)’ > *ðitiká apó* ‘to the west of’
 (74) *vória / nótia / anatoliká / ðitiká apó to céndro* ‘to the north/south/east/west of the centre’

şçetiká ‘relatively’ > also *şçetiká me* ‘regarding’
 (75) *erotímata şçetiká me tin evropaiكى énosi* ‘questions about the European Union’

ásçeta ‘unrelatedly’ > *ásçeta apá / pros / me* ‘regardless of’¹⁰
 (76) *ásçeta me (or apó or pros) tin ilikía* ‘regardless of age’

parálima ‘in parallel’ > *parálima me* ‘in parallel with’
 (77) *ðulévun parálima me tis spuðés tus* ‘they have a job, in parallel with their studies’

aneksártita ‘independently’ > *aneksártita apó / me* ‘independently of’
 (78) *aneksártita apó ti ynómi tu* ‘independently of his opinion’

10 Again, more archaic forms of the adverb can be used (*ásçeta* > *asçétos*), in which case the more archaic syntax, i.e. the use of the genitive case instead of an added simple adposition, becomes an additional option. So, for example, ‘regardless of age’ can be expressed as *asçétos pros tin ilikía* but also as *asçétos ilikías* (the plain genitive case). This, again, indicates a parallel between one-word and multi-word adpositions. The more modern form *ásçeta*, as indicated above, can only be used compounded with a simple adposition.

andiθeta ‘contrariwise’ > *andiθeta apó / me / pros* ‘unlike, contrary to’¹¹
 (79) *andiθeta apó* (or *pros* or *me*) *aftó pu nóμισa* ‘contrary to what I thought’

On the issue of where to draw the line, we are faced with a further question: which *types* of two-word strings should, in principle, be accepted or ruled out? Above, we have reviewed the [ADV + P] type (e.g. *káto apó*) somewhat extensively because, in Modern Greek, this is the most common type of adposition, after the plain simple adposition [P]. But Modern Greek has also rare instances of other types of two-word sequences that are used in adpositional fashion, such as:

[Conj] + P] e.g. *óso ja* ‘regarding’ (which largely matches, in structure and meaning, the French *quant à*, and the English *as for*)

[N + P] e.g. *hári s(e)* ‘thanks to’ (which matches, in structure and meaning, English *thanks to*, French *grâce à* and, if one replaces adposition with inflection, also German *dank* +DAT/GEN, Polish *dzięki* +DAT, Romanian *mulțumită* +DAT, Bulgarian *благодаря* +DAT, etc.)

Modern Greek has, above all, also several *longer* strings used in (arguably) adpositional fashion. This is what we turn to next.

4 More complex strings: Phrasal adpositions

In Modern Greek, numerous multi-word phrases that are distributionally comparable to prototypical adpositions are also in use. These are not only morphologically more complex, but also, as such, more semantically specific (see Fagard et al., this volume). We can review a few types, assessing their status in light of the considerations we have made about compound prepositions.

The types of Modern Greek multi-word strings that (especially on the basis of what linguists working on other languages have suggested) one might consider for adpositional status are primarily the ones listed below. They will be looked at one by one in the following sections.

¹¹ Here too, given the influence of archaising styles, older forms can be used instead – either as simple adposition with the genitive case or in combination with a simplex preposition. So *andiθeta* can be replaced by the older *andiθétos* +GEN or by *andiθétos pros*, etc.

Format	§	Example	Meaning	Numerical frequency ^a
[P + Det + N] + NPGEN	4.1	<i>katá ti diárcia</i>	during	131.11 per million words
		<i>ipó to fos</i>	in light of	0.11 per million words
[P1 + N + P2] + NPACC	4.2	<i>se scési me</i>	in relation to	119.8 per million words
		<i>se sinðiazmó me</i>	in conjunction with	50.03 per million words
[P + N] + NPGEN	4.3	<i>ek mérus</i>	on [the] part of	33.34 per million words
		<i>eks etías</i>	because of	74.64 per million words
[P + N] + NPACC	4.4	<i>me stóho</i>	with [the] goal of	59.07 per million words
		<i>me théma</i>	on the topic of	47.26 per million words
[NPDAT/ACC] + NPGEN	4.5	<i>méso</i>	by means of	342.87 per million words
		<i>vási</i>	on the basis of	55.1 per million words

^aThese figures have been extracted from the TenTen database of Sketch Engine by Lexical Computing Limited.

4.1 Complex adpositions of the type [P + Det + N] + NPGEN

The label P, in this formula, is used with diachronic latitude: most of the simple prepositions that appear in the strings of this category, such as *eks*, *en*, *is*, *epí*, *ipó*, are disused *Ancient* Greek adpositions. These strings are, at least in origin, whole adpositional *phrases*, complete with a noun. They constitute a closed category. Examples include:

pará to plevró +GEN ‘by the side (of)’ (lit. ‘by the flank’)
epí ti efcería +GEN ‘on the occasion (of)’ (lit. ‘on the opportunity’)
epí ti emfanísi +GEN ‘upon presentation (of)’ (lit. ‘upon the exhibition/production’)
ipó tin ejíða +GEN ‘under the aegis (of)’ (lit. ‘under the shield’)
ipó to fos +GEN ‘in light (of)’ (lit. ‘under the light’)
ipó to próscima +GEN ‘under the pretext (of)’
katá ti diárcia +GEN ‘during’ (‘over the duration’)

(80) *to sinédrio élave hóra ipó tin ejíða tu tmímatos*
‘the conference took place under the aegis of-the department’

These phrases can only be followed by a genitive, which marks the possessor of the preceding noun. Governing the genitive is against the rules of Modern Greek prepositional usage, and perhaps weakens the claim that these strings constitute adpositions.

Many of these strings, and their uses in an adpositional fashion, have parallels in other European languages, for example:

Modern Greek	English	French	Italian	German
<i>ipó tin ejíða</i> +GEN	<i>under the aegis of</i>	<i>sous l'égide de</i>	<i>sotto l'egida di</i>	<i>unter der Ägide von</i>
<i>ipó to fos</i> +GEN	<i>in light of</i>	<i>à la lumière de</i>	<i>alla luce di</i>	<i>im Licht von</i>
<i>ipó to próσçima</i> +GEN	<i>under the pretext</i>	<i>sous prétexte de</i>	<i>col pretesto di</i>	<i>unter dem Vorwand</i>

Modern Greek has phrases of this type not taken from the archaising style, but formed with modern adpositions, e.g. *s to éleos* +GEN ‘at the mercy (of)’ – which, again, is like the English *at the mercy of*, the French *à la merci de*, the Italian *alla mercé di*, or the Polish *na łasce* +GEN.

Some of the considerations that we made earlier about [ADV + P] sequences apply also here: we may feel inclined to regard Modern Greek phrases of this longer type as complex prepositions because they are functionally and semantically equivalent to adpositions. Indeed, many of them correspond to *monolectic* adpositions of many other European languages: *katá ti ðiárcia* (+GEN), ‘over the duration (of)’, corresponds to the monolectic English *during*, the monolectic French *pendant*, the monolectic German *während*, the monolectic Italian *durante*, and the monolectic Swedish *under*.

As with the [ADV +P] strings above, here too the same question arises: do such sequences form a unit? We might test this by checking, for example, whether modifiers may or may not be added (cf. Quirk et al. 1972: 302; Quirk et al. 1985: 671–2). If they may not, this suggests that these phrases have been grammaticalised. When a string gets grammaticalised, paradigmatic and syntagmatic variability is lost (cf. Lehmann 1985: 305ff.). However, grammaticalisation is a very gradual process, and this renders the categorial analysis of many sequences problematic. Besides, making alterations to the archaising phrases cited above (for instance, by inserting an adjective) is possible, but tricky if it requires the on-the-fly application of archaic grammatical forms alien to Modern Greek. For example, Modern Greek speakers can easily insert an adjective in phrases such as these:

- (81) *ipó tin ejíða* ‘under the aegis (of)’
 > *ipó tin epísimi ejíða* ‘under the official aegis (of)’
- (82) *ipó to próσçima* ‘under the pretext (of)’
 > *ipó to jelío próσçima* ‘under the laughable pretext (of)’

This is because the inserted adjective only needs to be in a form that is identical in the modern language. The same freedom to make insertions can be seen in the phrases that are entirely modern, for example:

- (83) *s to éleos* ‘at the mercy (of)’
 > *s to apólito éleos* ‘at the total mercy (of)’

However, insertions and variations are difficult and unlikely in archaising phrases that entail obsolete forms such as the dative case. For instance, one could not make changes to (84):

- (84) *epí ti efcería* ‘given the opportunity; apropos’
 on the.DAT opportunity.DAT
 > **epí ti próti efcería* ‘at the first opportunity’

If, instead of *epí ti efcería*, one uses the equivalent modern phrase *me tin efcería* (consisting of *me*, an adposition in productive use, and the accusative case), there is more flexibility:

- (85) *me tin efcería* ‘given the opportunity’
 with the.ACC opportunity.ACC
 > *me tin próti efcería* ‘at the first opportunity’

4.2 Complex adpositions of the type [P₁ + N + P₂] + NPACC

The initial adposition of such [P₁ + N + P₂] strings is usually *s(e)*; the final adposition can be *me*, *s(e)*, *ja*, or *pros*. Examples:

- se scési me* ‘in relation to’
se andíthesi me/pros ‘in contrast with/to’
se analojía me ‘in proportion to’
se andálaγma ja ‘in return for’
se sinðiazmó me ‘in conjunction with’
se síngrisi me/pros ‘in comparison with/to’
me anáfora se ‘with reference to’

- (86) *i aðerfí su, se andíthesi me séna, mas*
 DET.F.SG sister POSS.2SG in contrast with you, us.ACC

voíthise *polí*
 help.PST.IND.3SG much
 ‘your sister, unlike you, helped us a lot’

And again, in such phrases, which are modern in syntax and morphology, one can easily insert adjectives:

(87) *se scési me* ‘in relation to’
 > *se ámesi scési me* ‘in direct relation to’

Some of these phrases can be replaced by etymologically-related [ADV + P] compounds of the type we saw earlier:

se analojía me ‘in proportion to’ > *análoya me* ‘in proportion to’
se scési me ‘in relation to’ > *scetiká me* ‘in relation to’
se síngrisi me ‘in comparison with’ > *síngritiká me* ‘in comparison with’

Some of these phrases can also be given an archaic colouring, mostly by replacing *se* +ACC with the archaic *en* +DAT. The old dative forms often differ from the modern accusative ones, notably in their accentuation (and in their Greek orthography¹²):

[[*se andíthesi*_{ACC}] *me*] > [[*en andíthési*_{DAT}] *pros*]
 [[*se sinðiazmó*_{ACC}] *me*] > [[*en sinðiazmó*_{DAT}] *me*]
 [[*se anaforá*_{ACC}] *pros*] > [[*en anaforá*_{DAT}] *pros*]
 [[*se analojía*_{ACC}] *me*] > [[*en analojía*_{DAT}] *me*]
 [[*se scési*_{ACC}] *me*] > [[*en scési*_{DAT}] *me*]

Many such Modern Greek strings have parallels in other European languages:

Modern Greek	English	Dutch	French
<i>se scési me</i>	<i>in relation to</i>	<i>met betrekking tot</i>	<i>par rapport à</i>
<i>se analojía me</i>	<i>in proportion to</i>	<i>in verhouding tot</i>	<i>en proportion de</i>
<i>me anaforá se</i>	<i>with reference to</i>	<i>ten opzichte van</i>	<i>en référence à</i>

¹² Thus, the dative form is written συνδυσασμῷ (in full ancient spelling, συνδυσασμῶ) while the modern accusative is συνδυσασμό, the dative form is ἀντιθέσει while the modern accusative is ἀντίθεση, the dative form is σχέσει while the modern accusative σχέση, etc.

<i>se andálayma ja</i>	<i>in return for</i>	<i>in ruil voor</i>	<i>en échange de</i>
<i>se sindíazmó me</i>	<i>in conjunction with</i>	<i>in verband met</i>	<i>en conjonction avec</i>
<i>se síngrisi me/pros</i>	<i>in comparison with</i>	<i>in vergelijking met</i>	<i>par comparaison avec</i>
<i>se andíthesi me</i>	<i>in contrast with/to</i>	<i>in tegenstelling tot</i>	<i>en contraste avec</i>

German	Italian	Swedish	Polish
<i>in Bezug auf</i>	<i>in rapporto a</i>	<i>i relation med</i>	<i>w związku z</i>
<i>im Verhältnis zu</i>	<i>in proporzione a</i>	<i>i proportion till</i>	<i>w stosunku do</i>
<i>unter Bezugnahme auf</i>	<i>in riferimento a</i>	<i>med hänvisning till</i>	<i>w odniesieniu do</i>
<i>im Gegenzug für</i>	<i>in cambio di</i>	<i>i gengäld för</i>	<i>w zamian za</i>
<i>in Verbindung mit</i>	<i>(in aggiunta a)</i>	<i>i samband med</i>	<i>w połączeniu z</i>
<i>im Vergleich mit/zu</i>	<i>in confronto a</i>	<i>i jämförelse med</i>	<i>w porównaniu z</i>
<i>im Gegensatz zu</i>	<i>in contrasto con</i>	<i>i kontrast med</i>	<i>w przeciwieństwie do</i>

Such parallels are worth pointing out also because it is possible that the use in Greek of some such phrases might have originated in calques – if so, most likely from French, in which such constructions are rife.

We should note, incidentally, that some English phrases of this type, such as *in spite of* are not only equivalent to monolectic adpositions like *despite*, but are commonly treated by linguists as complex adpositions (cf. Quirk et al. 1985: 669ff.).

4.3 Complex adpositions of the type [P + N] + NPGEN

Examples:

ðia mésu ‘through’ (lit. ‘through the middle (of)’)

ek mérus ‘on the part (of)’ (lit. ‘from part (of)’)

eks aformís ‘arising from, in connection with’ (lit. ‘from the departure point (of)’)

eks etías ‘because (of)’ (lit. ‘from cause (of)’)

en ópsi ‘in view (of)’ (lit. ‘in sight (of)’)

en aynía ‘unaware (of)’ (lit. ‘in ignorance (of)’)

is epíkoon ‘within earshot (of)’ (lit. ‘to hearing (of)’)

is város ‘at the expense (of)’ (lit. ‘to burden (of)’)

kata míkos ‘along’ (historically, ‘down/along length (of)’, see below)

pros epírosin ‘in confirmation (of)’ (lit. ‘towards / to corroboration (of)’)

pros timín ‘in honour (of)’ (lit. ‘towards honour’)

pros ófelos ‘for the benefit (of)’ (lit. ‘towards advantage (of)’)
ípo morfín ‘in the form (of)’ (lit. ‘under shape (of)’)

These are all archaising phrases, made of ancient prepositions combined with non-structural cases (and, often, with archaic case forms). For example:

(88) *aftó íne pros ófelos ólon mas*
 this bePRS.IND.3SG towards advantage all.GEN.PL we.GEN.PL
 ‘this is in the interest of all of us’

(89) *pes çeretismús ek mérus mu*
 say.IMPRT.2SG greetings from part.GEN me.GEN
 ‘say hallo from me’

Again, perhaps a reason for regarding Greek phrases of this kind as complex adpositions is that many of them correspond, in terms of meaning and distribution, to monolectic adpositions of the main European languages: the Modern Greek *ðia mésu* +GEN, lit. ‘through the middle (of)’, translates the monolectic English *through*, German *durch*, Swedish *genom*, Polish *przez*, etc. Indeed, without phrases of this type, Modern Greek would have no equivalent of (monolectic) English prepositions such as *along* or *during*. Modern Greek orthography indeed now spells *ðia mésu* as *ðiamésu* (διαμέσου), just as it spells *en ópsi* ‘in view’ as *enópsi* (ενόψει), and *eks etías* ‘because’ as *eksetías* (εξαιτίας). Some phrases of this type, moreover, express meanings that Greek itself, in earlier stages in its history, expressed with monolectic adpositions.

Other phrases of this category, on the other hand, have similarly polylectic parallels in several other European languages. The Greek *ek mérus* +GEN matches, semantically and literally, the English *on the part of*, the Italian *da parte di*, and the German *auf Seiten* +GEN. The Greek *en ópsi* +GEN matches the English *in view of*, the German *in Anbetracht* +GEN, the French *en vue de*, and the Italian *in vista di*. The Greek *eks etías* +GEN is comparable to the English *on account of*, the French *par/à cause de*, the Italian *per/a causa di*, the German *auf Grund* +GEN, the Polish *z powodu* +GEN, the Swedish *på grund av*. The list could go on.

Unsurprisingly, in Modern Greek, strings of this type, since they are taken from Ancient Greek, are not open to alterations. When changes such as insertions are possible, the results are rather telling. Let us take, for example, the expression *katá mikos*, which historically meant ‘down/along (the) length’, but in Modern Greek it simply means ‘along’. As it was common in Ancient Greek, but not in Modern Greek, the noun here has no article. Now, an article can be inserted: *katá to mikos* ‘along the length’. This, however, makes the idiom into

a non-grammaticalised modern phrase, in which every word contributes its full meaning. Compare:

(90) *katá mīkos ton sinóron* ‘along the border’

(91) *katá to mīkos ton sinóron* ‘along the length of the border’

The grammaticalisation of the phrase *katá mīkos*, as opposed to *katá to mīkos* (and the importance of the former as a compound adposition) is also made clear by sheer numbers: *katá mīkos* is over a hundred times more common: it occurs 12.36 times every million words, while *katá to mīkos* occurs only 0.1 times.¹³

The limited flexibility of the archaic [[P + N] + NPGEN] constructions which we are considering in this section, though it can be generically explained as due to grammaticalisation, may also be interpreted in a different way: invariability is often a characteristic of phrases borrowed from other languages, and Ancient Greek – despite what Greek archaisers claimed – is a different language from Modern Greek.

Some Greek [[P + N] + NPGEN] strings that are used in adpositional fashion, and that therefore belong to this subcategory, are modern – though they are mostly translated from *Katharévousa*. Their lexemes, their morphology, and their syntax are all current; only the lack of article imparts to them a more formal and somewhat archaic flavour. Examples include:

se períptosi +GEN ‘in case (of)’
se anazítisi +GEN ‘in search (of)’
se morfí +GEN ‘in the shape (of)’
apo ápopsi +GEN ‘in terms (of)’

(92) *i sinavliá tha anavliθí se períptosi vroçís*
 DET concert FUT postponePASS.3PS in case rain.GEN
 ‘the concert will be postponed in case of rain’

Such phrases are paradigmatically and syntagmatically more flexible, and one could argue that this casts doubts on their adpositional status. For example,

¹³ These figures have been extracted from the TenTen database of Sketch Engine, by Lexical Computing Limited.

the phrase at the end of (92), *se períptosi vroçís* ‘in case of rain’, could be turned into (93):

- (93) *se períptosi, paraðíymatos hárin, ðinatís vroçís*
 in case example.GEN sake.ACC strong.GEN rain.GEN
 ‘in case, for example, of heavy rain’

Finally, let us note – not as an indication of adpositional status but of possible calques – that the modern phrases of this category too have extensive parallels across European languages:

Modern Greek	English	Dutch	French
<i>se períptosi</i> +GEN	<i>in case of</i>	<i>in geval van</i>	<i>en cas de</i>
<i>se anazítisi</i> +GEN	<i>in search of</i>	<i>op zoek naar</i>	<i>en recherche de</i>
<i>se morfí</i> +GEN	<i>in the shape/guise of</i>	<i>in de vorm van</i>	<i>en guise de</i>

German	Italian	Swedish	Polish
<i>im Falle von</i>	<i>in caso di</i>	<i>i händelse av</i>	<i>w przypadku</i> +GEN
<i>auf der Suche nach</i>	<i>in cerca di</i>	<i>på jakt efter</i>	<i>w poszukiwaniu</i> +GEN
<i>in Form von</i>	<i>sotto/in forma di</i>	<i>i form av</i>	<i>w postaci</i> +GEN

4.4 Complex adpositions of the type [P + N] + NPACC

These are entirely modern phrases. Examples:

- me théma* ‘on the subject of’ (lit. ‘with, [as] topic, ...’)
me stóho ‘aiming at’ (lit. ‘with, [as] goal, ...’)
me vási ‘on the basis of’ (lit. ‘with, [as] basis, ...’)
me aformí ‘prompted by’ (lit. ‘with, [as] point of departure, ...’)
me proorismó ‘bound for’ (lit. ‘with, [as] destination, ...’)
me afetíriá ‘starting from’ (lit. ‘with, [as] starting point, ...’)
me ekséresi ‘except’ (lit. with, [as] exception, ...’)

- (94) *tu tilefónise me prófasi ta jenéthliá tu birthday.ACC he.GEN/DAT*
 he.GEN/DAT phone.PST.IND.3SG with pretext.ACC DET.N.PL.ACC
 birthday.ACC he.GEN/DAT
 ‘(s)he phoned him with the pretext of his birthday’

where the object of the simple adposition *me* is actually ‘the birthday’, and the noun *prófasi* ‘pretext’ is an apposition to it. The phrase could also be constructed as in English:

- (95) *me tin prófasi ton jeneθlíon tu*
 with DET.ACC.SG. pretext.ACC.SG DET.N.PL.GEN birthday.GEN he.GEN/DAT
 ‘with the pretext of his birthday’

4.5 Complex adpositions of the type [NPDAT/ACC] + NPGEN

These are frozen case-marked nouns from the archaising style, which are used unproductively as adpositions – though they are usually not counted as adpositions. Their archaic features include disused case-endings and the absence of articles. They are followed by the genitive of the referent noun. We are mentioning them here, although they consist of just a single word (and of case marking on the following constituent), only for the sake of completeness, because they are the synthetic equivalent of a whole phrase. Examples include:

Forms in the obsolete dative

vási +GEN ‘on (the) basis (of)’

apúsía +GEN ‘in (the) absence (of)’

parúsía +GEN ‘in (the) presence (of)’

sinepía +GEN ‘caused by’, lit. ‘as (the) consequence (of)’

elípsi +GEN ‘given (the) lack (of)’

lóyo +GEN ‘because (of)’, lit. ‘by reason (of)’

méso +GEN ‘via’, lit. ‘by middle (of)’

Old accusative forms

irímin +GEN ‘without consulting [someone]’

đícín +GEN ‘by way (of)’

hárin +GEN ‘for the sake (of)’

- (96) *píyame s ti jenévi méso ziríçis*
 go.PST.IND.1PL to DET Geneva via Zürich.GEN
 ‘we went to Geneva via Zürich’

- (97) *hárin ðiafánias*
 grace.ACC transparency.GEN
 ‘for the sake of transparency’

5 The diachrony: The emergence of complex adpositions in Greek

Greek is the Indo-European language with the longest and best-attested history: three thousand years with relatively minor gaps. This makes Greek a uniquely valuable terrain for studying, among other things, language change. In the history of Greek, a whole system of adpositions can be observed coming into existence, then being discarded, and being replaced.

Greek, across its diachrony, displays different strategies for marking the nominal relations considered typical of adpositions. The first thing to notice, relevant to our present topic, is that Greek has undergone a shift from synthesis to analysis:

(98) plain case → Archaic Greek: *oúresin* (Hom. Il. 5.52)
 mountain.DAT.PL
 ‘on the mountains’

(99) preposition + case → Classical Greek: *ep’ óresin*
 on mountain.DAT.PL
 ‘on the mountains’

(100) preposition + case → Modern Greek: *s ta vuná*
 on/at DET.N.PL mountain.ACC.PL
 ‘on the mountains’

(101) complex preposition → Modern Greek: *páno s ta*
 up at/on DET.N.PL
vuná
 mountain.ACC.PL
 ‘on the mountains’

At the time of Homer, there were adverbs occasionally used in adpositional fashion, both prepositionally and postpositionally, a usage occasionally seen also in later Classical literature. Adpositions thus appeared in Greek first in combination with case forms that were capable of expressing many spatial relations and typically adpositional meanings also on their own (cf. Bortone 2002: 70–72). Adpositions came to be associated with case-inflected nouns mainly to disambiguate or reinforce the meaning (Bortone 2010: 16–19). By the Classical period, in Greek we find many simplex prepositions, several of which could combine with multiple cases.

However, in those combinations, the semantic contribution of the case form gradually faded (Bortone 2010: 156–9), and the use of multiple case combinations (P_1+case_1 , P_1+case_2 , etc.) was slowly abandoned. By the Middle Ages, Greek had generalised the use of one case, the accusative, which marked prepositional case-government but added nothing to the meaning of the whole prepositional phrase. More importantly, new compound adpositions enabled the system to express lost semantic nuances. Such compounds have usually been neglected by philologists, despite their demonstrable importance in the history of Greek adpositions and for any theoretical analysis of it (Bortone 2014: 46). As early as in Homeric Greek we see adpositions occasionally combined (e.g. *hup+ék*, *apo+pró*); we also see, already then, rare constructions such as *ektós apó* ‘outside’, lit. ‘outside from’, a type of combination that is attested also in later Greek: e.g. *héōs eis* ‘until’, lit. ‘until to’ (Polybius 1.11.14). Both of these compounds, in format and meaning, have counterparts also in Modern Greek (Bortone 2010: 132), where we find *ékso apó* ‘outside’, lit. ‘outside from’, and *méhri s(e)* ‘until’ lit. ‘until to’. And both Classical and Hellenistic Greek had, occasionally, the more complex sequences of the $[[P + N] + NGEN]$ type that we saw also in Modern Greek. The developments seen in Greek largely parallel those seen in the trajectory from Latin to Romance (Bortone 2010: 215–7).

Since, initially, in Greek adpositional phrases, the case inflection on the noun carried a considerable semantic load, one could say that even simple adpositions were *complex* in structure. Furthermore, there is evidence, both within Greek and from comparative philology, that a number of Ancient Greek simple (monolectic) adpositions were originally inflected nouns. For example, the Ancient Greek *perí* ‘around’ was probably a noun in the locative case, just as the Modern Greek *jíro* ‘around’ was originally a noun in the dative case, thus meaning ‘in (the) circle/ring (of)’. The distinction between simple adpositions and more complex constructions, when observed closely, has therefore always been far less clear-cut than it is commonly assumed.

6 Conclusion: The status of compound and more complex adpositions

Basic compound prepositions (of the *píso apó* ‘behind’ type) are well-established in Modern Greek, and – at least in terms of sheer numbers – can even be considered to be now the main type of adposition. Modern Greek more complex adpositions (e.g. *se sçési me* ‘in relation to’) are likely to have arisen as a result of several factors, including a general typological shift from more synthetic towards more

analytic. And although this is harder to prove, they may occasionally have been learned calques from languages. The main reason for classifying such strings as complex adpositions is that they are distributionally similar to monolectic adpositions and perform similar functions as prototypical monolectic adposition, although such functions are distributed across their multiple components: an introductory (simple) adposition introducing a NP that specifies the nature of the relation expressed, and another (simple) adposition (or a case inflection) connecting the phrase to another NP (which in Greek can just be an article introducing a whole sentence).

One can also, however, take the position that the fact that the functions deemed typical of a single adposition are, in the case of the (putative) complex adpositions, performed by different components in the phrase undermines the very idea of a ‘compound’. And the fact that such phrases allow, though within limits, some variations also suggests that they are, to say the least, somewhat loose compounds. Besides, the semantics of such strings are also rather specialised: a complex adposition is semantically more specific – somewhat more literal, some might say – than simple (monolectic) adpositions typically are: the latter are usually more polysemous. Adpositions, furthermore, are usually thought to constitute a closed class, but such Greek phrases are an ensemble with unclear and porous borders. One may ask whether they really should be considered adpositions and, if so, where the boundaries of the ‘adposition’ category should be drawn.

But given that many, if not all, simple adpositions have developed from more complex forms, and others are at variable points along the same process, the distinction between the two types is one of degree. Ultimately, it is all perhaps a matter of definitions, rather than of ontology.

Abbreviations

ACC	accusative	MPASS	medio-passive
ADV	adverb	N	neuter
CONJ	conjunction	NP	noun phrase
DAT	dative	P	preposition
DET	determiner	PASS	passive
F	feminine	PL	plural
FUT	future	POSS	possessive
GEN	genitive	PRS	present
IMPRT	imperative	PST	past
IND	indicative	SG	singular

References

- Bortone, Pietro. 2014. Adpositions. *The encyclopedia of Ancient Greek language and linguistics*. Leiden: Brill, 40–8.
- Bortone, Pietro. 2010. *Greek prepositions – from antiquity to the present*. Oxford: Oxford University Press.
- Bortone, Pietro. 2003. On the Ancient Greek nominative. *Studies in Greek Linguistics* 23, 325–236.
- Bortone, Pietro. 2002. The status of the Ancient Greek cases. *Studies in Greek Linguistics* 22(1), 69–80.
- Bortone, Pietro. 2009. Greek with no history, no standard, no models. In Alexandra Georgakopoulou & Michael Silk (eds.), *Standard languages and language standards: Greek, past and present*. London: Ashgate, 67–89.
- Brøndal, Viggo Rasmus. 1940. *Præpositionernes Theori – indledning til en rationel Betydningslære* [A theory of prepositions – Introduction to a rational semantics]. Copenhagen: B. Lunos.
- Fagard, Benjamin, José Pinto de Lima, Dejan Stosic, & Elena Smirnova. This volume. Introduction: Complex adpositions and complex nominal relators.
- Fillmore, Charles. 1968. The case for case. Reprinted in René Dirven & Günter Radden (eds.), 1987. *Fillmore's case grammar – a reader*. Heidelberg: Groos, 21–33.
- Lehmann, Christian. 1985. Grammaticalization: synchronic variation and diachronic change. *Lingua e stile* 20, 303–318.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, & Jan Svartvik. 1972. *A grammar of contemporary English*. London: Longman.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.
- Theophanopoúlou-Kondoú, Dimitra. 1994. Σύνθετες προθέσεις της νέας ελληνικής – μια προβληματική κατηγορία [The compound prepositions of Modern Greek – A problematic category]. *Studies in Greek Linguistics* 15, 233–244.

Guillaume Bonnet and Benjamin Fagard

6 Complex prepositions in Albanian: A first assessment

Abstract: In this chapter, we investigate the marking of nominal relations in Albanian, from case marking to simple and complex prepositions, on the basis of the (limited) literature and of a corpus study. We rely on two corpora, the Albanian National Corpus (ANC) and the parallel corpus OPUS, totaling 20 and 46 million words respectively. This enables us to provide frequency lists and distributional features of simple and complex prepositions, making it possible to account for the formal and functional differences between different types of simple and complex adpositions. We also address the question of the origin of CAs in Albanian, showing a possible effect of language contact.

Keywords: corpus linguistics, Balkan languages, contact linguistics, case

1 Introduction

Albanian is a fascinating language, Indo-European but relatively isolated, with a few archaic Indo-European features but actually useless for Indo-European reconstruction on account of its long history of language contact, with numerous borrowings from Classical Latin (Bonnet 1999) and later from other neighboring languages, making it an important member of the Balkan Sprachbund (Sandfeld 1930). Concerning nominal relations, which are the focus of the present volume, Albanian is an interesting language, with case marking on nouns and adjectives, as well as an important paradigm of simple and compound prepositions – actually, simple and complex prepositions, strictly speaking (Boissin 1975, Buchholz and Fiedler 1987). This makes it an ideal test-case for the development of Complex Prepositions.

We begin with a genetic and typological overview of Albanian (Section 1), then present the main features of case marking and simple prepositions in Albanian (Section 2), before describing the methodology and results of our corpus study on complex prepositions (Section 3). Finally (Section 4), we discuss these results in light of structural and diachronic considerations, and conclude (Section 5).

Guillaume Bonnet, University of Bourgogne, Paris, France. Guillaume.Bonnet@u-bourgogne.fr
Benjamin Fagard, CNRS, ENS & Sorbonne Nouvelle; PSL, Montrouge, France. benjamin.fagard@ens.psl.eu

<https://doi.org/10.1515/9783110686647-007>

2 Albanian

Albanian (*shqip*, in Albanian) is an Indo-European language, mainly spoken in Albania, with important minorities in Macedonia, Montenegro, Kosovo, Sicily and the south of Italy, and smaller communities in Greece, Turkey, Germany and the United States. There are altogether 4¹ million native speakers of Albanian varieties. It has two main dialects, Gheg (1.2 million speakers in (the north of) Albania, 3.9 overall) and Tosk (1.6 million speakers in (the south of) Albania, 1.9 overall). Other varieties include those spoken in Italy (Arbëreshë, roughly 0.5 million speakers of four varieties which are not mutually intelligible) and Greece (Arvanitika, 0.05 million speakers, in rapid decline). In this chapter, we focus on literary, standard Albanian, a variant which is mostly based on Tosk.

Though it was recognized as pertaining to Indo-European in the mid-19th c. (Bopp 1854, Meyer 1891, Pedersen 1900), “Albanian shows no obvious close affinity to any other Indo-European language; it is plainly the sole modern survivor of its own subgroup” (Hamp 1972), despite putative links to Illyrian, Dacian and Messapian. It thus seems to represent an isolated branch of Indo-European, much like Greek and Armenian; however, Hamp (1972) also notes a likely link to the Balto-Slavic group. There has been intense language contact with neighboring languages, chiefly Latin, Greek, Turkish and Slavic (especially Bulgarian), to such an extent that Albanian cannot be used reliably for Indo-European reconstructions (Bonnet 2010). Indeed, it seems that Albanian has been very hospitable towards loanwords, which brought about important phonological changes.

More generally, alongside typical Indo-European features such as heavy nominal and verbal morphology, Albanian seems to have developed several new traits, linked in part to its geographic situation, with Balkan features (Sandfeld 1930) such as a postposed definite determiner (also found in Romanian and Bulgarian), an analytic future tense formed with a grammaticalized particle (*do < dua* ‘I will’) followed by the subjunctive (similar patterns are found in Serbian, Greek, Romanian and Bulgarian), the absence of infinitive forms (as in other Balkanic languages), differential object marking through clitic doubling (also in Romanian, Macedonian, Bulgarian, Greek, and marginally in Serbian, cf. Kallulli and Tasmowski 2008, Hill and Mardale forthcoming), etc. (see Table 1).

¹ 6 million by Ethnologue counts (Eberhard et al. 2019), which we used for the number of speakers of the different Albanian varieties.

Table 1: Shared grammatical innovations in the Balkans (Lindstedt 2000: 232).

	Greek	Albanian	Balkan Slavic	Balkan Romance	Balkan Romani
Argument marking					
ENCLITIC ARTICLES	(+)	+	+	+	(+)
OBJECT REDUPLICATION	+	+	+	+	+
PREPOSITIONS INSTEAD OF CASES	(+)	(+)	+	(+)	(+)
DATIVE / POSSESSIVE MERGER	+	+	+	+	–
GOAL / LOCATION MERGER	+	+	+	+	(+)
RELATIVUM GENERALE	+	+	+	+	+
Verb system					
AUX (+ COMP) + FINITE VERB	+	(+)	+	(+)	+
VOLO FUTURE	+	(+)	+	+	+
PAST FUTURE AS CONDITIONAL	+	+	+	(+)	(+)
HABEO PERFECT	+	+	(+)	(+)	–
EVIDENTIALS	–	+	+	(+)	(+)
Other					
ANALYTIC COMPARISON	(+)	+	+	+	+

As a result, from a typological perspective, Albanian is a degraded polysynthetic language. It is an SVO language, albeit with a relatively free word order. As far as the nominal morphology is concerned, there are six cases, with consistent morphological distinctions only between Nominative (from which Vocative differs only for a reduced set of names), Accusative and Dative (Genitive is identical to Dative; Ablative differs from Genitive and Dative only in the plural indefinite); Plural is marked by a suffixal element to which case marking is added very much as in agglutinative systems such as Turkish. Adjectives agree in gender (masculine/feminine) and number (singular/plural) with the determiner; 1st and 2nd possessive adjectives also agree in case.

Albanian also has a rich verbal system, with six persons and two voices (active and medio-passive), six modes (indicative, subjunctive, conditional, optative, imperative, admirative) and eight tenses (three simple tenses, present, imperfect and aorist; a complex future tense formed on the subjunctive; and four complex passive tenses based on a participle form and the auxiliary *kam* ‘to have’) formed on two basic forms, present and aoristic, with many cases of suppletion.

3 Case marking and simple prepositions

There are to our knowledge few studies dedicated to the analysis of simple prepositions in Albanian, and none on complex prepositions in Albanian alone. For our presentation, we therefore relied mostly on the reference grammars of Albanian, mainly Boissin (1975) and Buchholz and Fiedler (1987), on the few papers we found on simple prepositions, and on a contrastive analysis of prepositions in Serbian and Albanian (Krijezi 2012, which lists quite a few CPs), combined with the analysis of simple and complex prepositions in two corpora, the Albanian National Corpus (ANC) and the parallel corpus OPUS, totaling 20 and 46 million words respectively (see Section 3.1. for further details).

3.1 Case marking

3.1.1 Main features

Case is marked with nominal and adjectival suffixes, and with different forms on the pronoun; the Genitive also includes a specific “linking article” which varies in definiteness, gender, number and case (*i/e/të/së*). There are consistent formal differences for four different cases: Nominative, Accusative, Genitive and Dative. Additionally, some nouns have a distinct Vocative form, and there are different endings for the Ablative only in the plural. The extent of formal overlap between cases is thus important, mostly between Nominative and Vocative, and between the three oblique cases (Dative, Genitive and Ablative) (Likaj and Çabej 2013).

The use of case marking is mostly limited to grammatical functions, e.g. the marking of subjects and similar functions for the Nominative (1), of direct objects for the Accusative (2), of adnominal relations such as possession and belonging for the Genitive (3), of indirect objects and beneficiaries for the Dative (4), and of agents (Ablative).

- (1) *Gjoni* *është* *neurokirurg* *i* *shquar*.
John.NOM be.IND.PRS.3SG distinguished.NOM DET.NOM.M neurosurgeon
 ‘John is a distinguished neurosurgeon’
- (2) *Mbrëmë* *e* *takova* **Gjonin** *në* *teatër*.
 yesterday.evening CL.ACC meet.IND.AOR.1SG **John.ACC** in theatre.ACC
 ‘Last night I met John in the theatre’

- (3) *Bashkëshortja e Gjonit është mësuese*
 wife.DET.NOM DET.NOM.F **John.GEN** be.IND.PRS.3SG teacher.NOM
e gjuhës angleze.
 DET.NOM.F language.GEN English.F
 ‘**John’s** wife is a teacher of English’
- (4) *Unë ia dhashë Gjonit disa romane.*
 I.NOM CL.ACC.PL+DAT.SG give.IND.AOR.1SG **John.DAT** several.ACC
 novels.ACC.PL
 ‘I gave **John** several novels’ (adapted from Çanta 2017: 228)

3.1.2 Combination with simple prepositions

Cases are also used in combination with simple prepositions, to mark a wide range of semantic relations. In various classical and modern Indo-European languages, there are cases of distributed semantics (Sinha and Kuteva 1995) between case and preposition: a given preposition may then combine with two or more cases to express different semantic relations. This is what happens for instance when Classical Greek *pará* heads a noun phrase in the Genitive-Ablative, Dative-Locative or Accusative, or when Modern German contrasts *auf die_{ACC} Straße* ‘onto the road’ with *auf der_{ACC} Straße* ‘on the road’. A few Albanian prepositions may thus introduce noun phrases with different case endings, such as *ndaj* ‘towards’, (*për*)*tej* ‘beyond’ and *për* ‘for, about’. For instance, *ndaj* may be followed by a noun phrase in the Accusative (5) or in the Dative-Ablative (6).

- (5) *Ndaj të gdhirë, kur sollën telegramin*
 toward DET dawn.ACC when send.IND.AOR.3PL telegram.ACC
e fundit, ata ishin ende duke pirë.
 DET.ACC.SG end.DET.GEN they.NOM be.IND.AOR.3PL still ADV.GER² drink
 ‘Towards dawn, when they sent the last telegram, they were still drinking’
 (ANC, *Gjenerali i ushtrisë së vdekur*, Ismail Kadare, 1963)
- (6) *Disa nga ato projekte janë edhe*
 several.NOM from DEM.F.NOM.PL project.NOM.PL be.IND.PRS.3PL also

² Gerund-like adverbial form.

rezistenca e fillit sintetik ndaj zjarrit (...)
 resistance.NOM DET fibre.GEN synthetic toward fire.OBL
 ‘Some of these projects are also about synthetic fibre resistance to fire’
 (ANC, *Gazeta Shqiptare*, 2007.01.13)

Though there is a clear semantic difference between the two examples, it seems to be mostly linked to the nature of the noun (concrete vs abstract) rather than to the difference in case marking. In other cases, the semantic difference seems limited or even absent, as in (7–8).

(7) *Se ka Shqipëri edhe tej Drinit*
 DEM.PRO AUX.IND.PRS.3SG Albania still beyond Drin.OBL
 ‘That it is still Albania beyond the Drin’
 (ANC corpus, *Gjysmëfotografia, Ali Podrimja*, 1995)

(8) *tej Drinin*
 beyond Drin.ACC
 ‘beyond the Drin’

The use of preposition and case with distributed semantics thus seems impossible – or at least very limited – in Modern Albanian. As a consequence, in such contexts, the semantic relation seems to be mostly supported by the preposition; it could be said that, rather than distributed semantics, there is semantic agreement between preposition and case. Thus, many prepositions are systematically followed by noun phrases in an oblique form. Though it is formally a genitive-dative marking, Buchholz and Fiedler (1987) show that it should be considered as an ablative, given that pronouns introduced by these prepositions appear in the ablative (which, in pronouns, is formally distinct from the genitive-dative). Such prepositions include *afër* ‘near’, *drejt* ‘towards’, *gjatë* ‘during’, *larg* ‘far from’, *midis* ‘between’, *ndanë* ‘near’, *para* ‘in front of’, *prapa* ‘behind’, *prej* ‘from’, *ndaj* ‘towards’, *pranë* ‘next to, beside’, *pas* ‘after, behind’, *sipas* ‘according to’, *veç* ‘save’, etc., as in (9).

(9) *Qielli merr ngjyrë prej syve tu*
 sky.DET.NOM take.IND.PRS.3SG color.ACC.SG from eye.OBL.PL your.OBL.PL
 ‘The sky takes its color from your eyes’ (ANC, *Zëri*, 2013.03.16)

Other prepositions are followed by the accusative, such as *në* ‘in’, *mbi* ‘on, about’, *përmbi* ‘over’, *me* ‘with, by’, *më* ‘on’, *ndër* ‘between, among’, *n(d)ë* ‘in, into’, *nën* ‘under’, *pa* ‘without’, *nëpër* ‘through, across’ as in (10–11).

(10) *erdhi me të motrën*
 come.AOR.3SG with DET³ sister.ACC
 ‘he came with his sister’ (internet)

(11) *peshk pa halë s’ka*
 fish.NOM without fishbone.ACC NEG.exist.PRS.3SG
 ‘a fish with no bones, that doesn’t exist’ (internet)

A more uncommon feature of Albanian is that some prepositions are followed by noun phrases in the nominative. This is another example of distancing from the original IE combinations of preposition and case. Only two prepositions are concerned: *tek* (Gheg *ke*) ‘to, at (X’s place)’ and *nga* (Gheg *kah*) ‘from, since, because of’, as illustrated in (12–13).

(12) *shko tek unë!*
 go.IMP.2SG by me.NOM
 ‘go to my place!’ (internet)

(13) *sot nuk dola nga shtëpia*
 today NEG exit.AOR.1SG from house.NOM
 ‘today I didn’t go out’ (internet)

The use of the dative case after a preposition is rare (Agalliu et al., 2002, Çanta 2017: 234).

Thus, as a general rule, prepositions head noun phrases with one case only, most frequently the ablative, as indicated in Table 2.

3.2 Simple prepositions

3.2.1 High frequency items: Identifying functional prepositions

Within the word class of prepositions, across languages, there are often important differences between frequent items which have grammatical uses and are generally highly polysemous – so-called *functional* or *bleached* prepositions – and less frequent items which have a more consistent meaning component and are generally considered to be *lexical* prepositions. Taking into account close to a hundred

³ This is a possessive use of the determiner.

Table 2: Prepositions and case marking in Albanian.

Nominative	Accusative	Ablative
<i>nga</i> ‘from, since, because of’, <i>tek</i> ‘to, at’ [examples: <i>nga familja</i> (from/by the family), <i>nga e diela</i> (from Sunday), <i>nga perëndimi</i> (from the west), <i>tek unë</i> (to me/at my house), <i>te shkolla</i> (at school), <i>te fabrika</i> (at the factory), etc.]	<i>deri</i> ‘until’ <i>n(d)ë</i> ‘in, into’ <i>mbi</i> ‘on, about’ <i>përmbi</i> ‘over’ <i>me</i> ‘with, by’ <i>më</i> ‘on’ <i>ndër</i> ‘between, among’ <i>nën</i> ‘under’ <i>nëpër</i> ‘through, across’ <i>pa</i> ‘without’ <i>qysh</i> ‘since’ [examples: <i>në punë</i> (at work), <i>mbi tavolinë</i> (on the table), <i>përmbi çati</i> (on the roof), <i>nëpër oborr</i> (across the yard), <i>me kënaqësi</i> (with pleasure), <i>pa ndrojtje</i> (without coyness), etc.]	<i>afër</i> ‘near’ <i>brenda</i> ‘inside’ <i>drejt</i> ‘towards’ <i>gjatë</i> ‘along, during’ <i>jashtë</i> ‘outside’ <i>kundër</i> ‘against’ <i>larg</i> ‘far from’ <i>midis</i> ‘between, among’ <i>ndanë</i> ‘near’ <i>ndërmjet</i> ‘among’ <i>nëpërmjet</i> ‘by means of, through’ <i>para</i> ‘in front of’ <i>pas</i> ‘after, behind’ <i>poshtë</i> ‘below’ <i>pranë</i> ‘next to, beside’ <i>prapa</i> ‘behind’ <i>prej</i> ‘from’ <i>rreth</i> ‘around’ <i>sipas</i> ‘according to’ <i>veç</i> ‘save’ [examples: <i>ndaj miqve</i> (towards friends), <i>pranë mysafirëve</i> (next to the guests), <i>afër liqenit</i> (near the lake), <i>para shitores</i> (in front of the shop), <i>pas mësimit</i> (after class), <i>prej shtëpisë</i> (from the house), <i>nga lumturia</i> (from happiness), etc.]
	<i>ndaj</i> ‘towards’, <i>për</i> ‘for, about’ <i>(për)tej</i> ‘beyond’	

items listed either in grammars (e.g. Boissin 1975, Buchholz and Fiedler 1987) or papers dedicated to Albanian prepositions, and checking their frequency in two different corpora (OPUS and ANC), we were able to verify that Albanian is no exception, with stark contrasts in frequency between the most frequent items and the least frequent ones.⁴ While two prepositions have a relative frequency of more than 15,000 occurrences per million words (henceforth *pmw*) in both corpora (*në* ‘in, into’ and *për* ‘for, about’), a few items have a relative frequency under 5

⁴ Note that the frequencies given by the corpora should be taken with caution, since some items are highly polyfunctional, as is typical of prepositions in general.

pmw in both corpora (*tatëpjetë* ‘downward’, *tinëz* ‘furtively’ and *përposh* ‘below’, as well as the nouns *brinjë* ‘ribs’ and *vithe* ‘root’, with some uses as prepositions).

The five most frequent prepositions in the OPUS corpus are *në* ‘in, into’, *për* ‘for, about’, *me* ‘with, by’, *nga* ‘from, since, because of’ and *më* ‘on’; in this corpus, they have a mean relative frequency of 14,791 occurrences per million words, meaning that they roughly account for one tenth of the words in the corpus. There is evidence that this frequency is linked to the fact that they are indeed functional prepositions, i.e. that they have grammatical uses in Modern Albanian. For instance, *për* is frequently used as a composition element for new prepositions, and the same can be said of *në* (see the following section, 2.2.2, for examples). Another clue to their functional use is the fact that these frequent prepositions tend to be highly polysemous, as is the case of *për*, which can be glossed by ‘for’ but also ‘about’, ‘to’ or ‘on’, taking on semantic nuances of goal, argumentation, implicature, cause, means, etc.

Table 3: Highly frequent simple prepositions in the OPUS corpus.

<i>në</i> , <i>n(d)ë</i> + ACC ‘in, into’
<i>për</i> + ACC/ABL ‘for, about’
<i>me</i> + ACC ‘with, by’
<i>nga</i> + NOM ‘from, since, because of’
<i>më</i> + ACC ‘on’

3.2.2 Simple prepositions and compounds: A synchronic analysis

Besides these few very frequent and polysemous prepositions, Albanian has a wide range of (less frequent) simple prepositions. The frequency of these items ranges from roughly 1,500 (*mbi* ‘on, about’, 1,775 pmw in OPUS and 1,209 in ANC) to zero (a few prepositions listed by Boissin 1975 or Buchholz and Fiedler 1987 are absent from at least one of the two corpora). It is hard to distinguish classes, be it on the basis of their frequency, semantics or use-types. These ‘lexical’ prepositions include rather frequent items such as *mbi* ‘on, about’, *gjatë* ‘along, during’ or *pas* ‘after, behind’, etc. (with a mean of roughly 1,100 pmw), less frequent lexical prepositions such as *para* ‘in front of’, *nën* ‘under’ or *brenda* ‘inside’, etc. (with a mean of roughly 300 pmw) and rare items such as *ndërmjet* ‘among’, *poshtë* ‘below’ or *larg* ‘far from’, etc. (with a mean of roughly 30 pmw), as shown in Table 4.

Table 4: Relative frequency (pmw) of less frequent simple prepositions in the OPUS corpus (including non-prepositional uses).

rel. fq.	item & gloss
< 2,000	<i>mbi</i> + ACC 'on, about'; <i>gjatë</i> + ABL 'along, during'; <i>pas</i> + ABL 'after, behind'; <i>prej</i> + ABL 'from'; <i>deri</i> + ACC 'until'; <i>sipas</i> + ABL 'according to'; <i>rreth</i> + ABL 'around'; <i>tek</i> + NOM 'to, at'
< 1,000	<i>ndaj</i> + ACC/ABL 'towards'; <i>midis</i> + ABL 'between, among'; <i>kundër</i> + ABL 'against'; <i>përpara</i> + ABL 'before'; <i>pa</i> + ACC 'without'
< 500	<i>para</i> + ABL 'in front of'; <i>nën</i> + ACC 'under'; <i>brenda</i> + ABL 'inside'; <i>jashtë</i> + ABL 'outside'; <i>drejt</i> + ABL 'towards'; <i>nëpërmjet</i> + ABL 'by means of, through'; <i>pranë</i> + ABL 'next to, beside'; <i>përveç</i> + ABL 'except'; <i>ndërmjet</i> + ABL 'among'; <i>poshtë</i> + ABL 'below'; <i>larg</i> + ABL 'far from'; <i>tej</i> + ACC/ABL 'beyond'
< 100	<i>nëpër</i> + ACC 'through, across'; <i>prapa</i> + ABL 'behind'; <i>afër</i> + ABL 'near'; <i>përtej</i> + ACC/ABL 'beyond'; <i>përmes</i> + ABL 'through'; <i>qysh</i> + ACC 'since'; <i>përballë</i> + ABL 'opposite'; <i>tutje</i> + ABL 'beyond'; <i>veç</i> + ABL 'save'; <i>rrotull</i> + ABL 'around'; <i>sipër</i> + ABL 'above'
< 10	<i>përkrah</i> + ABL 'alongside'; <i>ndanë</i> + ABL 'near'; <i>ndër</i> + ACC 'between, among'; <i>përmbi</i> + ACC 'over'; <i>kundrejt</i> + ABL 'across, opposite'; <i>krahas</i> + ABL 'near, along with'
< 1	<i>tatëpjetë</i> + ABL 'downward'; <i>karshi</i> + ABL 'opposite' (ADV); <i>gjer</i> + ACC 'till, until'; <i>përposh</i> + ABL 'below'; <i>përpos</i> + ABL 'among'

A few items are used as prepositions but appear to be synchronically analyzable as nouns. These items appear to have a very low frequency, as shown in Table 5.

Table 5: Relative frequency (pmw) of nouns used as prepositions in the OPUS corpus (including non-prepositional uses).

rel. fq.	item & gloss
< 5	<i>buzë</i> + ABL 'near, on the edge of'; <i>majë</i> 'top'; <i>anës</i> + ABL 'side > near, beside, besides'; <i>përqark</i> + ABL 'around'
< 1	<i>byth</i> + ABL 'at the bottom of'; <i>brinjë</i> + ABL 'ribs'; <i>vithe</i> + ABL 'root > at the bottom of'

Table 6: Relative frequency (pmw) of sequences used as prepositions in the OPUS corpus (including non-prepositional uses).

rel. fq.	item & gloss
< 10	<i>me gjithë</i> + ACC 'despite (lit. with all)'; <i>gjer në</i> + ABL 'inside'; <i>për nga</i> + ABL 'for'; <i>në bazë</i> + GEN/ABL 'on the basis of'; <i>në buzë</i> 'alongside (lit. 'on the edge of)'; <i>me anën</i> + GEN 'near, alongside'
< 1	<i>fill e</i> + ABL 'straight at'

Finally, a few constructions appear to be marked as compounds, though they follow patterns common to syntactically simple (but morphologically complex) prepositions. Again, these items appear to have a very low frequency (Table 6).

A number of simple prepositions – besides the five most frequent ones, for which it is obviously the case – are synchronically opaque and thus appear to be morphologically simple. This is the case for instance of *afër* ‘near’, *jashtë* ‘out of’, *mbi* ‘on, upon’, *me* associative and instrumental, *ndër* ‘in (with motion), amid’, *pa* ‘without’, *tej* ‘beyond’, *brenda* ‘inside’, *kundër* ‘against’. However, a number of apparently simple prepositions are actually synchronically transparent compounds; for instance, as mentioned above, the simple prepositions *në* and *për* are recognizable as elements of the following prepositions: *përballë* ‘opposite’, *përmbi* ‘over’, *përpara* ‘before’, *përpos* ‘among’, *përposh* ‘below’, *përqark* ‘around’, *përtej* ‘beyond’, *nëpër* ‘through, across’, *nëpërmjet* ‘by means of, through’.

The semantic determination of such items seems to lie mostly with the second item, given the polysemy of *për* and *në* and the fact that the compounds seem semantically equivalent to the corresponding simple forms. There may be semantic and/or functional differences between the two elements of such pairs, as in the case of *për* ‘for, about’ / *nëpër* ‘among’, *ballë* ‘in front of’ / *përballë* ‘opposite’, *tej* ‘beyond’ / *përtej* ‘beyond, across’, *kundër* ‘against’ / *përkundër* ‘despite’, *para* ‘in front of, before’ / *përpara* ‘before’. However, even in such cases it is hard to pinpoint the semantic contribution of the first element of the compound, and in some cases, as with *mbi* and *përmbi* ‘over’ or *veç* and *përveç* ‘except’, there seems to be no clear semantic difference between the two elements of the pair.

Sometimes, only one element of the pair is attested in the corpus with prepositional uses. It is therefore impossible to speak of a regular construction, e.g. [Prep: ‘meaning_i’] → [*për* Prep: ‘meaning_{i/j}’], be it with an associated semantic change (i→j) or not (i=i). On the contrary, this type of compounding is apparently associated with various possibilities, from both semantic and functional points of view. For instance, while *përposh* ‘below’ has prepositional uses, *posh* ‘back’ is apparently used only as a noun. The case of *qark* and *përqark* ‘around’ is even more complex, since the latter is well attested with prepositional uses (14), while *qark* appears regularly in a complex sequence *rreth e qark* ‘all around’ (lit. ‘around and in a circle’) (15).

- (14) *Edhe toka përqark shtëpive ishte*
yet earth.DET.F.NOM.SG around house.PL.OBL.DEF be.IND.IMP.3SG
e njëjta.
DET same.F.NOM.SG
‘Even the earth around the houses was the same’
(ANC, *Natyra e natyrës*, Faruk Myrtaj, 1984–2009)

- (15) *Lehnin edhe qentë e kullave pranë përroit, lehnin qentë rreth e qark katundit...*
 bark.IND.IMP.3PL also dogs.NOM DET.NOM.PL old.towers.GEN near
 stream.DET.OBL bark.IND.PRS.3PL dog.NOM around DET circle village.DET.OBL
 ‘Dogs from the old towers also barked near the stream, dogs barked all
 around the village...’ (ANC, Ritet e Nishanes, Kim Mehmeti, 2004)

Besides, for some simple prepositions, such as *pas* ‘behind, after’, there does not seem to be a “*për*” compound (*përpas* is unattested as a preposition in the ANC corpus).

There are slightly more complex cases, such as that of *brendapërbrenda* ‘inside’, compound with three perfectly analyzable elements: *brenda* + *për* + *brenda*, i.e. the preposition *për* and the preposition *brenda* ‘inside’; this compound is marked as formal and quite rare (52 occurrences in the ANC corpus). Another similar case is that of *ndërmjet* ‘between, among’, attested early on (in Buzuku, 16th c.) and *nëpërmjet* ‘through’. In both compounds, the first element is a preposition, simple (*ndër*) or compound (*nëpër* = *në* + *për*), and the second is *mjet* ‘until, amid’, an archaic preposition (Demiraj 1997: 275), still used in dialects or as a noun. In this case, synchronically, there is thus a clear functional and semantic difference between the simple form *mjet* ‘means’, used only as a noun (16), and the compounds *ndërmjet* ‘between, among’, *përmjet* and *nëpërmjet* ‘by means of, through’ (17), used as prepositions.

- (16) *Edhe muzika mund të bëhet një mjet*
 also music can.IND.PRS.3SG COMP become.IND.PRS.3SG one means
torture.
 torture.GEN
 ‘Music can also become an instrument of torture’
 (ANC, *Gazeta Shqiptare*, 2007.01.05)

- (17) *Zëri i Çekut i erdhi nëpërmjet telefonit nga Tirana.*
 voice.DET DET Çeku.GEN CL come.IND.AOR.3SG through
 phone.OBL from Tirana.NOM
 ‘Çeku’s voice came through the phone from Tirana’
 (ANC, Rrënojat e iluzioneve, in *Kështu ecën njeriu*, Kim Mehmeti, 2007)

Other items which could be synchronically analyzed as compounds are less transparent, and we count them as simple prepositions. This is the case for *kundrejt* ‘in front of’, which could be analyzed as a compound of *kundër* ‘against’ with *drejt*

‘towards’, or with a suffixal element *-ej* (on *andej* or *tej* ‘beyond’). Similarly for *matanë* ‘along’ (*mat* ‘bank’ + *anë*), *mbanë* ‘along’ (*mbi* *anë*) and *andej* ‘beyond’ (*anë* + *tej* ‘beyond’), all including an adverbial form of *anë* ‘side, margin’.

3.3 Complex sequences

Some simple prepositions (*me*, *më*, *në*, *nga*, *para*, *për*, *prej*) may be preceded by various items: *sa* (indefinite quantifier), *që* (complementizer, here with an inchoative meaning ‘since’), *gjër* (adverbial, meaning ‘until’; Gheg *deri*) and *ngjithë* (participle of *ngjit* ‘to attach’). The resulting sequences include the following:

- (18) *sa për mua*
 QUANT for me.ACC
 ‘as far as I am concerned’
- (19) *që në mëngjes*
 since in morning.ACC
 ‘since the morning’
- (20) *Kryeministri u shpreh se një gjë e tillë është përcaktuar në Kodin Zgjedhor me konsensus që në vitin 2004.*
 Prime.minister CL.REF be.manifest.IND.AOR.3SG COMP one thing DET
 such be.IND.PRS.3SG determined in code.ACC electoral
 with consensus since in year.ACC 2004
 ‘The Prime Minister said that this was set by consensus in the Electoral Code since 2004’ (ANC, Zëri, 2013.07.17)
- (21) *gjër në Durrës*
 until in Durrës.ACC
 ‘all the way to Durrës’
- (22) *gjër para disa ditëve*
 until before some day.OBL.PL
 ‘until a few days ago’
- (23) *ngjithë me shtëpinë*
 attach.PTCP with house.DET.ACC
 ‘right by the house’

Some sequences, though they are formally similar, should not be included in this list, for instance *fill e në* (24), listed as a complex preposition by Boissin (1975: 206). In such uses, the confirmative meaning of *fill* ‘surely’ (25), combined with the conjunction *e* ‘and’, result in an idiomatic, emphatic sequence.

(24) *shkoi fill e në shtëpi*
 come.PRET.3SG directly DET in home
 ‘he came home directly’

(25) *jam fill vetëm*
 be.PST.1SG absolutely alone
 ‘I am completely alone’

The frequency of such constructions is hard to establish with corpora, however, given the multifunctionality of the first element; for instance, the sequence *që në* is very frequent in the ANC corpus (over 24,000 hits), but much less frequently followed by a noun (less than 6,000 hits), and even then, *që* often seems to retain its use as a complementizer or relative pronoun.

4 Complex prepositions: A corpus study

Complex prepositions are generally not well described in grammars, and rarely the object of dedicated studies. The grammars of Boissin (1975) and Buchholz and Fiedler (1987) do provide lists of complex prepositions, but do not discuss the existence and productivity of patterns such as those found in most European languages, e.g. [Prep (Det) N Prep]. The lists include sequences such as *për nder të* ‘for the sake of (lit. *for honor of*)’, *në anë të* ‘at the end of (lit. *in end of*)’, *në emër të* ‘in the name of (lit. *in name of*)’, *me anë të* ‘by means of (lit. *with side of*)’, etc. These sequences include near synonyms, for instance *në faqe të* and *në fytyrë të* ‘on the surface of’, *me anë të* and *në ndihmë të* ‘by means of’, *në pikë të* and *në pisk të* ‘at the break of (e.g. dawn)’. In all such sequences, the third element of the sequence is the proclitic linking article, a genitive marker (cf. section 2.1), here in the undetermined form.

In this section, our aim is to check the existence and frequency of the complex prepositions listed in the literature, but also to uncover patterns of CPs and evaluate their productivity. In order to do this, we relied on two corpora, the ANC Corpus and the OPUS corpus, as described in (3.1.).

4.1 Methodology

The Albanian National Corpus is a relatively large corpus of fiction and journalistic texts, containing approximately 20 million tokens. There is a “user-friendly morphological markup consisting of tags assigned to individual tokens”, which is quite reliable (ambiguous tokens may be associated to multiple tags; in some cases, no analysis is provided) and can be used in queries. Indeed, queries can be based on word forms, lexemes or grammatical and morphological attributes, and multi-word queries are also possible. This was very helpful for our corpus study.

The OPUS corpus (in its second version, i.e. OPUS2) is a parallel corpus with data from 40 languages (Tiedemann 2012). Actually, it is a set of text corpora with sentence alignment. We used only the Albanian section of the corpus, which contains roughly 40 million tokens. There is no POS tagging for Albanian, but the query interface remains very useful, since both strings and CQP expressions can be used for queries.

Our first step was to check both corpora for the presence of all sequences listed in the literature, with targeted queries. For instance, in order to check the frequency of the sequence *në bazë të* ‘on the basis of’, we ran the following queries: (a) on the ANC corpus, lexeme “në” + lexeme “bazë” + wordform “të”; (b) on the OPUS corpus, [word=“(n|N)ë”] [word=“bazë”] [word=“të”]. The results then had to be inspected for the presence of noise, i.e. of formally identical but functionally different sequences. For instance, the search for *në vend të* ‘in place of’ (26) also yields examples which are clearly not prepositional, for instance when the following word is an adjective, as in *në vend të parë* ‘in first place’.

- (26) *Kam parë gjithashtu nën diell që në*
 have.IND.PRS.1SG see.PST.PTCP also under sun CONJ in
vend të arsyes kishte paudhësi, që
 place DET reason.GEN have.IND.IMP.3SG iniquity, CONJ
në vend të drejtësisë kishte paudhësi.
 in place DET justice.GEN have.IND.IMP.3SG iniquity
 ‘I have also seen under the sun that there was iniquity instead of reason,
 iniquity instead of justice’. (ANC, Dhjata e vjetër, Predikuesi,
 përkthimi i Shoqërisë Biblike e Shqipërisë (Albanian Bible Society), 1994)

Once we had the frequencies of all sequences, our second step was to identify patterns of CPs, such as [Prep N Det], as well as subpatterns. For instance, in order to check the frequency of the pattern [*në N të*], we ran the following queries: (a) on the ANC corpus, lexeme “në” + [grammatical attribute = Noun] + wordform “të”; (b) on the OPUS corpus, [word=“(n|N)ë”] [] [word=“të”]. However, given the

limitations of the ANC interface, which does not permit the retrieval of all occurrences, for this step of our investigation we mainly used the OPUS corpus. We thus identified a long list of potentially prepositional sequences, which then had to be checked one by one, since there was a large amount of noise, with sequences such as *në qytete të* ‘in the city of’, which is obviously not a complex preposition.

The final step was to check the existence of formal variation, for all sequences identified as potential CPs. Various features are associated with the grammaticalization of CPs, i.e. with the transition from free syntax to CA, as noted by various authors (e.g. Lehmann 1985, Adler 2001, Hoffmann 2005). Some important factors are the tendency to lose the morpho-syntactic features usually expected for the word class of each item of the sequence (*decategorization*, cf. Hopper and Traugott 2003), the loss of syntactic variability, and the tendency towards greater semantic bleaching (*desemantization*, Hopper and Traugott 2003). The fact that the noun in [Prep N *të*] sequences does not appear in its definite form is a clue that there is indeed some measure of decategorization. For instance, the sequence *në drejtim të* ‘towards, in the direction of’ almost exclusively appears in the corpus as such (27); there are only four occurrences of the expected form, *në drejtimin të* ‘in the direction (lit. direction.DET) of’.⁵

- (27) *Mbi qiellin e Zajazit do të kalonte*
 over sky.ACC DET Zajaz.GEN PRT⁵ PRT pass.IND.IMP.3SG
një skuadër aeroplanësh luftarak, që fluturonte
 one squadron airplane military PRO fly.IND.IMP.3SG
në drejtim të malit Drenovë.
 in direction DET mountain.GEN Drenova
 ‘Over the sky of Zajaz would pass a squadron of warships flying in the direction of Mount Drenova’ (ANC, Koha.mk, 2012.01.03)

In order to identify the first two factors, we checked for the presence of variants of all sequences, with targeted queries, for instance, for *në bazë të*: (a) in the ANC corpus, lexeme “*në*” + lexeme “*bazë*”; (b) in the OPUS corpus, [word=“(n|N)ë”] [word=“*bazë*”]. Semantic bleaching is harder to identify; we only excluded sequences which seemed to display absolutely no trace of semantic extension, for instance if the noun always retains its original, referential meaning; this seems to be the case, for instance, of *në mënyrë të* ‘in the manner of’, which is frequent but referential and often used with a modifier (28).

⁵ *Do* is a particle used to form the future tense; *të* is used here as a particle to form the subjunctive.

- (28) *Por ai ishte i vetmi që pati*
 but PRO.3SG be.IND.IMP.3SG DET only PRO have.IND.AOR.3SG
guximin ta refuzojë në mënyrë të qetë dhe
 courage.ACC COMP refuse.SBJ.PRS.3SG in manner DET calm and
dinjitoze ta braktisë besimin e vetë.
 dignified COMP abandon.SBJ.PRS.3SG faith DET PRO.3SG
 ‘But he was the only one who had the courage to refuse in a calm and dig-
 nified fashion to abandon his faith’ (ANC, *Spastrimet etnike. Politika gjenocidale serbe ndaj shqiptarëve* (Shqyrtime, komente, elaborate, dokumente) Hivzi Islami 2000–2001)

4.2 Results

Our analysis led us to exclude many sequences corresponding to the pattern [Prep N *të*], which showed no evidence of functional equivalence with prepositions and/or no sign of grammaticalization. Quite a few sequences listed in the literature were either not attested or not found with prepositional uses in our corpora. We thus eliminated some sequences, and identified others as CPs (Table 7). Only twenty have a relative frequency over 10 (pmw). The proportion of prepositional uses is not easy to pinpoint, though, given the formal similarity with other uses; this means that the relative frequency is only indicative.

Table 7: PNP sequences (i.e. potential CPs) and their frequency in the two corpora (relative frequency, pmw).

sequence	ANC (rel. fr.)	OPUS (rel. fr.)
<i>në afërsi të</i> ‘in the vicinity of, near’	41	4
<i>në bazë të</i> ‘on the basis of’	124	15
<i>në dëm të</i> ‘to the detriment of’	15	5
<i>në drejtim të</i> ‘towards, in the direction of’	87	7
<i>në fillim të</i> ‘at the beginning of’	88	82
<i>në interes të</i> ‘in the interest of’	22	18
<i>në kërkim të</i> ‘in search of’	25	10
<i>në kuadër të</i> ‘in the framework of’	130	5
<i>në mbrojtje të</i> ‘in defense of’	15	4
<i>në mungesë të</i> ‘in the absence of’	25	4
<i>në nderim të</i> ‘in honor of, in tribute to’	2	4

Table 7 (continued)

sequence	ANC (rel. fr.)	OPUS (rel. fr.)
<i>në përgjigje të</i> 'in response to'	2	11
<i>në periferi të</i> 'on the periphery of, near'	8	9
<i>në prag të</i> 'on the verge of, near'	35	13
<i>në pritje të</i> 'in anticipation of'	24	16
<i>në rast të</i> 'on the occasion of'	18	11
<i>në shkelje të</i> 'in violation of, in breach of'	0	5
<i>në var(t)ësi të</i> 'depending on' (depending on)	7	12
<i>në vend të</i> 'in place of'	90	90
<i>në anë të</i> 'at the end of'	5	1
<i>në emër të</i> 'on behalf of, in the name of'	82	26
<i>në favor të</i> 'in favor of'	34	39
<i>në fund të</i> 'at the end of'	182	291
<i>në (sh)këmbim të</i> 'in exchange for'	8	17
<i>në krye të</i> 'at the end (lit. head) of'	83	88
<i>në mbështetje të</i> 'in favor of, in support of'	12	35
<i>në mes të</i> 'in the middle of'	141	64
<i>në përfitim të</i> 'in exchange for, in return for'	0	8
<i>në përputhje me (/të)</i> 'in compliance with'	43	20
<i>në pikë të</i> 'in the middle of'	3	0
<i>në qendër të</i> 'at the center of'	49	50
<i>në zemër të</i> 'at the heart of'	7	5
<i>në lidhje me + ACC</i> 'in connection with'	118	175
<i>me përjashtim të + GEN</i> 'except'	24	27
<i>për nder të + GEN</i> 'for the sake of'	8	2
<i>mbi kokë të + GEN</i> 'at the bedside of'	1	0

Some of these sequences are relatively frequent and display a low proportion of occurrences with modification. The fact that the noun *may* appear with a referential meaning, as in (29), does not preclude a given sequence from having uses as a CA (30).

- (29) *E në atë kohë ajo parcelë nuk ka*
 PRT in that.ACC time.ACC that.NOM parcel NEG AUX.IND.PRS.3SG

genë në emër të 'Bujqësisë'

be.PTCP in name DET Agriculture.GEN.DET

'At that time, that parcel was not in the name of "Agriculture"'

(ANC, *Zëri*, 2013.09.10)

- (30) *Në emër të redaksisë së javores «Fokus», para kolegëve dhe miqve të pranishmëve foli bashkëpunëtori i tij në këtë gazetë, Shase Dimevski, i cili evokoi momente nga puna e përditshme me të.*
 in name DET redaction DET weekly Focus in.front.of colleague.OBL.DET.PL and friend.OBL.DET.PL DET present.OBL.PL speak.IND.AOR.3SG colleague.NOM.SG DET PRO in this.ACC.SG newspaper, Shase Dimevski, DET PRO evoke.IND.AOR.3SG moment.ACC.PL from work DET daily with PRO

'On behalf of the weekly magazine 'Focus', in front of the present friends and colleagues, spoke his co-worker in this newspaper, Shase Dimevski, who evoked moments from their daily work with him'

(ANC, *Koha.mk*, 2013.03.28)

As in other languages, the use of names of body parts (e.g. *zemër* 'heart' in *në zemër të* 'at the heart of', *krye* 'head' in *në krye të* 'at the end (lit. head) of') (31) and relational nouns (e.g. *mes* 'middle' in *në mes të* 'in the middle of', *qendër* 'center' in *në qendër të* 'at the center of') (32) is quite frequent in the (limited) paradigm of Albanian CPs.

- (31) *Tash disa vite mbretëron një klimë ndërtimesh dhe mbindërtimesh pa leje mu në zemër të qytetit*
 now several.ACC.PL year.ACC.PL reign.IND.PRS.3SG one climate.NOM construction.OBL.PL and over.construction.OBL.PL without permission.ACC until in heart DET city.OBL.DET
- 'Now reigns right in the heart of the city a climate of construction and illegal over-construction'
- (ANC, *Zëri*, 2013.10.27)⁶

6 *Mu* is glossed 'bis' (i.e. 'until') by Meyer (1891), who indicates that it is a preposition.

- (32) *Stavreski, ka qenë të hënën në qendër të Stavreski AUX.IND.PRS.3SG be.PTCP DET Monday in center DET akuzave të opozitës, të cilët accusation.OBL.PL.DET DET opposition.GEN DET PRO iniciuan votëbesim për punën e tij. initiate.IND.IMP.3SG vote.of.confidence.acc for work.ACC DET PRO*
 ‘Stavreski was, on Monday in the center of accusations from the opposition, who launched a vote of confidence for his work’ (ANC, Koha.mk, 2012.05.29)

But there are also quite a few CPs with an abstract noun as lexical item, for instance *mbështetje* ‘support’ in *në mbështetje të* ‘in favor of, in support of’, which seems to appear frequently or even mostly in very formal registers (33).

- (33) *Në mbështetje të nenit 16 të ligjit nr. 7491, in support DET paragraph.OBL 16 DET law.OBL number 7491 datë 29.4.1991 «Për dispozitat kryesore date.ACC 4.29.1991 about disposition.ACC main.ACC kushtetuese», propozim të Këshillit të constitutional.ACC proposal.NOM DET council.OBL DET Ministrave, Kuvendi popullor i Republikës minister.OBL.PL convention.NOM popular.NOM DET Republic.OBL së Shqipërisë vendosi: DET Albania.OBL decide.IND.AOR.3SG*
 ‘Following Article 16 of law # 7491 of 4.29.1991 ‘On the Main Constitutional Provisions’, on the proposal of the Council of Ministers, the People’s Assembly of the Republic of Albania decided (...)’ (ANC, *Kodi civil i Republikës së Shqipërisë*, Ligj nr. 7850, datë 29.7.1994)

In some cases, a [Prep N të] sequence appears to be part of idiomatic expressions, as with *në pikë të* ‘in the middle of’, cf. *në pikë të fundit* ‘at the very end’, *në pikë të dimrit* ‘in the dead of winter’ (34).

- (34) *Komuna e ka gjetur zgjidhjen që commune PRT AUX.IND.PRS.3SG find.PTCP solution.ACC.DET COMP nuk e lë më të varur nga NEG DET leave.IND.PRS.3SG more DET dependent from ujësjellësi i cili reduktimet nuk na i water.line.OBL in which reduction NEG CL.1PL.OBL CL.ACC*

of the nucleus and the corpus in which they appear: CPs with a spatial or temporal noun as a nucleus are typically much more frequent in the ANC corpus (Figure 1).

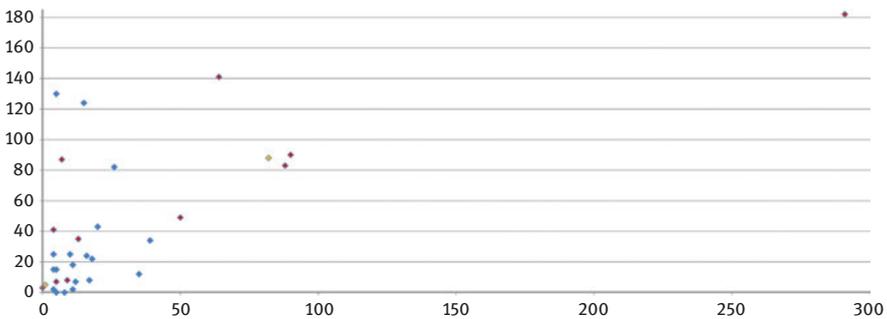


Figure 1: Relative frequency (pmw) of CPs in the ANC (on the horizontal axis) and OPUS (vertical axis) corpora, depending on the semantic nature of the nucleus: spatial (red), temporal (yellow), other (blue).

This could be a function of language contact: given that the OPUS corpus contains mostly translations, there is probably some amount of translation bias, which could lead to the overrepresentation of sequences which appear under the influence of the source languages, which are probably mostly European languages with CPs such as English, German or French. If this is true, our data would constitute an illustration of the tendency for CPs to appear in European languages as a result of language contact (cf. Van der Horst 2013, Hüning 2014).

5 Diachrony

From a diachronic perspective, Albanian is both fascinating – as a rather isolated branch of Indo-European (henceforth IE) – and highly complex – given the amount of language contact at various periods of its (pre)history. Concerning case marking and prepositions, Albanian has expectedly drifted away somewhat from IE, and now has (a) a reduced case-marking system and (b) a renewed paradigm of simple prepositions (with compounds but also new items resulting from grammaticalization phenomena), albeit in both cases with important traces of its IE origin. The rise of complex prepositions, for which we have not yet been able to identify the period of emergence, could be seen as a third step away from IE. These four paradigms – case markers, simple, compound and complex prepositions – constitute stops along the way from completely opaque grams

(cf. Svorou 1994) to partly or completely transparent sequences. Thus, case markers are completely opaque; simple prepositions are mostly opaque but do retain some semantic saliency, as they are sometimes also used as preverbs; compound prepositions are partly motivated, as their composition is often accessible to speakers; finally, in this respect, complex prepositions display a wide range of variation: some are rather opaque, others rather transparent (in the latter case, the main clue that they are complex prepositions is their syntactically erratic behavior, i.e. the absence of definite article: *për nder të* ‘for the sake of (lit. for honor of)’ vs **për nderin e* (for honor.DET of)). In this section, we provide an account of the evolution of these four subtypes of Albanian markers of nominal relations, focusing on problems of etymological reconstruction and paradigmaticization effects (Lehmann 1995 [1982]).

5.1 Case marking: Moving away from the Indo-European case-marking system

5.1.1 Loss of the locative case, partial fusion of Dative and Ablative

We described the case system of Standard Modern Albanian above (Section 2.1). Though there are traces of a locative case, it can no longer be considered an element of the paradigm. It seems to have been in use in Old Albanian, and can still be identified in literary texts (mostly Gheg) of the 16th-17th centuries, as a specific *-t* ending in the singular (Buchholz-Fiedler 1987: 242–243, comment 24). It is still sporadically in use in dialects, for prepositions generally requiring the accusative (except *ndër* and *pa*), but has otherwise not survived in the literary standard which appeared in the course of the 20th c.

For nouns, various cases have merged into what we called (after Boissin 1975) an *oblique* case, corresponding to the dative and ablative (for pronouns): genitive, ablative, dative, instrumental and locative.⁷ There are few traces of this origin in synchrony, as this evolution seems to have occurred in the prehistory of Albanian, with an important renewal of case marking before the first texts.

⁷ This overlap is also found in synchrony: Çanta (2017: 234) notes for instance that “the genitive case with the function of indicating the origin is sometimes interchangeably used with the ablative case”.

5.1.2 Loss of contrast between cases

As we noted in Section 2.1.1., the use of two cases with one preposition is no longer possible, or at least not productive. However, there are traces of this alternation in ancient texts, and it seems to have been a feature of Old Albanian. For instance, the preposition *në* ‘in(to)’ was apparently used with both accusative and locative (i.e. the result of merging between locative and instrumental⁸), as in *në jetët tonë* ‘in our life’ (Buchholz-Fiedler 1987: 243). This may have entailed a semantic difference, as in Latin *in* + accusative ‘into’ / + ablative ‘in’. However, Buzuku (LXXXV, 16th c.) already has the accusative in static examples, as in (37) (= Modern Albanian *ndë shtëpi të Atit tim*, Latin *in domo Patris mei*, Jn 14,2).

- (37) *endeh stepii teh atit tim*
 in house.ACC DET father.DET.OBL my.OBL
 ‘in my father’s house’

Some remnants of this alternation can be found in lexicalized sequences. For instance, the locative preposition *më* can now only take complements in the accusative, but is found with the ablative case in fixed constructions (38–39).

- (38) *e marr më sysh*
 PRO.M.ACC grasp.PST.1SG with eyes.ABL
 ‘I cast the evil eye on him’

- (39) *pres më dysh*
 cut.PST.1/2SG. in two.ABL
 ‘I (or you) cut in two’

⁸ This case is called Ablative in Buchholz and Fiedler (1987), Oblique in Boissin (1975). The singular and plural forms of personal and demonstrative pronouns take a specific ending, distinct from other oblique cases (genitive and dative). In the plural, they can also take another specific ending, similar to that found in the indefinite declension of nouns and adjectives; in that case, though, this ending almost always alternates with the oblique ending *-ve*, which is reminiscent of the ending *-vet* found in the definite declension of nouns, and is semantically close to the ablative: *një mur gurësh* ‘a wall made of stone (lit. a wall stone.ABL.PL)’. With prepositions, both endings are found: *thellësi prej gjashtë mijë metrash* ‘deep 6,000 meter.ABL.PL’, *njëra prej çështjeve* ‘one from problem.ABL.PL’. From an etymological perspective, the plural ablative ending is a continuation of the Indo-European plural locative ending, *-si*.

Another trace of this construction can be found with the highly polysemous preposition *për*, which can be used with the oblique case; it then takes on a specifically instrumental meaning (40–42).

(40) *doli për dere*
 exit.AOR.3SG to door.OBL
 ‘(s)he came out through the door’

(41) *e kapi për flokësh*
 PRO.M.ACC catch.AOR.3SG by hair.OBL
 ‘(s)he caught him by his hair’ (Buchholz-Fiedler 1987: 376)

(42) *e kap për dore*
 PRO.M.ACC take.PRS.3SG by hand.ACC
 ‘(s)he takes him by the hand’ (Kokona 1977, s.u. *për*)

This use of *për* is noted as dated in Boissin (1975: 204), who explains that the use of *për* with different cases may be linked to its complex etymology, as the convergence between different IE prepositions (Gr. *perí*, Lat. *per* and Gr. *pró*, Lat. *pro*, cf. Demiraj 1997: 317), and/or to the direct influence of Lat. *per* and its Italian, Dalmatian and Aromanian avatars.

5.1.3 An innovative nominative

A striking feature of Albanian prepositions – and another clear innovation with respect to IE – is the existence of a few items which take a nominative complement, as noted in Section 2.1.2. This is a typologically rare feature, which is linked to the fact that these items are originally – and partly remain – conjunctions. As such, they may introduce a clause with a verb, and its subject naturally stands in the nominative; when the verb is left unexpressed, the construction is thus superficially equivalent to a [preposition + noun phrase] sequence.⁹ The diachronic sequence thus appears to be the following: (main verb +) (i) **conjunction** + (NP_{SUBJECT} + verb)_{CLAUSE} > (ii) (main verb +) **conjunction** + NP_{SUBJECT} (+ verb) > (iii) (main verb +) **preposition** +

⁹ As e.g. with Spanish *según* in *según yo* ‘according to me.NOM’ (vs for instance *contra mi* ‘against me.ACC’). In Spanish, for that reason, some grammarians consider that *según* is not a preposition in its own right (see e.g. RAE 2009:2229 (II)).

NP_{COMPLEMENT} with a reanalysis between phases (ii) and (iii) (43–44). A full account of this evolution can be found in Petit (2015).

(43) *shko tek unë (banoj)*
 go.IMP.2SG where I.NOM live.PST.1SG
 ‘you came to my house (/ to where I live)’

(44) *sot nuk dolla nga <është> shtëpia*
 today NEG exit.AOR.1SG whence <be.PST.3SG> house.DET.NOM
 ‘today I went out of my house (/ of where my house is)’

5.2 Simple prepositions and compounds

5.2.1 A few Indo-European prepositions

Among the paradigm of simple prepositions in Albanian, quite a few are of IE origin, many prepositions having cognates in Greek and Latin. Thus, *mbi* can be traced back to IE **H₂mb^hi* (cf. Lat. *ambi*, *am*) or **H₁epi* (cf. Gr. *epi*) (cf. Demiraj 1997: 261), *jashtë* seems to be a (possibly suffixed) variant of IE **H₁egs* (cf. Lat. *ex*, Gr. *eks* ‘out (of)’, Demiraj 1997: 42). The case of *në* / *ndë* is slightly more complex, as it could be related to Lat. *in* (IE **H₁en*) and its reinforced variant *endo* (found in archaic and Republican texts) or traced back to a form **entí* (cf. Demiraj 1997: 331). In that case, *ndë* would be seen as an articulatory reinforcement of the initial phoneme of a proclitic word. This phenomenon can be seen in the pairs *pas* / *mbas* ‘behind’ and *prapa* / *mbrapa* ‘in front’. Also slightly complex, *brenda* ‘inside’ comes from an ancient compound, perhaps *për* + **entí*. Another example of compounding at an ancient period is *afër*, though the details of its composition are not obvious (they probably include the IE privative prefix, cf. Çabej 1976: 29). The dialectal preposition *nënë* ‘among’ (not related to *nën* ‘under’), in which speakers spontaneously perceive a reduplication of *në* ‘in(to)’,¹⁰ could be a variant of *ndanë* ‘next to’, itself the result of a compound (*n(d)ë* + *anë*, lit. ‘in side/flank of’). The last of this series is *kundër* ‘against’, also IE, but borrowed from Lat. *contra* ‘against’ (directly or indirectly, cf. Bonnet 1999: 358).

¹⁰ This was also the idea of the great Albanian specialist Norbert Jokl (1877–1942), as reported by Sandfeld (1930: 158).

5.2.2 An important paradigm renewal: P/Adv and P/N

Besides these prepositions with IE origins, Albanian has many prepositions which take the oblique case, and are presumably the result of more recent grammaticalizations from adverbs and nouns (typically spatial, esp. relational nouns).

Among the adverbs, some have IE origins, as is the case of *afër* ‘nearby / near’ (*afër teje* ‘near you’ / *banoj fare afër* ‘to live nearby’), *jashtë* ‘outside / outside of’ (*jashtë kufijve* ‘beyond the boundaries’ / *fle jashtë* ‘to sleep outdoors’), *brenda* ‘in / inside’ (see previous section). The existence of adverbial and prepositional uses is found for many items, among which *drejt* ‘straight ahead / towards’, *falas* / *falë* ‘free / thanks to’, *gjatë* ‘a long time / for the whole duration of’, *krahas* ‘side by side / near’, *larg* ‘far away / far from’, *sipër* ‘up / above’, *tutje* ‘far / beyond’, *veç* ‘apart / except’. In such cases, the question of which use came first – preposition or adverb – has been much debated. Though, as a rule, the grammaticalization of adverbs into prepositions is expected rather than the reverse, the (already mentioned) difficulties of Albanian reconstructions¹¹ rule out the possibility of a conclusive analysis (see e.g. Julia 2010 for the case of *kundër* and Latin equivalents).

There is also a series of nouns with uses as prepositions, in which they are followed by an NP in the oblique (Table 8).

Table 8: Nouns used as prepositions in Modern Albanian.

Item	Meaning (as noun)	Preposition (meaning when followed by an oblique NP)
<i>ballë</i>	front	in front of
<i>bri</i>	horn	in front of
<i>buzë</i>	lip	at the edge of
<i>byth</i>	bottom	at the bottom of
<i>majë</i>	tip	on top of
<i>rrëzë</i>	base	at the foot of

A marginal case is *tinëz* ‘furtively’ (based on *tinzi* ‘dissimulation’), used as an preposition (*tinëz meje* ‘unbeknownst to me’, Boissin 1975: 205) but also attested as an adverb (*largohej tinëz* ‘I took French leave (lit. *leave.PST.1SG furtively*)’, cf. Kokona 1977). Another marginal case is *vithe* ‘rump’, only used in a fixed construction *vithe kalit* ‘on horseback’ (Boissin 1975). Similar cases are that of *rreth* ‘ring, circle’, which is used both as noun and adverb (meaning ‘around’), and

¹¹ See e.g. *jashtë*, etymologically related to Latin *ex* but with a final element *-të* which is difficult to account for precisely.

forms part of the sequences *rreth e qark*, *rreth e rrotull* ‘around (lit. ‘round and (in) round’).

In all cases of prepositional uses of a noun, the fact that they stand in the indefinite form (45–46) is a sign of decategorization, and thus shows that there has been some degree of grammaticalization.

(45) *majë malit është një pishë*
 at.the.top mountain.NOM.DET be.PST.3SG a pine.NOM
 ‘there is a pine atop the mountain’

(46) *baxho i vërtetë rri rreth lopësh*
 shepherd.NOM DET real.NOM.M.SG stay.PST.3SG around cows.ABL
 ‘a real shepherd stays with his cows’

An important difference between these two subtypes, adverbs vs nouns used as prepositions, is that only adverbs may be found without a complement (47–50).

(47) *banoj afër teje*
 live.PRS.1SG near you.OBL
 ‘I live near you’

(48) *banoj afër*
 live.PRS.1SG near
 ‘I live nearby’

(49) *isha ballë një muri*
 be.IMP.1SG in.front.of a wall.ACC
 ‘I was in front of a wall’

(50) *?isha ballë*
 *be.IMP.1SG front.ACC

5.2.3 *për* + N/Adv in compound Ps

Beyond the fact that the reinforcement of adverbs and prepositions is a widely attested phenomenon, at least in IE languages (see e.g. the case of Romance, cf. Lehmann 2019, and more generally the other chapters of this volume), the emergence of a secondary pattern [Noun → [*për* + Noun]_{ADV}] (besides the pattern [Prep → *për* + Prep]) in Albanian could be seen as a way to clarify potentially

ambiguous cases linked to the multifunctionality of the items described in the previous section.¹² This pattern – which could be ancient, given the existence of cases such as *brenda* in which there has been some phonetic reduction of the compound – seems to have gained significantly in productivity in Modern Albanian, as shown in Section 2.2.2. The superposition of this pattern with the prepositional use of (some) adverbs and nouns can lead to complex systems, as in the case of *ballë* / *përballë* (51).

- (51) *ballë* ‘front’ (noun) / *ballë* + OBL ‘in front of’ (prepositional)
 → *përballë* ‘on the other side’ (adverb) / *përballë* + OBL ‘in front of’ (prepositional)

Another interesting case is the noun *krah* ‘arm’, which also forms a compound following this pattern, apparently only used as a preposition: *përkrah* ‘with’ (*përkrah popullit* ‘with people.OBL’, *përkrah shokut* ‘with/near comrade.OBL.PL’, see Buchholz and Fiedler 1987).

5.3 Complex prepositions: Loss and renewal

The limited diachronic depth of our corpus makes it impossible to provide an accurate appraisal of the emergence of CPs in Albanian. However, there are a few indirect clues in the literature and the data, which seem to indicate that this emergence is a rather recent and limited phenomenon. The first clue is precisely that CPs, at least as far as we could make out in our corpus study, constitute a modest paradigm in comparison to those found in other IE languages, esp. Romance, Germanic and Slavic (see the dedicated chapters in this volume). Compared with French, for instance, Modern Albanian seems to share various features with Old French as far as the system of nominal relations is concerned: remnants of the IE case system (more limited in Old French) and an incipient paradigm of CPs.

The fact that some CPs listed in Boissin (1975) or Buchholz and Fiedler (1987) are absent or nearly absent from our corpora could be interpreted as indicating a

¹² This happens to be a well-known IE feature, perfectly illustrated by the following example (i), in which three different word orders are possible, as indicated by Saussure (1916: VIII, 179: *óreos baínō káta* ≡ *katà óreos baínō* ≡ *kata-baínō óreos*).

- (i) *óreos* *baínō* *káta*
 mountain.G.SG walk.1SG down
 ‘I come down from the mountain.’

high degree of renewal in a short period of time – in other words, it could mean that this is not yet a lexicalized set of sequences.

Finally, the importance of language contact suggested by the differences between the two corpora we used (Section 3.3.) also seems to corroborate the impression that the paradigm of CPs is still in its first stages.

6 Conclusion

In this paper, we provide an overview of nominal relation marking in Albanian, including its various subsystems, i.e. case marking, simple prepositions, compound prepositions and complex prepositions. We show that nominal relations are mostly signaled with the help of case marking and simple or compound prepositions. However, case marking seems to have lost its semantically contrastive function (still found e.g. in German and most Slavic languages, and apparently still in use in Old Albanian), while there is evidence of the emergence of a (mostly incipient) paradigm of complex prepositions. Our corpus study shows that the number and frequency of these sequences (mostly Preposition + Noun + genitive marking *të*) is still limited, but suggests that there are sequences, among those already signaled in the literature, which display evidence of lexicalization and grammaticalization. The differences between the two corpora we used, OPUS and ANC, also suggests that language contact may have played a role in the emergence of such sequences. Finally, we show that there are still a number of clearly IE traits in Albanian, even if the long history of language contact has made them harder to distinguish.

Abbreviations

ABL	ablative	F	feminine
ACC	accusative	GEN	genitive
ADV	adverb	GER	gerund
AOR	aorist	Gr.	Greek
AUX	auxiliary	IE	Indo-European
CL	clitic	IMP	imperative
COMP	complementizer	IND	indicative
DAT	dative	Lat.	Latin
DEF	definite	M	masculine
DEM	demonstrative	NEG	negative
DET	determiner	NOM	nominative

OBL	oblique	PST	past
PL	plural	PTCP	participle
PRET	preterite	QUANT	quantifier
PRO	pronoun	REF	reflexive
PRS	present	SBJ	subjunctive
PRT	particle	SG	singular

References

- Adler, Silvia. 2001. Les locutions prépositives : questions de méthodologie et de définition. *Travaux de linguistique* 42–43(1). 157–170.
- Agalliu, Fatmir, Engjëll Agoni, Shaban Demiraj, Ali Dhimo, Enver Hysa, Emil Lafa & Ethem Likaj. 2002. *Gramatika e gjuhës shqipe, I: Morfologjia* [Grammar of the Albanian language, I: Morphology]. Tirana: Akademia e Shkencave.
- Boissin, Henri. 1975. *Grammaire de l'albanais moderne*. Paris: Chez l'auteur.
- Bonnet, Guillaume. 1999. *Les mots latins de l'albanais*. Paris: L'Harmattan.
- Bonnet, Guillaume. 2010. L'albanais. In Alain Peyraube, Emilio Bonvini & Joëlle Busutil (eds.), *Dictionnaire des langues du monde*, 511–520. Paris: PUF.
- Bopp, Franz. 1854. Über das Albanische in seinen verwandschaftlichen Beziehungen, *Akademie der Wissenschaften Berlin, Abhandlungen Phil.-Hist. Klasse*. 459–549.
- Buchholz, Oda & Wilfried Fiedler. 1987. *Albanische Grammatik*. Leipzig: VEB Verlag Enzyklopädie.
- Çanta, Agnesa. 2017. The Category of Case in English and Albanian Nominal System: A Contrastive Analysis. *International Journal of English Linguistics* 7(1). 226–239.
- Demiraj, Shaban. 1997. *Albanische Etymologien*. Amsterdam / Atlanta: Rodopi.
- Eberhard, David M., Gary F. Simons & Charles D. Fennig (eds.). 2019. *Ethnologue: Languages of the World*. Twenty-second edition. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>.
- Hamp, Eric P. 1972. Albanian. In Thomas A. Sebeok (ed.), *Current Trends in Linguistics, vol. 9: Linguistics in Western Europe*, 1626–1692. The Hague & Paris: Mouton.
- Hill, Virginia & Alexandru Mardale. Forthcoming. *The diachrony of Differential Object Marking in Romanian*. *Oxford Studies in Diachronic & Historical Linguistics*. Oxford: OUP.
- Hoffmann, Sebastian. 2005. *Grammaticalization and English Complex Prepositions. A corpus-based analysis*. Routledge, London & New York.
- Hopper, Paul J. & Elizabeth C. Traugott. 2003 [1993]. *Grammaticalization*. Cambridge: CUP.
- Horst, Joop van der. 2013. *Taal op drift: lange-termijnontwikkelingen in taal en samenleving*. Amsterdam: Meulenhoff.
- Hüning, Matthias. 2014. Over complexe preposities en convergentie [About complex prepositions and convergence]. In Freek Van de Velde, Hans Smessaert, Frank Van Eynde & Sara Verbrugge (eds.), *Patroon en argument. Een dubbelfeestbundel bij het emeritaat van William Van Belle en Joop van der Horst* [Pattern and argument. A double festschrift on the occasion of the retirements of William Van Belle and Joop van der Horst], 433–445. Leuven: Universitaire Pers Leuven.
- Julia, Marie-Ange. 2010. *Ante et circa: de la préposition à l'adverbe. De lingua Latina* 3.

- Kallulli, Dalina & Liliane Tasmowski. 2008. *Clitic Doubling in the Balkan Languages*. Amsterdam & Philadelphia: John Benjamins.
- Krijezi, Merima H. 2012. *Značenje predloga u albanskom i srpskom jeziku*. [Meaning of prepositions in Albanian and Serbian]. PhD Dissertation. Belgrade.
- Lehmann, Christian. 1985. Grammaticalization: Synchronic Variation and Diachronic Change. *Lingua E Stile* 20. 303–318.
- Likaj, Ethem & Maklena Çabej. 2013. *Morfologji e shqipes standarde* [Morphology of standard Albanian]. Tiranë: Julvin.
- Lindstedt, Jouko. 2010. Linguistic Balkanization: Contact-Induced Change by Mutual Reinforcement, *Studies in Slavic and General Linguistics* 28. 231–246.
- Lloshi, Xhevat. 1999. Albanian. In Uwe Hinrichs & Uwe Büttner (eds.), *Handbuch der Südosteuropa-Linguistik*, 277–299. Wiesbaden: Harrassowitz Verlag. Wiesbaden Slavistische Studienbücher, Band 10. [<https://web.archive.org/web/20110709114947/http://mirror.undp.org/albania/download/pdf/albanian.pdf>]
- Meyer, Gustav. 1891. *Etymologisches Wörterbuch der albanesischen Sprache*. Strasbourg: Trübner.
- Newmark, Leonard, Philip Hubbard & Peter Prifti. 1982. *Standard Albanian. A Reference Grammar for Students*. Stanford: Stanford University Press.
- Pedersen, Holger. 1900. Die gutturale im Albanesischen. *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete der Indogermanischen Sprachen* 36 (3). 277–340.
- Petit, Daniel. 2015. Les prépositions nominatives de l'albanais. *Bulletin de la Société de linguistique de Paris* CX(1). 359–443.
- Samara, Miço. 1999. Parafjalët në shqipen e sotme (Vështrim leksiko-semantik) [Parallels in Modern Albanian – a lexico-semantic analysis]. Tirana: Panteon.
- Sandfeld, Kristian. 1930. *Linguistique balkanique, problèmes et résultats*. Paris: Champion.
- Sinha, Chris & Tania Kuteva. 1995. Distributed Spatial Semantics. *Nordic Journal of Linguistics* 18. 167–199.
- Svane, Gunnar. 1992. *Slavishe Lehnwörter im Albanischen*. Aarhus: Aarhus University Press.
- Svorou, Soteria. 1994. *The Grammar of Space*. Amsterdam & Philadelphia: John Benjamins.
- Tiedemann, Jörg. 2012. Parallel Data, Tools and Interfaces in OPUS. *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'2012)*, 2214–2218.



Non Indo-European languages

Anni Jürine

7 What makes a complex postposition: The cyclic development of postpositions in Estonian

Abstract: The paper gives an overview of complex adpositions (CAs) in Estonian with a special focus on complex postpositions (CPOs). The development of CPOs in Estonian is a process still in its early stages and has (thus) so far received very little attention. However, in available previous research, CPOs have been claimed to have the following characteristics: unit interpretation, desemanticization, decategorialization, and extension to new contexts. These characteristics, which are also observed in many other accounts of CAs, may be used as criteria to distinguish Estonian CPOs from their source form (the simple postpositional phrase). As there are many parallels in the development of complex and simple postpositions in Estonian, the present paper observes these characteristics in CPOs against the backdrop of functional and formal properties of simple postpositions. The results indicate that the development of the simple and the complex structure share a number of characteristics, including the stage of lexicalization, which is argued to be vital for emerging of new function words in Estonian. It is suggested that the development of CPOs can be viewed as an instance of ‘recycling grammar’ – a recurrent phenomenon where new complex grammatical items develop as a result of lexical and grammatical components merging together.

Keywords: grammaticalization, lexicalization, recycling grammar, complex postpositions

1 Introduction

The purpose of the present paper is to give an overview of functional and formal properties of complex adpositions (CAs) in Estonian. More specifically, the paper aims to describe complex adpositions as a part of the system of nominal relators – cases and simple adpositions.

The development of CAs in Estonian is in its very early stages, and the phenomenon is not well charted in the literature. The traditional grammars have denied the existence of complex adpositions (Palmeos 1985) or have failed to

Anni Jürine, University of Tartu, Tartu, Estonia

<https://doi.org/10.1515/9783110686647-008>

mention CAs at all. The possible existence of CAs has been discussed by a handful of more recent studies (Habicht and Penjam 2007; Jürine 2011; Jürine and Habicht 2013; Jürine 2016), as a result of which, CAs have also been included in the most recent reference grammar of Estonian (Veismann and Erelt 2017). One of the most detailed analyses of CAs in Estonian is Jürine (2016), which among other things, lists a set of features to define complex postpositions (CPOs) and to distinguish them from their source forms. The aim of the present paper is a critical analysis of these features by contrasting the results with that of simple postpositions (SPOs).

SPOs prove to be a suitable material of comparison because there are many similarities between CPOs and SPOs. As SPOs are already an established category, which includes items that are in general further down the grammaticalization cline than CPOs, comparing the simple and the complex structures may give us valuable insights into the realm of CPOs, which has been less investigated. The critical analysis of the defining features serves two purposes – validation of the defining features of (complex)postposition-hood and observing the extent of their attestation. To validate the features, I observe the interplay of these features in CPOs (the complex structure) and SPOs (the simple structure). To better interpret the results of CPOs I will then compare the attestation of the features in both structures.

The paper is structured as follows. The following section gives an overview of nominal relation marking in Estonian, including typological background of Estonian, various ways of nominal relation marking, typological characterization of Estonian complex adpositions and non-adpositional use of CPOs. The next section gives an overview of the methodology, introducing the theoretical and methodological premises of this study. Section 4 presents the results of CPOs, contrasting them with that of SPOs. Section 5 discusses the results and presents the model of cyclic development of Estonian adpositions.

2 General description of Estonian (Complex) adpositions

2.1 Sociolinguistic and typological features of Estonian

Estonian belongs to the Finnic branch of the Uralic language family. Besides Finnish and Hungarian, Estonian is one of the few languages in the family that enjoys the status of being an official state language which is used in all public domains, including government and (higher) education. According to the 2011

census,¹ Estonian has close to 890,000 native speakers. Throughout its history, Estonian has been in close contact with various Indo-European languages, especially German, Swedish, and Russian.

Despite foreign influence over the years, Estonian carries many features typical of Finno-Ugric languages. As a typical example of the language group, Estonian is rich in morphology – grammatical markers are added to the word stems typically as suffixes. Estonian is primarily a head-final language. That is, in noun phrases, adjectival phrases as well as in adverbial phrases, modifier occurs typically before the head (e.g. *lapse raamat* ‘child’s book’, *uus raamat* ‘new book’). Typical for the language family, Estonian does not make use of articles, thus the NP is either a bare noun (*raamat*) or preceded by a demonstrative (see *raamat* ‘this book’), a possessive (*lapse raamat* ‘child’s book’, *tema raamat* ‘his/her book’) or a quantifier (*mitu raamatut* ‘many books’).

One of the characteristic typological features of Estonian relevant to this study is rich compounding. Compounding is a very productive way of word formation in Estonian. Compounds are more common than derivatives in Estonian and their formation is less restricted (Kasik 2015: 93). Many Estonian compounds are lexicalized and all of them form an orthographical word (Viitso 2003: 85; Kasik 2015: 97). Compounding is used in word formation of inflected (*häda* ‘emergency’ + *maanduma* ‘land’ ‘to emergency land’) as well as uninflected words, such as complex adverbs (*pool* ‘half’ + *püsti* ‘up’ ‘half-up’) (Kasik 2015: 176, 408). The formation of complex adverbs is not as regular as in inflected words. Kasik (2015: 409) suggests that the motivation for development of complex adverbs is a syntactically motivated change in parts of speech, whereby adverbials merge into adverbs. Based on the source forms, complex adverbs may be divided into three groups – NP based complex adverbs (*teine kord* ‘other + time’ > *teinekord* ‘sometimes’), VP based complex adverbs (*võib olla* ‘may be’ > *võib-olla* ‘maybe’), and PP based complex adverbs (*vahe peal* ‘on the gap’ > *vahepeal* ‘meanwhile’). The latter also serve as the source form for complex adpositions.

2.2 Nominal relation marking in Estonian

Nominal relations in Estonian can be marked by cases and adpositions, i.e. prepositions or postpositions. According to the traditional grammars, Estonian adpositions only come in simple forms (Palmeos 1985: 6). It is only in more recent studies (Habicht and Penjam 2007; Jürine and Habicht 2013; Jürine 2016)

1 http://www.stat.ee/64629?parent_id=39113

that the existence of complex adpositions has been postulated. In this section, I give an overview of the ‘traditional’ nominal relators in Estonian – cases and simple adpositions, which will be followed by description of complex adpositions (section 2.3.).

In Estonian, all nominals, i.e. nouns, pronouns, adjectives and numerals are inflected for number and case. Estonian has a 14-way case system that makes use of suffixes. The case system consists of three grammatical cases – nominative, genitive, and partitive; six locative cases, including three inner locative cases – illative, inessive, elative – and three outer locative cases – allative, adessive, and ablative. In addition, Estonian also has translative, terminative, essive, abessive and comitative cases. Among the grammatical cases (see Table 1), the nominative case is always unmarked. The partitive, on the other hand, has several possible endings. The genitive case has lost its historical ending (-n), and in contemporary Estonian, the genitive is unmarked for case, but may display changes in the stem. All other case forms are based on the genitive stem.

Table 1: The Estonian Case System.

Case	Case function	Example word <i>raamat</i> ‘book’
Nominative	Grammatical cases	<i>raamat</i>
Genitive		<i>raamatu</i>
Partitive		<i>raamatu-t</i>
Illative	Inner-locative cases	<i>raamatu-sse</i>
Inessive		<i>raamatu-s</i>
Elative		<i>raamatu-st</i>
Allative	Outer-locative cases	<i>raamatu-le</i>
Adessive		<i>raamatu-l</i>
Ablative		<i>raamatu-lt</i>
Translative	Other cases	<i>raamatu-ks</i>
Terminative		<i>raamatu-ni</i>
Essive		<i>raamatu-na</i>
Abessive		<i>raamatu-ta</i>
Comitative		<i>raamatu-ga</i>

Another way of marking nominal relations is using adpositions. In Estonian, adpositions are mostly postpositions (see example (1)). However, prepositions (2) and ambipositions (3) exist as well.

- (1) *maja* ***kõrval***
 house.GEN next to
 ‘Next to the house’
- (2) ***peale*** *kool-i*
 after school-PRT
 ‘After school’
- (3) ***mööda*** *teed* AND *teed* ***mööda***
 along way way along
 ‘Along the ‘Along the way’
 way’

Although Estonian adpositions are considered to belong to the class of uninflected words, they often come in three forms (Veismann and Erelt 2017: 448–449), standing for the lative (4), locative (5), and separative (6). This is especially common among postpositions (Palmeos 1985: 8).

- (4) *maja* ***kõrvale***
 house.GEN next to.LAT
 ‘To next to the house’
- (5) *maja* ***kõrval***
 house.GEN next to.LOC
 ‘Next to the house’
- (6) *maja* ***kõrvalt***
 house.GEN next to.SEP
 ‘From next to the house’

The most common sources for Estonian adpositions are nouns (Veismann and Erelt 2017: 449). In Estonian, adpositions develop in a grammaticalization and lexicalization process (Habicht 2001; Jürine 2016) where nouns are fossilized in locative cases. For instance, in examples (4)–(6) above, the adposition *kõrvale/kõrval/kõrvalt* ‘next to’ has evolved from the source noun *kõrv* ‘ear’ in outer locative cases. However, a developmental path where the noun fossilizes in inner locative cases (7–9) and less often in other cases (10) is also possible. In rare cases, adpositions have (infinitive forms of) verbs as their source forms (see example (11)).

- (7) *juur-de*
root-ILL
'near.LAT'
- (8) *juure-s*
root-INE
'near.LOC'
- (9) *juure-st*
root-ELA
'near.SEP'
- (10) *kese-t*
centre-PRT
'in the middle of'
- (11) *hooli-ma-ta*
care-SUP-ABE
'despite'

Estonian adpositional phrases consist of a nominal complement and a head adposition. In the case of postpositional phrases, the head postposition can take nominals in a number of cases, the choice of which depends on the particular postposition. Most commonly, postpositions take a modifier in the genitive case (12), but partitive, nominative, elative and comitative are also possible (Veismann and Erelt 2017: 452–454). Prepositions govern most frequently the partitive (13), but the genitive, elative, allative, comitative, abessive, and terminative are also possible (Veismann and Erelt 2017: 454–456). For postpositional phrases, the postposition and the head of the complement noun phrase need to be immediately adjacent (as in (12)). In the case of prepositions, the requirement of immediate adjacency can be violated, if the complement noun is modified (as in (13)).

- (12) *puu juures* BUT NOT: **puu sügava-s juures*
tree.GEN near tree.GEN deep-INE near
'Near the tree'

- (13) *keset* *paks-u* *mets-a*
 in the middle of thick-PRT forest-PRT
 ‘In the middle of a thick forest’

As mentioned above, Estonian has more postpositions than prepositions. However, as far as I know, no estimations have been made about the number of postpositions in Estonian, nor about the proportion of prepositions and postpositions in Estonian. Determining the number of Estonian adpositions is further complicated by the fact that adpositions are considered to be a relatively open class in Estonian (Jürine 2016: 24), meaning that it is continuously expanding through the process of grammaticalization.

Furthermore, as can be observed in examples (4–6), Estonian adpositions tend to retain the form of their sources, which are typically still in use as lexical items, making the source of the adpositions formally transparent. This complicates making formal distinctions between the source form and the adposition because the two structures are distinguishable only based on semantics and (sometimes) sentential context (see also Erelt et al. 2000: 145; Grünthal 2003: 56; Ojulkangas 2001: 47). This will be further elaborated below in section 2.3.

Even though formal reduction is not typical of Estonian adpositions, it can be found in some cases. For example, in (14) the postposition *sees* ‘inside’ has developed from a noun, which has vanished from most of the Finnic languages. The development of this postposition cannot be observed in the written records, and it has been assumed that formally reduced and intransparent adpositions (such as *sees*) developed before the early 16th century (Habicht 2000: 23), i.e. before the beginning of the Estonian Literary tradition.

- (14) *maja* *sees*
 house.GEN inside
 ‘Inside the house’

There is a considerable functional overlap between case suffixes and adpositions in Estonian (Veismann and Erelt 2017: 446) and to some extent, adpositions and cases can be viewed as forming a synchronic continuum of relational markers (cf. Lehmann 1985: 304) that range from more grammatical (case suffixes) to less grammatical (postpositions) items (Jürine 2016: 20). However, it should be noted that diachronically, case suffixes do not typically originate from postpositions (except for the comitative, see Habicht 2000: 43–44; Erelt 2009: 19). It has been pointed out by Tauli (1966: 12) that Finno-Ugric languages rather have the tendency of replacing case suffixes with postpositions.

2.3 Complex adpositions in Estonian

Even though the traditional grammars make no reference to complex adpositions as a (sub)category, the possibility of using complex adverbs ‘in the function of complex prepositions’ is mentioned already by Palmeos (1985: 6). Moreover, the practical need for such a (sub)category has been voiced by lexicographers (Karelson 2005) and is further echoed by the fact that in descriptive dictionaries, some morphologically complex items are tagged for part of speech as postpositions². However, a theoretical and empirical account of the development of CPOs in Estonian has been put forward only in more recent studies (Habicht and Penjam 2007; Jürine and Habicht 2012; Jürine 2016). Based on these findings, CPOs have made their way into the contemporary grammar of Estonian (Veismann and Erelt 2017: 450–451).

CPOs are formed by combining a noun in the genitive case and a simple postposition:

<i>selja</i>	<i>taga</i>
NOUN _{GEN}	SPO
‘back’	‘behind’

Thus, a CPO in Estonian consists of a noun, typically in the genitive case, and a simple postposition, which are analyzed as a holistic unit. CPOs act as the head of a complex postpositional phrase, where the CPO is preceded by a noun, a pronoun or a proper noun (see examples 15–17).

(15) *rahva selja taga*
 people.GEN back.GEN behind.LOC
 ‘Behind the people’s back’

(16) *meie selja taga*
 our back.GEN behind.LOC
 ‘Behind our back’

(17) *Obama selja taga*
 Obama.GEN back.GEN behind.LOC
 ‘Behind Obama’s back’

² <https://www.eki.ee/dict/ekss/index.cgi?Q=k%C3%A4ek%C3%B5rval&F=M>

As the development of CPOs is a process still in its initial stages, all Estonian CPOs are morphologically transparent. This means that all CPOs have two interpretations – a compositional interpretation (18) and a holistic interpretation (19). The two are distinguished based on sentential context. For instance in (18) *selja taga* is interpreted literally and is hence considered as a compositional phrase. In (19) *selja taga* carries the meaning ‘in secrecy’, which elicits a holistic interpretation of the phrase.

(18) *Poiss peitis end tüdruku selja taga*
 Boy hide himself girl.GEN back.GEN behind
 ‘The boy hid behind the girl’s back’

(19) *Otsus teh-akse rahva seljataga*
 decision make-IMPS people.GEN back+behind
 ‘The decision is made behind people’s back / in secrecy’

Similar to SPOs, CPOs also often come in three-form sets, expressing the lative, locative, or separative (see examples 20–22):

(20) *rahva selja taha*
 people.GEN back.GEN behind.LAT
 ‘To behind the people’s back/in secrecy’

(21) *rahva selja taga*
 people.GEN back.GEN behind.LOC
 ‘Behind the people’s back/in secrecy’

(22) *rahva selja tagant*
 people.GEN back.GEN behind.SEP
 ‘From behind the people’s back/in secrecy’

Estonian complex prepositions (CPREs) are formed by combining two uninflected words:

<i>all</i>	<i>pool</i>
uninflected word	uninflected word
‘down’	‘side’

Similar to CPOs, the two components are analyzed holistically. As such, they serve as the head of a complex prepositional phrase (see 23), where they are followed by a nominal (see examples 23–25).

(23) **allpool** *merepind-a*
 down+side sea level-PRT
 ‘Below the sea level’

(24) **väljaspool** *se-da*
 out+side this-PRT
 ‘Outside of this’

(25) **siinpool** *Saaremaa-d*
 here+side Saaremaa-PRT
 ‘On this side of Saaremaa’

Similar to (complex) postpositions, CPREs may occur in three-form sets (see 26–28). In this case, the lative, locative, and separative forms are used for both elements/components.

(26) **siia poole** *us-t*
 here.LAT side.LAT door-PRT
 ‘To this side of the door’

(27) **siinpool** *us-t*
 here.LOC side.LOC door-PRT
 ‘At this side of the door’

(28) **siitpoolt** *us-t*
 here.SEP side.SEP door-PRT
 ‘At this side of the door’

As can be observed in examples (26–28), the second component of a CPRE is always *poole*, *pool*, or *poolt* ‘to/in/from direction of’. Thus, *poole* seems to be the only uninflected word that together with a modifier can occur as a CPRE (Palmeos 1985: 5; Veismann and Erelt 2017: 450).

2.4 Complex adpositions as non-canonical members of the category of adpositions

Estonian CAs fulfill the typical functions of adpositions (as described in the introductory chapter of this volume), i.e. they

- a) elaborate³ a constituent,
- b) characterise a relation between this constituent and a higher constituent, or the sentence as a whole,
- c) contribute to determining the semantic nature of this relationship.

For example, it can be observed that in (29), the CPRE *siinpool* elaborates the constituent (*ust* 'door') and establishes that there is a locative relationship between the constituent and the higher constituent (*võti* 'key'). In examples (30)-(31), the CPOs *selja taga* and *käe all* express rather abstract notions of COVERTNESS/SECURITY and CONTROL/GUIDANCE.

(29) *Võti on siinpool us-t*
 key is here side.LOC door-PRT
 'The key is on this side of the door'

(30) *Otsus teh-ti teis-te firma-de selja taga*
 decision make-PST.IMPS other-GEN.PL company-GEN.PL back+behind
 'The decision was made behind other companies' back'

(31) *Jalgpallur mängi-b treener X-i käe all*
 footballer play-3SG coach X-GEN hand+under
 'The footballer plays under (the instruction of) coach X now'

Fagard et al. (this volume, introductory chapter) have suggested that in canonical CAs, there is a straightforward division of labour amongst the components that make up CAs. It must be noted that as Estonian complex postpositions have two

³ Fagard et al. (this volume) use the term *introduce*. However, *introduce* is not the most accurate term for describing *postpositions* because of their location relative to the constituent. Thus, when describing Estonian adpositions, it would be more appropriate to use Langacker's term *elaboration* defined as "the relationship borne to a schema by another structure that characterizes the same entity with greater specificity (i.e. finer precision and detail)" (Langacker 1991: 548). For example, in the prepositional phrase *near the door*, *near* elaborates the word *door* because it gives a more specific conception of the door by instantiating it as being part of a (spatial) relationship (Langacker 2013: 200).

(as opposed to three) components, there is no straightforward one-to-one mapping between the various functions and the individual components. In the following, I will explain why this is not the case, going through the functions one by one.

a) *elaborating the constituent*

Estonian CAs do not have a separate component for elaborating (or introducing) the constituent. Fagard et al. suggest that in canonical adpositions, this function is carried by P_2 (*of* in Figure 1). However, in Estonian, the equivalent function of P_2 would be the case ending marked on the constituent (*rahva* in Fig 1). However, as was stated earlier, the genitive ending has been lost. Thus, in this case, the elaboration would be marked on the complement noun, i.e. outside of the complex postposition itself. So even though the genitive stem connects the adposition to the complement noun, I argue that the elaboration is also attributable to the adposition itself. Thus, I propose that when observing the syntactic functions of Estonian CAs, it makes the most sense to observe the whole adposition in its context, which, in this case, is the adpositional phrase.

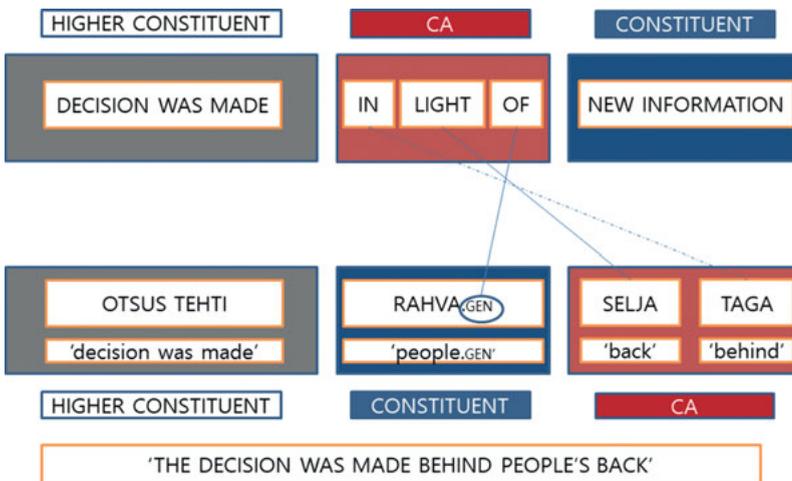


Figure 1: The components and division of labour of Estonian complex postpositions in comparison with canonical complex adpositions.

b) *characterising a relation between this constituent and a higher constituent, or the sentence as a whole*

In canonical CAs the relation between the constituent and higher constituent is coded in P_1 (*in*, see Figure 1). In Estonian CPOs, the same function is to some

extent coded by the simple position *taga* ‘behind’ (see Figure 1). However, in my previous work (Jürine 2016: 85) I have argued that the complex units develop through lexicalization, and are as such analysed holistically. According to this view, the relation between the constituent *rahvas* ‘people’ and the higher constituent *otsus tehti* ‘the decision was made’ would be coded in *selja taga* as a whole, in which case no distribution of meaning within the complex unit would be possible. It could be argued that because Estonian CPOs are still transparent, they are also compositional to a certain extent, meaning that the components of the CPO can carry different functions. However, I maintain that in the contexts that allow unit interpretation the relation between the constituent and higher constituent is characterised by the CPO *selja taga* holistically, because the simple postposition *taga* alone would characterise a different relation.

c) *contributing to determining the semantic nature of this relationship*

Similarly to my previous point, I argue that there is no single element responsible for the semantic nature of the adposition. Although it might be argued that the nominal nucleus *selja* ‘back’ contributes to determination of the semantic nature of the relation expressed by the adposition, it cannot be said that the simple postposition *taga* ‘behind’ plays a lesser role in this. To draw a parallel with another CPO, the type of relation *käe all* (‘under guidance of’) established in example (31) above is semantically motivated by both of its components. That is the nominal nucleus *käsi* ‘hand’ as well as the simple postposition *all* ‘under’ can be held responsible for the development of the meaning of the complex unit ‘under guidance/control/instruction of’. This applies particularly to Estonian CPREs, which lack nominal nucleus altogether (as in (29)).

2.5 Non-adpositional use of complex adpositions

Estonian complex adpositions can also function as other parts of speech, most commonly as complex adverbs (see examples (32)-(35)). The frequency of use in non-adpositional function depends on the specific complex item/word. There are complex units that predominantly function as complex adverbs, and adpositional use is rare, such as *selja taga* ‘back+behind’ when expressing spatio-temporal meaning (Jürine 2016: 87). However, there are also complex function words that rarely function as adverbs, such as *käe all* (hand+under) (Jürine 2016: 86). Nevertheless, belonging to several parts of speech is not reserved for complex structures, but is also widely attested among simple adpositions.

- (32) *Ta tee-b asj-u selja taga*
 S/he do-3SG thing-PL.PRT back+behind
 ‘She does things behind the back’
- (33) *Ta tee-b asj-u minu selja taga*
 S/he do-3SG thing-PL.PRT my back+behind
 ‘She goes behind my back’
- (34) *Ma ole-n siinpool us-t*
 I be-1SG here+side door-PRT
 ‘I’m on this side of the door’
- (35) *Ma ole-n siinpool*
 I be-1SG here+side
 ‘I’m on this side/over here’

In Estonian, morphologically complex grammatical words, which consist of several (formerly) independent words, are also found in other parts of speech. Most commonly, such items function as complex conjunctions, which may further develop into complex pragmatic markers (Jürine and Habicht 2017). The diachronic relationship between complex conjunctions and complex postpositions varies from case to case. However, most complex conjunctions are a product of lexicalization of the components, whereby the formerly freely combined items come to express a holistic meaning. For instance, the complex conjunction *selle pärast (et)* ‘because (of)’ (see example (36)) has developed from a nominal in the genitive case (*selle* ‘this’) and a simple postposition *pärast* ‘because’. It can be observed that *selle pärast (et)* ‘because (of)’ has the same components as a CPOs. However, this is not always the case. Complex conjunctions can also develop from other sources. In addition, not all lexicalized cases of a nominal and a simple postpositions that function as CPOs also function as complex conjunctions and vice versa.

- (36) *Tul-i-me sisse, sellepärast et väljas hakka-s*
 come-PST-1PL inside this+because+that outside start-PST.3SG
sadama
 rain-SUP
 ‘We came inside because it started raining outside’

Even though the paths of development (and sometimes even the source forms) of such complex items are similar (see also Jürine and Habicht 2017: 37, 39), there

are no cases known to me where a single complex unit would serve as all three parts of speech.

3 Methodology

3.1 Theoretical framework and methodological premises

The aim of the present paper is to give an overview of CAs in Estonian. More specifically, I am interested in the defining features of CAs (in Estonian) and to what extent these features are attested in the data. Because Estonian complex adpositions have many similar characteristics with simple adpositions, the results of CPOs are viewed in contrast with that of SPOs. The aim of this comparison is to provide more context for the analysis of the results of CPOs well as a critical analysis of the defining features for Estonian CPOs.

The theoretical framework of this study is provided by grammaticalization (e.g. Heine et al. 1991; Heine and Kuteva 2002; 2007; Hopper and Traugott 2003) and lexicalization (e.g. Brinton and Traugott 2005; Lehmann 2002). Both grammaticalization and lexicalization are types of language change that are taken to occur over time, and because of this, are often investigated diachronically. In this study, however, I take predominantly a synchronic approach to complex adpositions in Estonian.

The reason for choosing a synchronic approach is twofold. First, starting out from the early 1600s, Estonian has a relatively short literary history. The available corpora are relatively small⁴ and do not allow me to observe multi-word units in sufficient detail. Second, as the development of CPOs is considered to be a relatively new phenomenon in Estonian, it is not expected that a diachronic investigation would unveil strong uncontroversial evidence for this phenomenon (see also Jürine 2016: 54). Moreover, it can be said that it is especially the cases of incipient grammaticalization that easily lend themselves to synchronic investigation. Because diachronic change is often preceded by synchronic variation, the different usages of the grammatical(izing) item can be placed on a synchronic continuum, ranging from less grammatical to more grammatical uses. Because

⁴ The Corpus of Old Literary Estonian (1200–1800) (1.5 million words). Available at <http://www.murre.ut.ee/vakkur/Korpused/> (Accessed 20.07.2018) The Corpus of 19th century texts, 1800–1900 (0.5 million words). Available at <http://www.murre.ut.ee/vakkur/Korpused/Kwic2/paring19.htm> (Accessed 20.07.2018) The Corpus of Estonian Literary Language, 1890–1990 (4.5 million words). Available at <http://cl.ut.ee/korpused/baaskorpus/index.php?lang=en> (Accessed 20.07.2018)

grammaticalization (as well as lexicalization) adheres to the general principles of grammaticalization, such as the unidirectionality principle, it can be assumed that the synchronic continuum reflects diachronic change (Heine 1999: 179; Brinton and Traugott 2005: 101).

In this study, I proceed from the assumption that language variation and change is best captured by studying language use, i.e. the study adopts a usage-based approach (Kemmer, Barlow 2000; Bybee 2010). Thus, the present study is based on corpus data, which allows me to observe CPOs as they occur in natural language and take into account their frequency of occurrence. Even though the role of frequency in the grammaticalization process is far from clear (see for example Hoffmann 2004; Mair 2004), it is generally accepted that high frequency is associated with (a high degree of) grammaticalization. In this particular case of grammaticalization, which is concerned with multi-word units, besides absolute frequency associational strength is of interest. Associational strength shows how closely two (or more words) are associated to each other statistically (Evert 2005: 75). In this case, the statistical association demonstrates to what extent the complex postposition is fixed. This sort of fixedness (sometimes referred to as ‘fossilization’ or ‘freezing’) is associated with grammaticalization as well as with lexicalization (Brinton and Traugott 2005: 105), and as such, is taken to provide additional evidence of complex adposition-hood.

In regard to detection of CPOs, this study relies on the methodology developed in Jürine (2016). According to Jürine (2016: 33–34), Estonian body-part related CPOs exhibit the following parameters of grammaticalization, listed by Heine and Kuteva (2002; 2007) – desemantization⁵, decategorialization, and extension to new contexts^{6, 7}. These features, which are also described in many

5 In the case of Estonian complex postpositions, there are two types of semantic changes at play. For one, the grammaticalizing items go through desemantization, as the noun loses its referential capacity and develops a more abstract reference. At the same time, the abstract reference is acquired only as part of a larger unit, i.e. the formerly freely compositional phrase develops unit interpretation. As the two types of changes are intertwined and indistinguishable, they will not be separated in the analysis. Linguistic items that have gone through these changes will be referred to as having a ‘unit meaning’.

6 Thus, Estonian complex postpositions do not display the fourth parameter, erosion. On the one hand, this is explainable with the fact that the development of complex postpositions is an instance of incipient grammaticalization, while erosion typically takes place at the later stages of grammaticalization. On the other hand, (morphological) transparency is also common among simple postpositions (see above). Thus, it may be the case that Estonian postpositions are prone to retain their source form in general.

7 However, there have been studies that have taken note of a certain type of formal variation displayed by Estonian complex adpositions. It has been noted by Habicht and Penjam (2007)

other accounts of complex adpositions (e.g. Hoffmann 2005, Lehmann 2002, Moirón and Bouma 2003, Adler 2008), are used as defining features to distinguish Estonian CPOs from their source form.

The three parameters can be observed in example (37). The fact that *seljat-aga* (back+behind) as a whole refers to an abstract concept of SECRECY suggests desemantization and unit interpretation. Non-agreement between *firma-de* ‘companies-PL.GEN’ and *selja* (back.SG.GEN) manifests decategorialization of the body part noun. And finally, extension is manifested as the non-animate complement noun *firma-de* ‘companies-PL.GEN’ demonstrates that *selja taga* (back+behind) is no longer reserved to be used with words referring to entities with backs.

- (37) *Otsus* *teh-ti* *teis-te* *firma-de* ***selja***
 Decision make-PST.IMPS other-PL.GEN company-PL.GEN back.SG.GEN
taga
 behind
 ‘The decision was made behind other companies’ back’

3.2 Data set

The data comes from the etTenTen corpus, which consists of 270 million words⁸. The etTenTen corpus has been compiled from Estonian web texts, collected from 686,000 web pages, which fall into various categories (periodicals (25%), forums (20%), blog (10%), informative (7%), religion (3%), government (2%), unclassified (32%)). As such, the corpus includes edited texts that fall into the public domain of language use, and which adhere to formal register, as well as texts that represent written language that is more informal and often unedited. As such, the corpus is taken to represent contemporary written Estonian in all its variability and richness. Therefore, the corpus data are considered suitable to investigate variation that is taken to reflect linguistic change.

that some combinations of nouns and simple postpositions tend to be written as a single word. Jürine (2011) has demonstrated with a forced choice task that the single word spelling is associated with semantic change, i.e. non-literal interpretation of the combination of noun and simple postposition elicits single word spelling, and combinations that are interpreted as having a literal meaning are rather spelled as two words. However, following Jürine (2016: 26) who makes a convincing case that spelling can only be taken as a secondary parameter when investigating language change, the orthographical variation will be ignored here.

8 The corpus is available at www.keeveeb.ee (Accessed 20.08.2018)

The data investigated more thoroughly in this paper are restricted thematically as well as formally. That is, in this study I conduct a more thorough investigation of a small group of body part related complex and simple postpositions and their source forms (see Table 2). Because both CPOs as well as SPOs are to an extent still transparent and morphologically complex, they will be referred to as the complex structure and the simple structure.

Table 2: Complex and simple postpositions and their source forms.

Body part	Source Form	Complex postposition*			Source form	Simple postposition		
	N _{BODY PART NOUN} + PLOCATIVE				N _{BODY PART} + NOUN			
BACK	back.GEN + behind	<i>selja taha</i>	LAT	10958	back-ILL	<i>selga</i>	LAT	500
		<i>selja taga</i>	LOC		back-INE	<i>seljas</i>	LOC	
		<i>selja tagant</i>	SEP		back-ELA	<i>seljast</i>	SEP	
HAND	hand.GEN + munder	<i>käe alla</i>	LAT	4401	hand-ILL	<i>kätte</i>	LAT	500
		<i>käe all</i>	LOC		hand-INE	<i>käes</i>	LOC	
		<i>käe alt</i>	SEP		hand-ELA	<i>käest</i>	SEP	
	hand.GEN + beside	<i>käe kõrvale</i>	LAT	780				
		<i>käe kõrval</i>	LOC					
		<i>käe kõrvalt</i>	SEP					
SIDE	side.GEN + under	<i>külje alla</i>	LAT	2957	side-ILL	<i>külge</i>	LAT	500
		<i>külje all</i>	LOC		side-INE	<i>küljes</i>	LOC	
		<i>külje alt</i>	SEP		side-ELA	<i>küljest</i>	SEP	
NECK	neck.GEN + on	<i>kaela peale</i>	LAT	216	neck-ILL	<i>kaela</i>	LAT	500
		<i>kaela peal</i>	LOC		neck-INE	<i>kaelas</i>	LOC	
		<i>kaela pealt</i>	SEP		neck-ELA	<i>kaelast</i>	SEP	
Total				19312				2000

* See translations in Table 4.

In my previous work (Jürine 2016: 50–51) I have suggested that this group is suitable for the following reasons:

- 1) Body part words are a common source of adpositions in the world's languages (cf. Svorou 1994; Heine 1997; Heine and Kuteva 2002; 2007);
- 2) Body part related postpositions are a uniform set of linguistic items that provide a homogenous group for analysis;
- 3) Some of the selected postpositions have been discussed in former studies (Karelsen 2005: 65–66) as elements that function as complex postpositions;

- 4) Body part related complex postpositions provide a suitable sample for synchronic analysis because both their source forms and target forms are present in contemporary language.

All the data has been hand-coded for three categories – Semantic Interpretation, Grammatical Number of the Complement, and Semantic Class of the Complement (see Table 3). These codes are associated with the parameters of grammaticalization listed above (see section 3.1). The categories and all their possible levels as well as the parameters with which they associate are listed in Table 3.

Table 3: Coding scheme for complex and simple postpositions.

Category	Levels	Parameter
Semantic interpretation	Unit meaning	Desemanticization
	Compositional	
	Hybrid	
Grammatical number of the complement	Singular	Decategorialization
	Plural	
Semantic class of the complement	Animate	Extension
	Inanimate	

The data sample of CPOs, which consists of 19,312 examples, originates from Jürine 2016. No codes have been changed for the purpose of the present study. The considerably smaller data sample of SPOs, which consists of 2,000 examples (500 examples each) is used as a mean of validation and has been coded specifically for the present study.

4 Results: Body-part related complex postpositions in Estonian

4.1 Frequency and associational strength of complex postpositions

In this section, I present the results on frequency and associational strength of CPOs. Frequency here refers to absolute frequency of the construction $N_{\text{BODY}} \text{PART NOUN} + P_{\text{LOCATION}}$, i.e. the results presented here include all instances of the

construction irrespective of its linguistic analysis or part of speech – complex postposition, complex adverb, or a freely combined simple adpositional phrase. Associational strength is measured using the log-likelihood score.

It can be observed in Table 4 that the studied body-part related adpositions display great variation as to their frequency in the corpus. The most frequent of the studied constructions is *selja taga* (back+behind), which occurs over 10,000 times in the corpus of 270 million, *käe all* (hand+under) and *külje all* (side+under) are in the middle in our data set, and the least frequent constructions in our sample are *käe all* (hand+under) and *kaela peal* (neck+on).

Table 4: Absolute frequencies of body part related constructions $N_{\text{BODY PART NOUN} + P_{\text{LOCATIVE}}}$ (corpus size 270,000,000).

Body part SG	<i>f</i>	Body part PL	<i>f</i>	Other body part related prases (SG)	<i>f</i>
<i>selja taga</i> (back+behind)	10,958	<i>selgade taga</i> (backs+behind)	4	<i>nina kõrval</i> (nose+beside)	19
<i>käe all</i> (hand+under)	4,401	<i>käte all</i> (hands+under)	124	<i>pea all</i> (head+under)	338
<i>külje all</i> (side+under)	2,957	<i>külgede all</i> (sides+under)	10	<i>kaela all</i> (neck+under)	34
<i>käe kõrval</i> (hand+beside)	780	<i>käte kõrval</i> (hands+beside)	1	<i>selja all</i> (back+under)	46
<i>kaela peal</i> (neck+on)	216	<i>kaelade peal</i> (necks+on)	0	<i>käe kohal</i> (hand+over)	7

However, the frequency of a linguistic item is always relative, and the results can only be interpreted meaningfully when contrasted with comparable data. One of the criteria of CAs is the inability to display morphological variation, such as pluralization (see also Sigurd 1993: 199). When we compare the absolute frequencies of the studied phrases (body part noun in the singular) to phrases with nouns in the plural, it can be observed that pluralized variants are almost non-existent (Table 4, columns 3–4). For example, the absolute frequency *selgade taga* (backs+behind) is 4 (vs 10,958 with the singular form). The frequency of other constructions (with the exception of *käte all* (hands+under)) does not exceed 10. When we compare the frequencies of the studied phrases to that of other constructions that consist of a body part noun and a simple postposition, but which do not form a complex adposition, we also see that the frequencies of freely combined phrases are considerably lower (Table 4, columns 5–6). The discrepancy between the frequencies of similar constructions suggests that the

studied phrases are not freely combined phrases but instead fixed units, with (at least) a potential to grammaticalize.

Even though observing absolute frequencies may be useful when studying grammaticalization as well as lexicalization, they may not tell the whole story when studying multi-word expressions. For example, it has been claimed that absolute frequencies of multi-word units may be uninformative because absolute frequency may reflect the high frequency of only one component (Evert 2005: 20–21). This is why the absolute frequencies are often complemented by an associational strength measure, such as the log-likelihood measure. Taking into account the frequency of each component as well as the corpus size, the log-likelihood test measures the collocational strength between the components of the complex postposition. The results of the association measure test can be observed in Table 5, which ranks the studied phrases according to the value of the log-likelihood score.

Table 5: Associational strength of body part related constructions $N_{\text{BODY PART NOUN}} + P_{\text{LOCATIVE}}$ (N = 270,000,000).

Body part SG	LOG	Body part PL	LOG	Other body part related prases (SG)	LOG
<i>selja taga</i> (back+behind)	139,607	<i>selgade taga</i> (backs+behind)	12	<i>nina kõrval</i> (nose+beside)	0,5
<i>käe all</i> (hand+under)	26,167	<i>käte all</i> (hands+under)	148	<i>pea all</i> (head+under)	274,7
<i>külje all</i> (side+under)	24,769	<i>külgede all</i> (sides+under)	30	<i>kaela all</i> (neck+under)	-18,9
<i>käe kõrval</i> (hand+beside)	3,778	<i>käte kõrval</i> (hands+beside)	8	<i>selja all</i> (back+under)	-37,6
<i>kaela peal</i> (neck+on)	864	<i>kaelade peal</i> (necks+on)	NA	<i>käe kohal</i> (hand+over)	-20,0

The analysis shows that the ranking is the same as based on absolute frequency – the most frequent item *selja taga* (back+behind) also has the strongest association between its components. Complex items located in the middle of the frequency table – *käe all* (side+under) and *külje all* (side+under) – are considered intermediate based on their associational strength, and the least frequent items – *käe kõrval* and *kaela peal* – have the weakest score. When we compare these scores to that of body part related phrases with pluralized nouns (Table 5, columns 3–4) or other, freely combined postpositional phrases (Table 5, columns 5–6), we see that the scores of the potential CPOs under investigation are considerably higher. This provides further evidence that the studied items form fixed units.

4.2 Defining features of complex postpositions in Estonian

In this section, I will observe three defining features of CPOs – unit meaning, decategorialization, and extension – as well as their interplay in the dataset.

4.2.1 Unit meaning

Figure 2 presents the proportion of examples that display semantic change. It was observed that most (13,408) of the 19,312 examples analyzed express unit meaning. However, the proportion of usages varies across constructions, and there seems to be no correlation between high frequency and proportion of examples with unit meaning. For instance, in the case of *selja taga* (back+behind), which is by far the most frequent of the studied constructions, compositional meanings (as in (38)) make up almost half (45%) of all examples. On the other hand, *käe kõrval* (hand+beside), which is one of the least frequent of the studied constructions (with just 780 attestations), is almost exclusively (98%) used as a holistic unit (39). *Käe all* (hand+under) and *külje all* (side+under) which are placed in the middle of the frequency ranking, are also almost exclusively (92% and 85% respectively) used as holistic units (see example (40)).

- (38) *Jät-si-n kassi oma selja taha, et ta*
 leave-PST-1SG cat.GEN own-GEN back behind that it
sül-le võt-ta aga koer o-li kiirem /.../
 lap-ILL take-INF but dog be-PST.3SG quicker
 [etTenTen: www.lemmik.ee]
 ‘I left the cat behind my back so that I could pick it up, but the dog was quicker’

- (39) *Lugeja saa-b autori käekõrval osa loomeprotsessi-st /.../*
 reader can-3SG author.GEN hand+beside part creative process
 [etTenTen: vanajahea.ee]
 ‘Together with the author, the reader can take part in the creative process’

- (40) *Tudengi-te käe all o-n sündi-nud kuus ülikool-i*
 Student-PLGEN hand under be-3SG born-PST.PTCP six university-PRT
tutvustava-t lühivideo-t. [EtTenTen: www.tlu.ee]
 introducing-PRT short video-PRT
 ‘The students have created six short videos introducing the university’

However, the smaller proportion of uses with unit meaning does not necessarily indicate that the item is less grammaticalized. It can be assumed that the distributions depicted in Figure 2 are also influenced by the semantics of the source form. For instance, it can be assumed that some source forms have a broader and some a more specific reference. That is some constructions profile situations that are more common than others – something being located behind one’s back is perhaps a more common scenario than something being located beside one’s hand (see also Jürine 2016: 64).

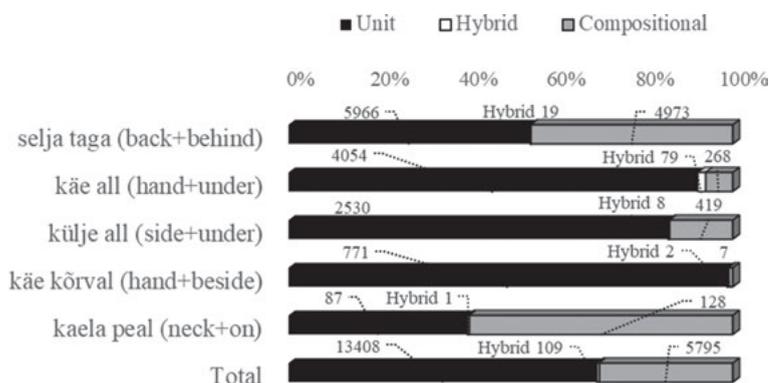


Figure 2: The proportion of examples that display unit meaning (n = 19,312).

The data also includes so-called hybrid forms, i.e. uses where the construction has two possible interpretations. Uses such as (41) can be, on the one hand, interpreted as CPOs because they have a unit meaning comparable with (40) above. On the other hand, they violate the immediate adjacency requirement (see section 2.2 above) set for all postpositions in Estonian, and therefore cannot be analyzed as such. Thus, such forms are analyzed as hybrids, i.e. cases that have characteristics of both the source and the target forms (see also De Smet 2012). Such hybrids are quite rare in the data, the largest proportion is found in the data of *käe all* (hand+under), where they make up just 1% of the data. Because they are too few in number to form a category of comparable size to the other two, they will be excluded from further statistical analysis.

- (41) /.../ 6-käigulis-t menüü-d mis valmi-s chef Rene Uusmees'i
 6 course-PRT menu-PRT that be made-3SG chef Rene Uusmees.GEN
valvsa pilgu ja kindla käe all.
 watchful.GEN glance and firm.GEN hand under

[etTenTen: www.vine.ee]

‘... 6 course menu that was made under the watchful eye and firm hand of chef Rene Uusmees’

4.2.2 Decategorialization and extension

In the present section, it is observed to what extent decategorialization and extension are attested in the data, and how they are associated with semantic interpretation, i.e. unit and compositional meaning of the $N_{\text{BODY PART}} + P_{\text{LOCATIVE}}$ constructions.

Decategorialization and extension can be observed in examples where the $N_{\text{BODY PART}} + P_{\text{LOCATIVE}}$ construction is preceded by a nominal, which belongs to the same phrase as the body part noun. As mentioned above, in this case, **decategorialization** is indicated by non-agreement in number of the body part noun and the preceding noun. Figure 3a gives the distribution of all *plural complements* that occurred in the data ($n = 1,672$) and their distribution among unit and compositional meaning.

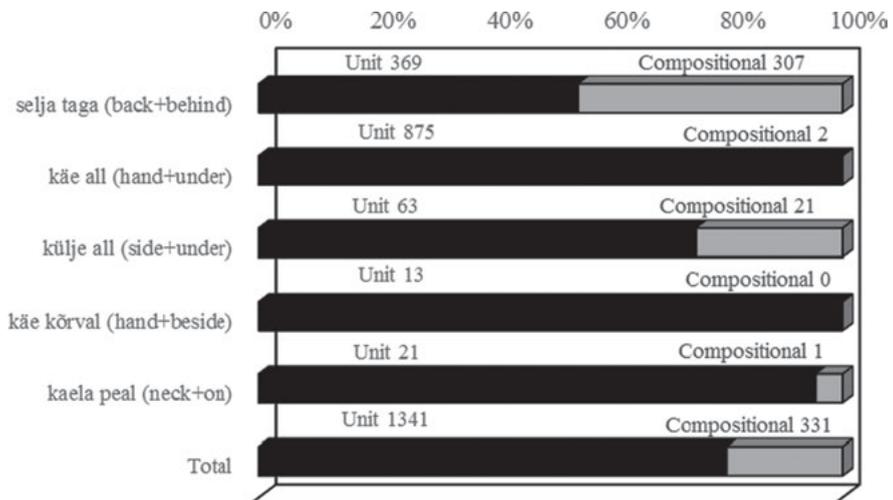


Figure 3a: Distribution of plural complements among usages with unit and compositional meanings ($n = 1,672$).

As a general tendency, non-agreement in number is associated with unit meaning. The data show that in total, 80% (1,341 out of 1,672) of the plural complements co-occurred with unit meaning. However, plural complements are not reserved for

such usages only. In example (42), it can be observed that non-agreement might occur also in instances where the construction has a compositional meaning. This can be explained with free variation or the possibility of interpreting *selja taga* (back+behind) in example (42) as a complex postposition that (as a whole) bears locative meaning ('in back of') (Jürine 2016: 115).

- (42) *Päkapiku-d pike-ma-d ja lühe-ma-d*
 elf-PL tall-COMP-PL and short-COMP-PL
nen-de selja taga /.../
 they-PL.GEN back-GEN behind
askelda-si-d vilkalt
 bustle-PST-3PL hastily
 [etTenTen: blog.maaleht.ee]
 'Elves – some taller, some shorter – were bustling about hastily behind their backs'

Thus, even though plural complements are not entirely reserved for instances that carry unit meanings, they have a strong tendency to co-occur with such instances. This gives strong evidence for regarding non-agreement as an indicator of decategorialization. However, it should be noted that decategorialization is not particularly common in my data set. Figure 3b, which presents the distribution of

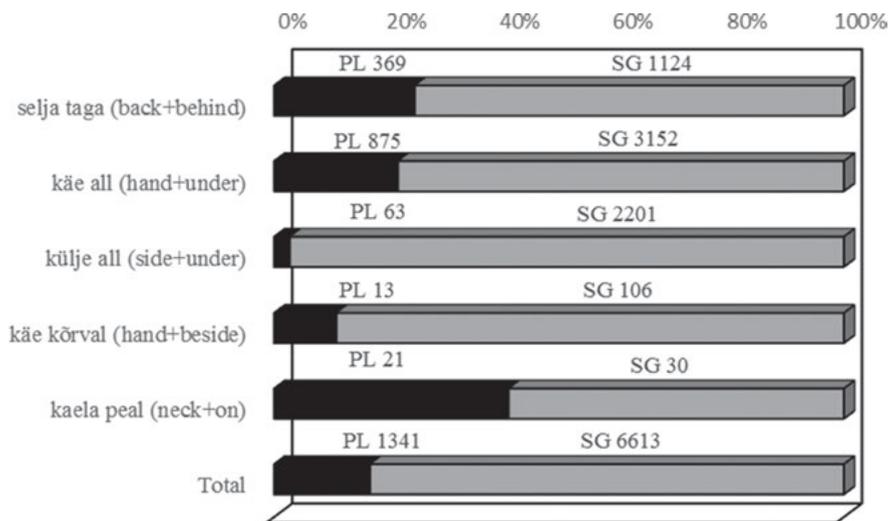


Figure 3b: Proportion of instances of decategorialization among CPOs (n = 7,954).

plural and singular complements among all CPOs ($n = 7,954$), shows that on the whole, decategorized uses make up a small proportion (17%; 1341) of all CPOs.

Extension in Estonian complex postpositions is observed in complement nouns, more specifically the semantic class of the complement noun. Figure 4a depicts all inanimate complements in the data ($n = 2,533$) and their distribution among usages with unit and compositional meanings. Similarly to plural complements discussed above, inanimate or non-individual complements tend to co-occur with those instances of $N_{\text{BODY PART}}+P_{\text{LOCATIVE}}$ constructions that are interpreted as semantically holistic units (see example (43)). In this case, there are only 20 instances of inanimate or non-individual complements that co-occurred with literal readings (see example (44)). As such, inanimate complements can be considered to manifest extension of the CPO to new contexts.

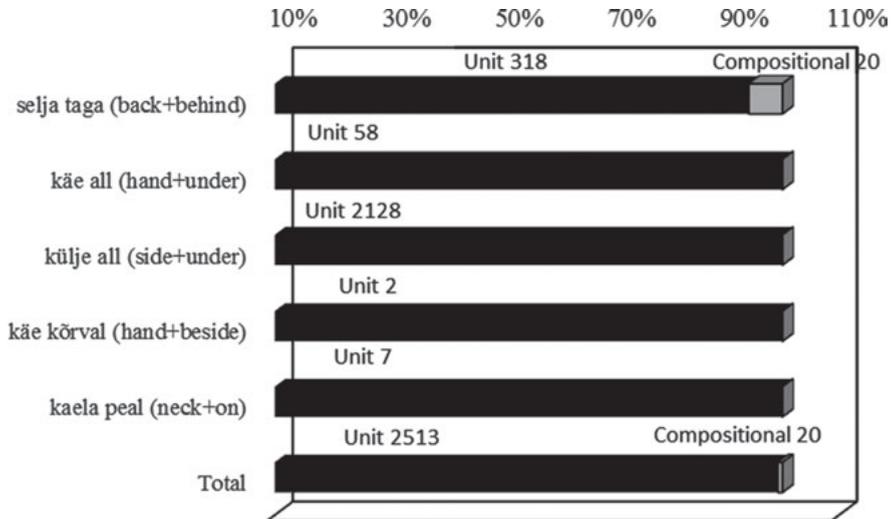


Figure 4a: Distribution of inanimate complements among unit and compositional meanings ($n = 2,533$).

- (43) /.../ *et uue valitsuse käe all tõusevad avaliku sektori palga-d* [etTenTen: tark.tooandjad.ee]
 that new.GEN government.GEN hand under raise-3PL public.GEN
 sector.GEN salary-PL
 ‘... that under the new government, the salaries in the public sector will rise ...’

- (44) *Unune-b olukord, et asu-me 2 km vaenlase seljataga,*
 forget-3SG situation that be located-1PL 2 kms enemy.GEN back behind
oma-de-st ära lõiga-tud. [etTenTen: www.virtsu.ee]
 own-PL-ELA off cut-PST.PTCP
 ‘One can forget the fact that we are located 2 kms behind (the back of) the enemy, being cut off from our own’

Even though inanimate complements are a clear indication of extension, they are also not very common. Figure 4b depicts all instances of extension in the data (n = 7,954) and the distribution of animate and inanimate complements. It can be observed in Figure 4b that extension is extremely frequent in the case of *külje all* (side+under), but is less common or almost nonexistent in other constructions. The total number of inanimate nominal complements that co-occur with instances of unit meaning is 2,513, which makes up about 31% of the data set.

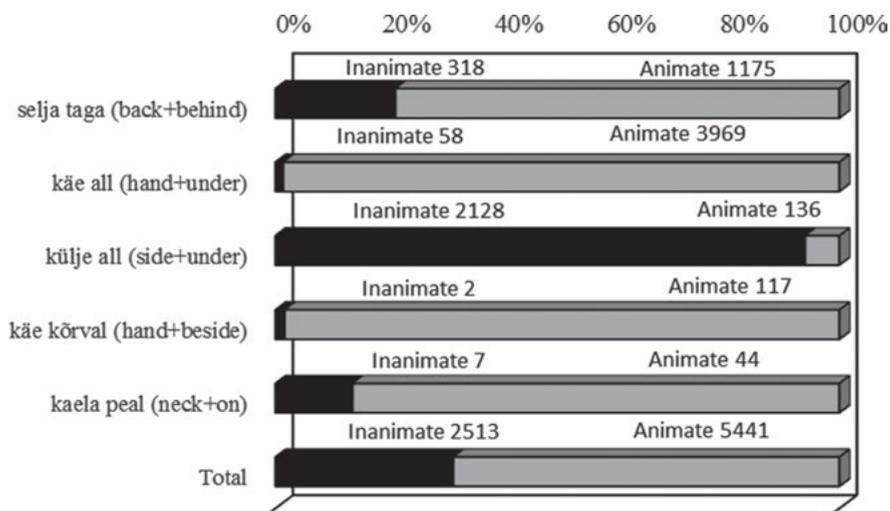


Figure 4b: Proportion of instances of extension among CPOs (n = 7,954).

In sum, the analysis of complex structures demonstrates that these criteria differentiate between unit meaning and compositional meaning rather well. Thus, plural complements and inanimate complements can be considered to manifest decategorialization and extension respectively, and as such, be indicators of grammaticalization. However, the data also demonstrates that the proportions

of decategorialized and extended forms are quite small. This would suggest that the development of CPOs is an instance of grammaticalization in its very early stages.

4.3 Comparison of the simple and the complex structure

In this section, I will compare the results of the complex structure (CPOs) presented above to that of the simple structure (SPOs). First, I will test the validity of plural and inanimate complements as indicators of decategorialization and extension by observing their distribution among unit and compositional interpretations in the data set of the simple structure. Then I will observe the proportions of examples displaying decategorialization and extension.

When we observe the distribution of plural and inanimate complements among unit and compositional meanings in the simple structure, we obtain a result similar to that of the complex structure presented above (see Figure 5a). Both plural and inanimate complements have a strong tendency to co-occur with unit meanings. While there is some variation in the individual forms, the proportion of plural complements that co-occur with compositional meanings remains under 3% and the proportion of inanimate complements under 1%. This is taken as further evidence that plural and inanimate complements suggest extension and decategorialization of the complex unit, and as such are indicators of grammaticalization and (complex) postposition-hood.

Figure 5b gives the proportions of examples with unit meaning, decategorialization, and extension in the simple structures and the complex structures. It can be observed that the results point in the same direction. Whereas unit meanings are quite frequent in both structures (74% in complex and 91% in simple structures), the proportions of decategorialized and extended usages of both simple and complex structures remain below 25%. Thus, based on this evidence, it seems that higher levels of decategorialization and extension cannot be expected of the complex structures.

Further comparison of the two structures shows that quite expectedly the simple structure displays slightly higher levels of unit meaning and decategorialization (74% vs 91% and 24% vs 16%). Those results are also statistically significant ($\chi^2 = 259.6257$, $p < .00001$ and $\chi^2 = 14.1063$, $p = .0001$ respectively)⁹. However, a more unexpected result is that extension is more common in complex structures

⁹ The chi-square tests were performed using the Social Science Statistic calculator, available at <https://www.socscistatistics.com/tests/chisquare2/Default2.aspx> (Accessed 28.07.2018)

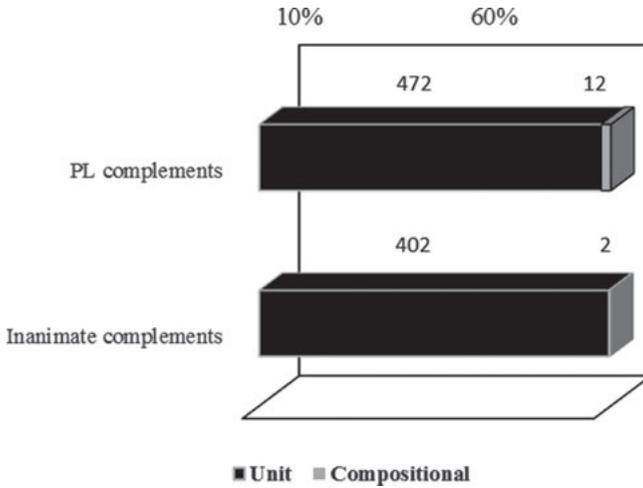


Figure 5a: The distribution of plural and inanimate complements among unit and compositional meanings in the *simple structures* ($n_{PL\ COMP.} = 484$, $n_{INANIMATE\ COMP.} = 404$).

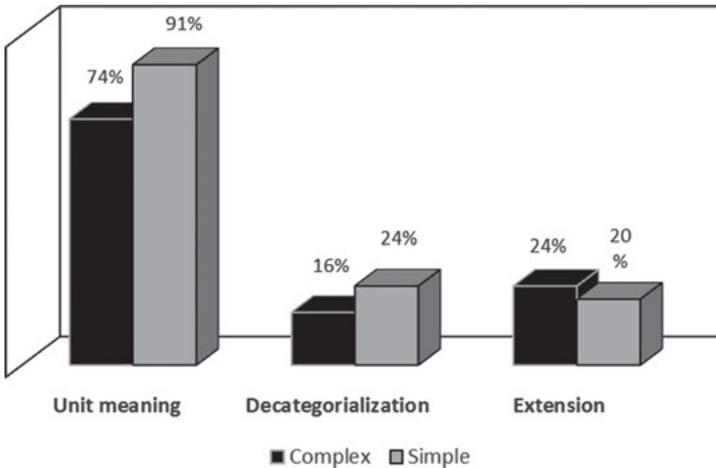


Figure 5b: The proportions of instances of unit meaning, decategorialization, and extension in *simple* and *complex structures* ($n_{SIMPLE} = 2,000$; $n_{COMPLEX} = 10,518$).

than in simple structures. This result is also statistically significant ($\chi^2 = 84.5507$, $p < .00001$). This indicates that based on this criterion, the observed complex structures could be considered further down the grammaticalization path than the observed simple structures.

5 Discussion: Lexicalization as the driving force for the cycle of postpositions

The present study has investigated Estonian CPOs with the main aim of discovering what makes a CPO. That is, the main purpose of this study has been to determine the defining features for CPOs in Estonian to learn more about the mechanisms at work in development of these grammatical items. To meet this purpose, I have revisited the parameters listed for CPOs by Jürine (2016) – unit interpretation, decategorialization, extension – and tested their validity on a small dataset of simple postpositions.

The results indicate that the defining feature for Estonian CPOs is unit interpretation. That is, if the construction $N_{\text{BODY PART}}+P_{\text{LOCATIVE}}$ has developed a unit meaning it can be analyzed as a CPO, given that it fulfills the criteria set for adpositions in general – it elaborates an immediately adjacent nominal (constituent) and establishes a relation between this nominal and the rest of the sentence (higher constituent).

The critical nature of the semantic change is highlighted by the fact that unit interpretation is by far the most common feature among complex postpositions as well as in simple postpositions, whereas formal changes, such as decategorialization and extension, are much more infrequent. As such, the development of CPOs in Estonian is a typical instance of reanalysis, a type of language change where semantic change precedes all formal changes (Langacker 1977: 58). It is only in the process of actualization (De Smet 2012) that the result of reanalysis becomes visible, as the change is then manifested by formal evidence. In the case of the phenomenon under study in this paper, the formal evidence for actualization is provided by a CPO's occurrence in a new sentential context, i.e. with plural or inanimate complements.

The present study has demonstrated that plural and inanimate complements can indeed be taken to manifest decategorialization and extension. This is evidenced by the fact that, in both complex and simple structures, plural and inanimate complements do not occur randomly in the data, but are associated with unit interpretation. Heine and Kuteva (2007: 34–35) have also pointed out that the development of new grammatical meanings takes place when linguistic items are extended to new contexts, making meaning change and extension two sides of the same coin. The results of simple and complex structures concur.

Further contrasting of the results of the complex and simple structures also elucidated some differences. Even though both structures demonstrated association between unit interpretation and plural and inanimate complements, the data

also indicate that this tendency is weaker among complex structures. However, as complex postpositions are taken to be less grammaticalized than simple postpositions, this outcome is rather expected. On the other hand, the results also demonstrated that decategorialization and extension are rather infrequent in comparison with unit interpretation in both structures, simple and complex. This suggests that formal changes may not be that frequent in the development of adpositions in general, and might not be necessarily a good indication of incipient grammaticalization or grammaticalization of complex items, as I have also suggested before (Jürine 2016: 196).¹⁰

However, I do maintain that it is semantic change which allows the CPOs to enter the grammaticalization path, i.e. to undergo decategorialization and/or extension. It is important to highlight that the main mechanism for semantic change in the development of Estonian (complex) postpositions is lexicalization. Lexicalization, in this case, is defined as a type of language change whereby a ‘new contentful’ form is created with formal and/or semantic properties that cannot be directly derived from its components (Brinton and Traugott 2005: 96). It must be noted that lexicalization is here taken to be a phenomenon that is not the opposite of the grammaticalization process, but rather the two are taken as processes that complement each other, and which can occur in a single instance of language change. To put it in Lehmann’s (2002) terms, lexicalization is here understood as ‘adoption into (mental) lexicon’, which means that the linguistic item is analyzed holistically, not analytically.

Lexicalization defined as such also allows for the so-called recycling of grammar (see Jürine and Habicht 2017: 39). Recycling grammar is understood as a process whereby formerly freely combined grammatical (and lexical) units are merged together in the process of lexicalization to form a new grammatical item. For instance, the development of complex postpositions can be viewed as a part of a cyclic path of development (see Habicht and Penjam 2007; Jürine and Habicht 2013), whereby new grammatical items are created in the recursive process of merging nominal relators (case endings and simple adpositions) with lexical items (see Figure 6).

The schema in Figure 6 depicts the cyclic model of development of the complex postposition *kaela peal* (neck+on) ‘as a burden to’, showing its path in three stages. The process starts out from a free combination of the body part term *pea* ‘head’ and a locative (adessive) case ending *-l* (as exemplified in (45)). In the first stage, the free combination of the body part noun and the case ending

¹⁰ Of course, incipient grammaticalization is a relative term and this topic deserves further investigation and, if possible, quantification.

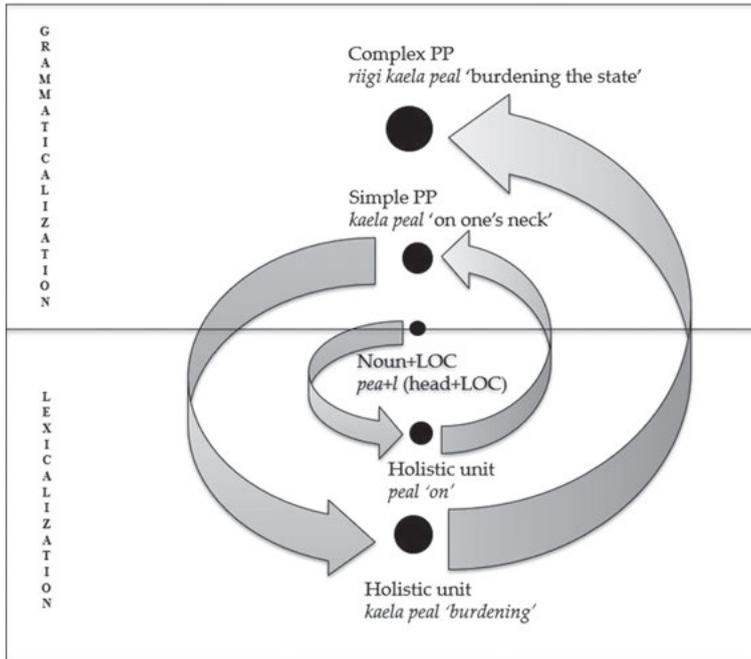


Figure 6: The cyclic development of function words in Estonian (modification of Jürine and Habicht 2013).

becomes a holistic unit *peal* 'on', which can be used either as a simple adverb or as a simple postposition (as in (46)). Once a postposition, *peal* 'on' extends to new contexts and may be freely combined with many nominal complements (such as *kael* 'neck', see example (47)). In the next stage, in specific contexts, the simple postposition and its complement noun are lexicalized and acquire a new holistic meaning. The lexicalized unit may also be used as a complex adverb or as a complex postposition (as in example (48)). This process is also aided by analogy – as the lexicalized item fits right into the slot previously available for the nominal relator. Immediately following the nominal in the genitive case, complex postpositions fit into the same slot as simple postpositions (*peal* in 46) and case endings (adessive ending *-l* in (45)).

- (45) *Sääsk maandu-s lapse pea-l.*
 Mosquito land-3SG kid.GEN head-ADE
 'A mosquito landed on the kid's head'

- (46) *Sääsk maandu-s kapi pea-l.*
 Mosquito land-3SG cupboard.GEN head-ADE
 ‘A mosquito landed on top of the cupboard’
- (47) *Sealjuures ei tohi masina lõikustera liiga*
 Therewith NEG can.CONNEG machine.GEN blade.PRT too
vastu nahka suruda, eriti kaela peal.
 against skin push especially neck on
tugevasti
 strongly
 [etTenTen: www.iluguru.ee]
 ‘In addition, one should not push the blade of the machine too strongly against one’s skin, especially on the neck’
- (48) *Kui te-id ema-isa on sellis-te-ks kasvata-nud,*
 if you-PRT mom-and-dad be.3SG this kind-PL-TRA raise-PST.PTCP
siis ol-ge aga nen-de kaela peal. [etTenTen: noortemaal.delfi.ee]
 then be-IMP but they-GEN neck on
 ‘If this is how your mom and dad have raised you then go ahead and live as a burden to them’

Lexicalization is involved in all stages of the process whereby a formerly freely combined unit becomes a holistic unit (lower part of Figure 6). As such, it is taken to be a vital part of the grammaticalization process (upper part of Figure 6) of complex units. This sort of recycling is not characteristic of Estonian postpositions only – a similar cyclic change has been described for example by Waters 2009 and McMichael 2006 for English. However, what makes Estonian postpositions a particular case is the fact that the source forms of the adpositions can be kept intact through many cycles.

Abbreviations

ABE	abessive	ELA	elative
ADE	adessive	GEN	genitive
COMP	comparative	ILL	illative
CONNEG	connegative	IMP	imperative
CPO	complex postposition	IMPS	impersonal
CPRE	complex preposition	INE	inessive

INF	infinitive	PTCP	participle
LAT	lative	SEP	separative
LOC	locative	SG	singular
NEG	negative	SG	singular
PL	plural	SPO	simple postposition
PRT	particle	SUP	supine
PST	past	TRA	translative

References

- Adler, Silvia. 2008. French compound prepositions, prepositional locutions and prepositional phrases in the scope of the absolute use. In Dennis Kurzon, Silvia Adler (eds.) *Adpositions: Pragmatic, semantic and syntactic perspectives*. Typological Studies in Language vol. 74, 17–36. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Brinton, Laurel, J. & Elizabeth Closs Traugott. 2005. *Lexicalization and language change*. Research Surveys in Linguistics. Cambridge: Cambridge University Press.
- Bybee, Joan L. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- De Smedt, Hendrik. 2012. The course of actualization. *Language* 88, 601–633.
- Erelt, Mati. 2009. Typological overview of Estonian syntax. *STUF*, Akademie Verlag, 62: 1/2, 6–28.
- Erelt, Mati, Tiiu Erelt, & Kristiina Ross. 2000. *Eesti keele käsiraamat [Handbook of Estonian]*. 2nd edn. Tallinn: Eesti keele sihtasutus.
- Evert, Stefan. 2005. *The statistics of word cooccurrences: Word pairs and Collocations*. Ph.D. Dissertation. Institut für maschinelle Sprachverarbeitung Universität Stuttgart.
- Grünthal, Riho. 2003. *Finnic adpositions and cases in change*. Suomalais-ugrilaisen seuran toimituksia 244. Helsinki: The Finno-Ugrian Society.
- Habicht, Külli. 2000. Grammaticalization of adpositions in Old Literary Estonian. In Mati Erelt (ed.) *Estonian: Typological Studies IV*. Tartu Ülikooli eesti keele õppetooli toimetised [Publications of Chair of Estonian language of University of Tartu], vol. 14, 19–58. Tartu: University of Tartu.
- Habicht, Külli. 2001. Lexicalization and grammaticalization – opposite phenomena? About some lexicalized verb forms of Old Written Estonian. In Mati Erelt (ed.) *Estonian: Typological Studies V*, vol 18, 26–76. Tartu Ülikooli eesti keele õppetooli toimetised [Publications of Chair of Estonian language of University of Tartu].
- Habicht, Külli & Pille Penjam. 2007. Kaassõna keeleuurija ja -kasutaja käsituses [Adpositions as viewed by a linguist and by a language user]. *Emakeele Seltsi aastaraamat* 52 (2006), 51–68.
- Heine, Bernd. 1997. *Cognitive foundations of grammar*. New York/Oxford: Oxford University Press.
- Heine, Bernd. 1999. Grammaticalization chains across languages: An example from Khoisan. In Spike Gildea (ed.). *Reconstructing grammar. Comparative linguistics and grammaticalization*. Typological Studies in Language 43, 177–200. Amsterdam/Philadelphia: John Benjamins Publishing Company.

- Heine, Bernd, Ulrike Claudi, & Frederike Hünemeyer. 1991. *Grammaticalization: a conceptual framework*. Chicago/London: University of Chicago Press.
- Heine, Bernd, & Tania Kuteva. 2002. *The World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press.
- Heine, Bernd, & Tania Kuteva. 2007. *The Genesis of Grammar. A Reconstruction*. Studies in the Evolution of Language. Oxford/New York: Oxford University Press.
- Hoffmann, Sebastian. 2004. Are low-frequency complex prepositions grammaticalized? In Hans Lindquist, Christian Mair (eds.) *Corpus Approaches to Grammaticalization in English*. Studies in Corpus Linguistics 13, 171–210. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Hoffmann, Sebastian. 2005. *Grammaticalization and English complex prepositions: a corpus-based study*. Routledge Advances in Corpus Linguistics. London: Routledge.
- Hopper, Paul J., & Elizabeth Closs Traugott. 2003 [1993]. *Grammaticalization*, 2nd edn. Cambridge: Cambridge University Press.
- Jürine, Anni. 2011. Vormierinevused tähenduserinevuse ilminguna: kaassõnaühendite grammatiseerumine konstruktsioonides [Change in Form as a Manifestation of Change in Meaning: Estonian Adpositional Phrases]. *Keel ja Kirjandus* 12, 899–916.
- Jürine, Anni. 2016. *The development of complex postpositions in Estonian: a case of grammaticalization via lexicalization*. Dissertationes Philologiae Estonicae Universitatis Tartuensis vol. 38. Tartu: University of Tartu.
- Jürine, Anni, & Külli Habicht. 2013. Kaassõnade tsükliline areng [The cycle of postpositions]. *Keel ja Kirjandus* 12, 736–749.
- Jürine, Anni, & Külli Habicht. 2017. Grammaticalization of complex items: Estonian *nii et* ‘so that’, ‘so’. *Journal of Estonian and Finno-Ugric Linguistics*, 8(2), 35–58.
- Karels, Rudolf. 2005. Taas probleemidest sõnaliigi määramisel [On the problems of how to determine the word class, once again]. *Estonian Papers in Applied Linguistics* 1 (2004), 53–70.
- Kasik, Reet. 2015. *Sõnamoodustus [Word Formation]*. Eesti Keele Varamu vol. 1. Tartu: University of Tartu Press.
- Kemmer, Suzanne, & Michael Barlow. 2000. Introduction: A usage-based conception of language. In Michael Barlow, & Suzanne Kemmer (eds.) *Usage-Based Models of Language*, 7–28. Stanford: Center for the Study of Language and Information.
- Langacker, Ronald W. 1977. Syntactic reanalysis. In Charles N Li (ed.) *Mechanisms of syntactic change*. Austin: University of Texas Press, 57–140.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar. Volume 2: Descriptive Applications*. Stanford: Stanford University Press.
- Langacker, Ronald W. 2012. *Essentials of Cognitive Grammar*. Oxford/New York: Oxford University Press.
- Lehmann, Christian. 1985. Grammaticalization: Synchronic variation and diachronic change. *Lingua e stile* 20(3), 303–318.
- Lehmann, Christian. 2002. New Reflections on Grammaticalization and Lexicalization. In Ilse Wischer, Gabriele Diewald (eds.) *New Reflections on Grammaticalization*. Typological Studies in Language 49, 1–18. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Mair, Christian. 2004. Corpus linguistics and grammaticalization theory: Statistics, frequencies, and beyond. In Hans Lindquist, Christian Mair (eds.) *Corpus Approaches to Grammaticalization in English*. Studies in Corpus Linguistics 13, 121–150.

- McMichael, Andrew. 2006. The A's and BE's of English Prepositions. In Patrick Saint-Dizier (ed.), *Syntax and Semantics of Prepositions*. Text, Speech and Language Technology. Vol. 29. Springer, 43–56.
- Moirón, Begoña Villada, Gosse Bouma. 2003. A corpus-based approach to the acquisition of collocational prepositional phrases. In *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002: Copenhagen, Denmark, August 13–17, 2002*, 153–158.
- Ojutkangas, Krista. 2001. *Ruumiinosannimien kieliopillistuminen suomessa ja virossa*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Palmeos, Paula. 1985. *Eesti keele grammatika II. Kaassõna [The Grammar of Estonian II. Adposition]*. Tartu: TRÜ trükikoda.
- Sigurd, Bengt. 1993. The problems of multi-word prepositions and subjunctions. *Working Papers in Linguistics*. Vol. 40, 197–212.
- Svorou, Soteria. 1994. *The Grammar of Space*. Typological Studies in Language 25. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Tauli, Valter. 1966. *Structural Tendencies in Uralic Languages*. Indiana University Publications. Uralic and Altaic Series vol. 17. London-The Hague-Paris: Mouton & Co.
- Viitso, Tiit-Rein. 2003. Structure of the Estonian language. In Mati Ereht (ed.) *Estonian Language. Linguistica Uralica Supplementary Series* vol. 1, 9–129. Tallinn: Estonian Academy Publishers.
- Veismann, Ann, & Mati Ereht. 2017. Kaassõnafraas [The Adpositional Phrase]. In Mati Ereht, Helle Metslang (eds.) *Eesti keele süntaks [Estonian Syntax]*. Eesti Keele Varamu vol. 3, 446–462, Tartu: University of Tartu Press.
- Waters, Catherine. 2009. The preposition cycle in English. In Elly van Gelderen (ed.) *Cyclical change*. Linguistics Today vol. 146, 285–301. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Maria Usacheva and Timofey Arkhangelskiy

8 Complex and simple postpositions: The split in Permic

Abstract: This paper is devoted to complex postpositions in Permic languages (namely, in Udmurt and in Komi varieties). We argue that only morphological complexity is attested in most Permic idioms, i.e. most postpositions attach case and possessive markers and/or their stems consist of two roots. However, postpositions in certain Udmurt dialects have the option of having nominal and pronominal dependents in the genitive, in addition to the nominative or unmarked form. Postpositions in Komi varieties, both morphologically complex and morphologically simple, strongly prefer nominative marking of dependents. Komi and Udmurt are quite closely related, but have existed in different contact situations for quite some time. Therefore, it could be possible that the difference has been partly induced by language contact, in particular, by Turkic influence, but this hypothesis appeared not to be supported by the data.

Keywords: Permic languages, syntax of postpositional phrases, contact influence

1 Introduction

This paper is devoted to complex postpositions in Permic languages (namely, in Udmurt and in Komi varieties).¹ We argue that only morphological complexity is attested in most Permic idioms, i.e. most postpositions attach case and possessive markers and/or their stems consist of two roots. However, postpositions in certain Udmurt dialects have the option of having nominal and pronominal dependents in the genitive, in addition to the nominative or unmarked form. We demonstrate that the genitive construction is more syntactically complex, and that syntactic complexity is therefore also attested. This phenomenon could be caused by Turkic influence, but this hypothesis appeared not to be supported by

¹ This research was supported by the Russian Foundation for Basic Research (grant No. 16-24-17003).

Maria Usacheva, The Institute of Linguistics, Russian Academy of Sciences, Moscow, Russian Federation. mashastroeva@gmail.com

Timofey Arkhangelskiy, Universität Hamburg / Alexander von Humboldt Foundation, Hamburg, Germany. timarkh@gmail.com

<https://doi.org/10.1515/9783110686647-009>

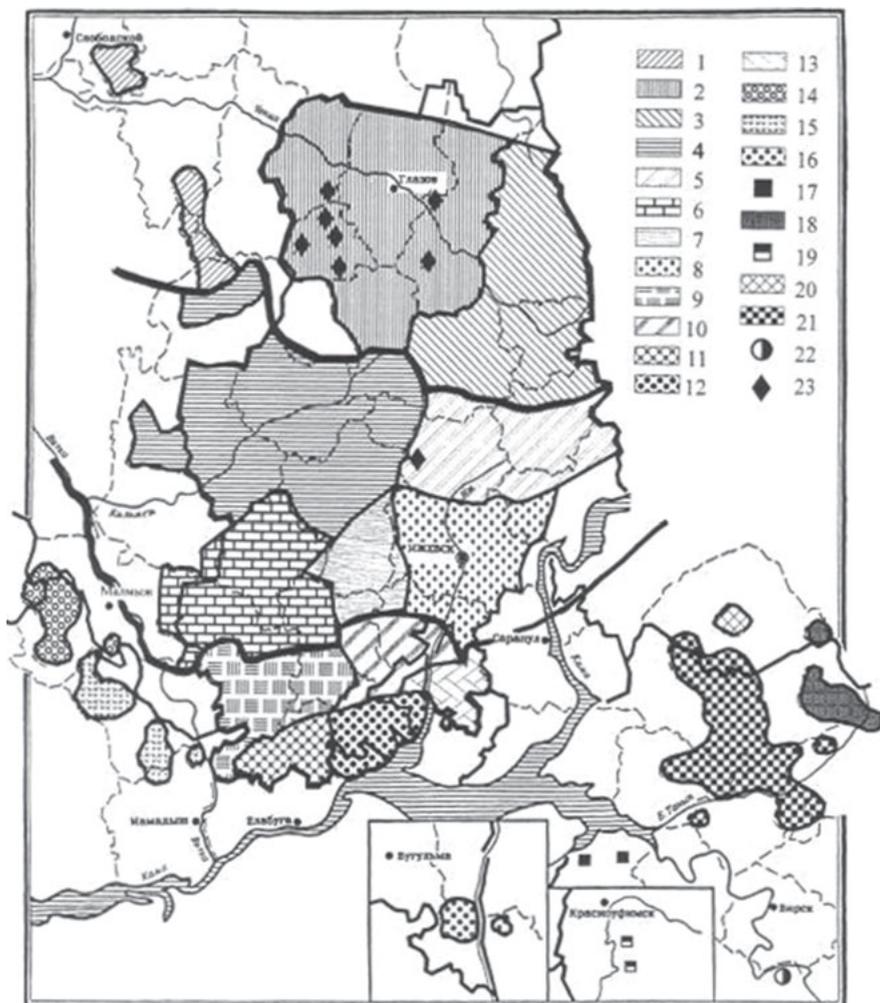
the data. The structure of the paper is as follows. In section 2, we provide a sketch of sociolinguistic, geographical and structural features of Permic languages and describe the sources of the data we use and the methods of our investigation. Section 3.1 contains a brief description of the types of postpositions in Permic languages. In this section, we explain why most Permic postpositions are complex in terms of the present book. Section 3.2 is devoted to syntax of complex postpositions in Udmurt, and section 3.3, in Komi. In section 4, we describe some grammatical features of Permic languages which can be connected to the syntax of postpositional phrases. In section 5, we argue that the complexity of syntax in postpositional phrases could be the result of a contact-induced change, caused by Turkic influence, but this hypothesis is not supported by the corpus data. Section 6 contains conclusions.

2 Languages and sources

2.1 Sociolinguistic and geographical features

Permic languages form a subgroup of the Uralic language family. This group consists of Udmurt and a continuum of closely related Komi varieties with two literary standards, Zyrian and Permyak. Permic languages are spoken in several regions of Russia to the west of the Ural mountains (Udmurtia, Tatarstan, Kirov region, Komi republic, Perm krai, Arkhangelsk province, Nenets autonomous district). It is a vast territory, so the dialectal variation is significant (though speakers of most of the dialects inside languages do understand each other). Udmurt has four main varieties (groups of dialects), which differ from each other primarily in vocabulary and phonetics (Kelmakov 1998) and, to a much lesser extent, in morphology. These varieties are commonly referred to as Northern Udmurt (zones 1–3 on the Map 1 below), Central Udmurt (zones 4–8), Southern Udmurt (zones 9–22) and Beserman Udmurt (zone 23).

There are also several dialects of the Komi languages. The most striking differences between them lie in the field of phonology (systems of phonemes and stress) and in case systems (Batalova 1982: 155–158). Zyrian dialects represented on Map 2 below are Luza-Letka (1), Upper Sysola (2), Upper Vychegda (3), Pechora (4), Izhma (5), Lower Vychegda (6), Udora (7), Vymj (8), Central Sysola (9) and Pryshtyvkhar (10). Permyak is divided into two varieties, Upper Kama (11) and Yazva (12), and 8 dialects: On (13), Lower Inva (14), Kudymkar-Inva (15), Nerdva (16), Kochevo (17), Kosa-Kama (18), Mysy (19) and Upper Lupya (20).



Map 1: Udmurt dialects (Nasibullin 2000: 73).

According to the 2010 Census, there are about 340,000 speakers of Udmurt and about 219,000 speakers of Komi.² Udmurt has undergone influence from neighboring Turkic languages and (to a lesser extent) from Russian, while Komi has been significantly influenced by Russian. The vast majority of Permian speakers are bilingual in Russian.

² http://www.gks.ru/free_doc/new_site/perepis2010/croc/perepis_itogi1612.htm.

concerning information structure. Compare (5) and (6) or in the cases of substantiation (7–8):

- (3) *Mi mąji-eś=ńi val=no.*
 we.NOM old-PL.ADJ=already be.PST=and
 ‘We were already old as well.’
 Northern Udmurt (corpus)
- (4) *Eśin-nez-įs i mij ńeveśoląj-aś.*
 window-PL-P.3SG and what sad-PL.ADJ
 ‘Even the windows are sad.’
 Upper Kama Permyak (Batalova 1975: 166)
- (5) *So lapeg-a-z tatĉâ vand-âl-âtek târ-o-m*
 that.NOM shallow-ILL-P.3SG here.ILL cut-ITER-CVB.CAR put-FUT-1PL
tiń piĉi pomidor-įos-tâ.
 here small tomato-PL-ACC.PL
 ‘In this shallow [plate] we will put small tomatoes, uncut.’
 Beserman Udmurt (corpus)
- (6) *Pomidor-ez tânad ząk-eś tiń otân wań.*
 tomato-P.3SG you.GEN big-PL.ADJ here there.LOC be.PRS
 ‘You have big tomatoes over there [and the small ones here].’
 Beserman Udmurt (corpus)
- (7) *Kąńa ulĉa-os=pe otįn? Vįl-įos-įz=na jeśo?*
 how.many street-PL=CIT there new-PL-P.3SG=also also
 ‘How many streets are there (they said)? There are also the new ones, right?’
 Northern Udmurt (corpus)
- (8) *Ďtpįr garįįšt-i, i jeźįt-ik-aś da joniĉ-ik-aś*
 once dig-PST.1SG and white-DIM-PL.ADJ and big-DIM-PL.ADJ
pet-i-s-a.
 come.out-PST-3-PL
 ‘I have dug once – white and big [potatoes] appeared.’
 Upper Kama Permyak (Batalova 1975: 166)

Noun phrases are always left-branching, which means that the nominal head is always on the right border of the noun phrase. One of the ways of marking a dependent in a noun phrase is to use one of the two genitive suffixes. Genitive marking is

strictly defined by the syntactic position of the noun phrase: dependents of nouns in direct object position are marked by a suffix *-IV³ś*, traditionally called ablative and labeled as “second genitive” in this paper. In all other positions, the “primary” genitive *-IVn* is used. Compare examples (9a) and (9b) below, which differ in syntactic position of the word *ćoj* ‘sister’. While the “primary” genitive is only used for marking dependents in noun phrases, the suffix of genitive 2 also has a range of other functions. It can mark material (10), it is used in comparative constructions (11) (only in Udmurt), marks arguments of several verbs of emotion and speech (12), etc.

- (9a) *Adža ćoj-ljś pałto-sə.*
 see.PRS.1SG sister-GEN2 coat-ACC.P.3SG
 ‘I see the sister’s coat.’
 Standard Komi Zyrian (Bubrikh 1946: 44)
- (9b) *Ćoj-lən pałto-įs mića.*
 sister-GEN coat-P.3SG beautiful
 ‘The sister’s coat is beautiful.’
 Standard Komi Zyrian (Bubrikh 1946: 44)
- (10) *Ku-leś siņjs-ez prišepka pjr pott-ono.*
 skin-GEN2 thread-P.3SG pin through let.out-DEB
 ‘The skin thread has to be fed through a pin.’
 Standard Udmurt (corpus)
- (11) *Kartoška tue so bātca-eś bud-em,*
 potato this.year that size-PL.ADJ grow-PST2
vedra-je das-leś tros ug tērđ.
 bucket-ILL ten-GEN2 many NEG.PRS.3 contain
 ‘Potatoes have grown so big this year that a (standard) bucket can contain no more than ten.’
 Beserman Udmurt (field notes)
- (12) *Urmem pudo.život ug kjška ađami-leś.*
 rabid domestic.animal NEG.prs.3 be.afraid.of human-GEN2
 ‘Rabid domestic animals are not afraid of a human.’
 Standard Udmurt (corpus)

3 Here and below *V* in roots and morphemes indicates vowels which are different in different dialects.

Third singular (in both languages) and second singular (more in Komi than in Udmurt) possessive suffixes are special in that they can perform functions beyond marking possessive relations. They are currently undergoing pragmaticization, starting to perform discourse and pragmatic functions such as marking focus of attention or anchoring to implicit elements, time, place, second and third order entities, or to linguistic and situational context (Fraurud 2001). For example, the third singular possessive suffix in the form *nâĺłosâz* in (13) marks focus of attention, while the same element in *odigâz* is an anchor to a referent ‘group’, which is not mentioned in the utterance:

- (13) - *Nâĺłos-âd=a sâł-o?*
 girl-PL-P.2SG=Q stand-PRS.3PL
 ‘- Is it your girls (the girls from your pictures) who are standing?’
 - *Nâĺłos-âz odig-âz plîta ušt-em aź-ân sâł-e.*
 girl-PL-P.2SG one-P.3SG stove open-NMZ place-LOC stand-PRS.3SG
 ‘- One [group] of the girls is standing near the stove opening.’
 Beserman Udmurt (corpus)

In example (14) from Pechora Komi Zyrian, the focus of attention is marked by the second singular possessive suffix:

- (14) *Mi Sarju.pas kuźa-ïd kaj-a-m.*
 we.NOM PN along-P.2SG move.up-PRS-1PL
 ‘We were going along the Sarju pas road (the speaker told about before).’
 Pechora Komi Zyrian (corpus)

In (15), the second singular possessive suffix is used to mark the referent previously mentioned by the listener (rather than the listener’s relative, as would be the case if that suffix had purely possessive meaning):

- (15) *Tož’-iko mon boĺno Paraś abi-de!*
 know-PRS.1SG I.NOM as.if PN granny-P.2SG.ACC
 ‘As if I knew that granny Praskovja (you have told me about before)!’
 Beserman Udmurt (corpus)

Permic languages are rich in spatial postpositions and cases. There are up to eight primary local cases (in Beserman Udmurt: Arkhangelskiy and Usacheva 2015) most of which also can combine with one (Udmurt: Lyukina 2008) or more (Komi-Permyak: Baker 1985) localization suffixes.

2.2 Sources and methodology

The present paper is based on the data of both literary varieties of Permian languages and their dialects. We use four main sources of data.

1. Corpus data come from the following sources:
 - Beserman Udmurt data comes from the Beserman spoken corpus of speech⁴ (ca. 65,000 tokens); the Multimedia-aligned Beserman corpus⁵ (ca. 38,000 tokens); and the Corpus of usage examples from the Beserman-Russian dictionary and field questionnaires (ca. 83,000 tokens).
 - The data from other Udmurt dialects comes from the Sound-aligned dialectal Udmurt corpus by Ekaterina Georgieva (Arkhangelskiy and Georgieva 2018).
 - Standard Udmurt data comes from the corpora of Standard Udmurt.⁶ When not explicitly indicated otherwise, the literary Udmurt examples in this paper come from the Corpus of Standard Udmurt.
 - Standard Komi data comes from the Komi-Zyrian literary corpus.⁷
 - Dialectal Komi data comes from the Pechora Komi-Zyrian spoken corpus.⁸
2. Yazva Komi data was mostly gathered during two series of referential communication experiments conducted by the authors in the villages of Vankova and Antipina (Krasnovishersk district, Perm krai, Russia) in 2016–2017. The series included retelling of the mute cartoon fragment which was taken from “Pyotr and Petrusha” (director Yury Pronin, “Pilot” studio, 2005) and adapted, and referential communication tasks of two types. During the first series of experiments, the Directors were asked to describe a black-and-white picture with colored figures of people, domestic animals, and rural technical equipment glued onto it. The Matchers had to put the identical colored figures on their example of the black-and-white picture in the same configuration. In the second series, the Directors described the motion of figures through a model of the area where the experiments took place, and the Matchers repeated the motion.

The same experiment was conducted by Maria Usacheva and Olga Biriuk in a Beserman village of Shamardan in 2010. Experimental stimuli are described in detail in (Biriuk and Usacheva 2012). We have thus two

⁴ <http://beserman.ru/corpus/search>.

⁵ <http://multimedia-corpus.beserman.ru/search>.

⁶ http://udmurt.web-corpora.net/index_en.html.

⁷ <http://komicorpora.ru>.

⁸ <http://web-corpora.net/KomiTexts/>.

comparable experimental subcorpora, Yazva Komi and Beserman, which simplifies the comparison of the two varieties.

3. We also use the elicitation data that we collected in a Beserman village of Shamardan (Yukamenskoye district, Udmurtia, Russia; 2003–2018) and in Yazva villages Vankova and Antipina (Krasnovishersk region, Perm krai, Russia; 2016–2017).

3 Syntax of postpositions in Permic languages

3.1 Inflected postpositions vs non-inflected ones

This volume is based on the definition of a complex adpositions (Fagard et al.: present volume). According to this definition, these language units are morphologically complex, and the parts they could be split into perform different functions, e.g. one of them introduces a constituent, while the others establish a syntactic and/or semantic relationship between the constituent and another element of the clause. Applying this definition to Permic (as well as many other Uralic) languages yields unusual results.

Most postpositions in Permic languages group into series that consist of a stem inflected for case and, possibly, other nominal categories. The stem generally denotes a localization, i.e. indicates the position of a certain spatial area with respect to the landmark, while the case suffixes that attach to them express either orientation (direction of motion) or a syntactic relationship. Nominative, or caseless, forms of such postpositions usually either do not exist or have a different, non-postpositional interpretation and syntactic properties. In Table 1, one can see an example of such series in Standard Udmurt, consisting of the stem *dor*, which could be translated as ‘the space near the landmark (L)’, in all spatial case

Table 1: Udmurt inflected postposition *dor* ‘near’.

Form	Case	Meaning
<i>dor-jn</i>	LOC	located near L
<i>dor-jš</i>	EL	from a place near L
<i>dor-j / dor-e</i>	ILL	to L, towards L
<i>dor-jšen</i>	EGR	starting from a place near L
<i>dor-ož</i>	TERM	up to the place near L
<i>dor-ti</i>	PROL	moving through the vicinity of L

literary Udmurt, in addition to its more general spatial meaning of close vicinity. It must be noted however that even dialects of the same language are different in this respect: for example, *dor* in Beserman can only mean ‘close vicinity’, even in the core cases. The difference between Permic languages is, of course, even greater: the same item *dor* in Komi has a much greater range of nominal uses than it has in any Udmurt dialect, cf. (19). When a postposition is used with the nominal plural suffix, this suffix normally refers to the plurality of its dependent (10). The possibility of plural marking varies across idioms and postpositions.

- (17) *Projekt-len valt-iš mug-ez mĭn-iš*
 project-GEN lead-PTCP.ACT aim-P.3SG go-PTCP.ACT
vetl-iš-jos-lĭ ĵm puš-len višon-jos-ĵz=no
 travel-PTCP.ACT-PL-DAT mouth inside-GEN disease-PL-P.3SG=ADD
so-os-leš utiškön šarĭš ivort-ĭnĭ.
 that-PL-GEN2 protection about inform-INF
 ‘The primary aim of the project is to inform the travelers about oral cavity diseases and about the methods of protecting from them.’
 Standard Udmurt (corpus)
- (18) *Ķĭlkak muket ĵugdur korka puš-lĭ arberi-os*
 completely other circumstance house inside-DAT thing-PL
lašan.
 from.side.of
 ‘As for house furniture (lit. furniture for inner space of house), it is quite a different story.’
 Standard Udmurt (corpus)
- (19) *Nastajĭin-lĭš ĵur dor-sə vəl-i, kul-əma, stav-na-s*
 PN-GEN2 head near-ACC.P.3SG be-PST.3SG die-PST2.3SG all-INS-P.3SG
šarkñit-əma.
 crack-PST2.3
 ‘He [bit] Nastaychin in the head, he died, all [his head] cracked.’
 Pechora Komi-Zyrian (corpus)
- (20) *Viĭ korka dor-jos-a-zĕ kăšpu puk-t-ĕl-ĕmĕn.*
 new house near-PL-LOC-P.3PL birch sit-CAUS-ITER-RES
 ‘Birches are planted near the new houses.’
 Beserman Udmurt (field notes)

Apart from the inflected postpositions, which share a lot of properties with nouns, there are also “true”, non-inflected postpositions in Permic languages. They cannot attach markers of case and number¹⁰ and therefore should be treated as simple postpositions in terms of the present book. Some of them govern a case, while others are much more like adverbs than like nouns in their distributional properties. These items, such as Beserman *wamen* ‘across’, can be used both without dependents and possessive suffixes (26–28).

- (26) *Reka wamen ez=uk pot-â.*
 river across NEG.PST.3=FOC go.out-NEG.SG
 ‘She did not cross the river, did she?’
 Beserman Udmurt (corpus)
- (27) *Wamen ez=uk pot-â.*
 across NEG.PST.3=FOC go.out-NEG.SG
 ‘She did not go straight across, did she?’
 Beserman Udmurt (field notes)
- (28) *Reka-ti wamen pot-i-z.*
 river-PROL across go.out-PST-3SG
 ‘She went across the river, straight across’
 Beserman Udmurt (corpus)

In both languages, all non-inflected postpositions still can have possessive suffixes, thus somewhat contradicting their label. This can happen under two different scenarios. Under the first scenario, the possessive marker references the dependent. Any of the six available possessive markers can cooccur with non-inflected postpositions that allow this construction, while case and number markers are disallowed. An Udmurt example can be seen in (29), where the postposition *ponna* ‘for’ bears a 1SG possessive suffix that stands for the omitted 1SG pronoun (‘me’). Such uses of *ponna* are widespread in some dialects, but rather infrequent in the standard language. Most simple postpositions, however, do not allow such use even in the dialects.

¹⁰ In Udmurt, all postpositions can attach possessive suffixes, but some of them only in the case of ellipsis (see below).

- (29) *Noš osk-em-di ponna-m duno*
 but believe-NMZ-P.2PL for-P.1SG valuable
 ‘But your trust is valuable for me.’
 Udmurt (Social media corpus)

The second scenario allows only for certain (etymologically) possessive markers, provided they are used in a non-possessive function. These are third singular possessive suffix for Udmurt and third and second singular possessive suffixes for Komi. The purpose of the possessive in this construction is different from the previous scenario, since it does not signal a reference to a dependent. Apart from having purely possessive meaning, these markers have developed a range of “discourse” meanings not associated with possession, as well as certain morphosyntactic properties that the “regular” possessives lack. For example, third singular possessive marker can be used in Udmurt to mark a contrastive topic or a semi-activated referent (Serdobolskaya, Usacheva, and Arkhangelskiy 2019); see also Section 2.

The circumstances under which those markers can appear on a simple postposition in Udmurt normally include head ellipsis. The third singular possessive marker that would appear on a nominal head modified by the postpositional phrase is moved to the head of the modifying postpositional phrase together with the case and number suffixes (if any), if that nominal head is omitted. In the last sentence of (30), the nominal head ‘video’ would be marked with third singular possessive suffix because it would be a contrastive topic, had it been there. Since it was omitted, its nominal morphology was transferred to the postpositional phrase ‘about me/you’, which modified it. Example (31) shows that the same simple postposition *šarâš* ‘about’ would not allow other possessive markers under the previous scenario:

- (30) – *Ta viđik-ân kwara-ez ľab. – Kće viđik-ân?*
 this video-LOC voice-P.3SG weak which video-LOC
 – [*Mon / ton šarâš*]-*a-z.*
 [I.NOM / you.NOM about]-LOC-P.3SG
 ‘The voice is low in this video. – In which video? – In the one about me / you.’
 Beserman Udmurt (field notes)
- (31) *Mon ćakla-šk-iško tĕnad anaj.ataj-os-ĕd šarâš /*
 I.NOM think-DETR-PRS.1SG you(SG).GEN parents-PL-P.2SG about
 **anaj.ataj-os šarâš-ed.*
 parents-PL about-P.2SG
 ‘I think about your parents.’
 Beserman Udmurt (field notes)

In Komi, all postpositions can attach 3SG and 2SG possessive suffixes under this second scenario, as can be seen in the example (32) featuring a simple postposition *kuža* ‘along’:

- (32) *Šor kuža-ys=śa i gort-a-s lokt-i-s.*
 brook along-P.3SG=EMPH and home-ILL-P.3SG come-PST-3SG.
 ‘Going along the brook is how he returned home.’
 Pechora Komi-Zyrian (corpus)

The first scenario does not exist in Komi. Unlike in Udmurt, possessors in Komi noun phrases and postpositional phrases normally do not trigger agreement in person and number on the head, which can have either no possessive marker at all (33), or one of the two “discourse” possessives (34).

- (33) *Kodi menam volpaś-in valajt-ć-ama?*
 who.NOM I.GEN bed-LOC roll-DETR-PST2.3SG
 ‘Who has been rolling in my bed?’
 Pechora Komi-Zyrian (corpus)

- (34) *O-z=kə šed, menam Vova-lən terpeńña-ǰd abu.*
 NEG.PRS-3=if OCCUR.NEG I.GEN PN-GEN patience-P.2SG NEG.EXIST
 ‘If he can’t [catch some fish], my [husband] Vova loses his patience.’
 Pechora Komi-Zyrian (corpus)

As one can see from the examples above, even non-inflected postpositions can attach certain possessive markers and, under some circumstances, even case and number suffixes. Nevertheless, we believe that they still fall under the definition of simple postpositions. Possessive morphology under the first scenario is just an alternative way of introducing the dependent. Under the second scenario, all morphology, including case markers, is external to the postposition and does not contribute to its meaning. This contrasts with the inflected postpositions, where the case marker always denotes an orientation.

In this section, we have shown that inflected postpositions in Permic languages are complex, and non-inflected postpositions are not. Additionally, the latter are not postpositions at all in a certain class of uses. So at a first glance, the continuum of postpositions in Permic languages can be represented as in Figure 1.

Although this continuum is organized according to the presence of certain morphosyntactic properties (broadly speaking, items on the left of the continuum allow more than those on the right), it also reflects the grammaticalization cline for many, although not all, postpositions. It is common for a non-inflected

fully inflected P-s (relational nouns) → partly inflected P-s → non-inflected P-s
 complex postpositions



Figure 1: Continuum of Permic postpositions.

postposition to originate from a petrified case form of an inflected postposition. For example, the Udmurt postposition *šarâš* clearly used to include the elative suffix *-âš*, which is not recognizable anymore because the rest of the series did not survive.

Here we come to the point of variation among Permic languages. We will show that this picture is true for Komi varieties. The situation is more complex in Udmurt, and the reason lies in the syntactic properties of postpositional and noun phrases. First we will describe the situation in Udmurt and then in Komi.

3.2 Udmurt. Fully inflected postpositions: Degrees and levels of complexity

3.2.1 Syntactic complexity

Let us now look at fully inflected postpositions in Udmurt more closely with a focus on how their dependents are marked. If we consider all Udmurt varieties, we observe that generally, their dependents can be marked either in the nominative or in the genitive. This is quite expected, given that their diachronic source is nouns and the fact that ordinary noun phrases with possessors allow for (superficially) the same two constructions. Nevertheless, the dependent marking is a point of significant variation between Udmurt dialects.

In Beserman, both types are attested and frequent (35, 36).

- (35) *Čašja ul-âš tože tetć-âsa pot-e*
 forest bottom-EL also jump-CVB come.out-PRS.3SG
bakća ber-âš piňal-los-len pumât-a-z žićâ.
 garden behind-EL child-PL-GEN across-ILL-P.3SG fox
 ‘From beneath the forest [on a picture], a fox comes jumping from behind the garden towards the kids [facing them].’
 Beserman Udmurt (corpus)

- (36) *Ben, proval-ez maġazĭn pumâĭ-âĭn=uk.*
 yes ravine-P.3SG store across-LOC=FOC
 ‘Yes, it’s opposite the store, that ravine.’
 Beserman Udmurt (corpus)

The nominative seems to be the most frequent option. Empirical observations show that the likelihood of having a dependent in the genitive is primarily determined by two factors:

- Lexical class of the dependent. First and second personal pronouns as dependents are much more likely to appear in genitive than in nominative. Third personal pronouns have a clear preference for the nominative, however, they can still be encountered with the genitive.
- Particular postposition. Postpositions differ in their syntactic properties, probably due to their different stages of grammaticalization. Some postpositions attach dependents in the genitive much more often than others.

In literary Udmurt, the nominative is the default choice for most postpositions, both for nominal and pronominal dependents. Table 2 demonstrates the relative frequencies (in items per million, further contracted as ipm) of dependents in the nominative and in the genitive for a sample of eight different inflected postpositions in Standard Udmurt. As can be seen from the table, seven out of eight postpositions have a very strong preference for the nominative. Moreover, most of the genitive examples seem to come from the non-standard language, such as transcripts of spoken interviews. This preference becomes only a little weaker if the dependent is a personal pronoun.

Table 2: Relative frequencies of dependents in different cases with different postpositional stems (Corpus of Standard Udmurt).

dependent	Postpositional stem							
	<i>dor</i> ¹¹ ‘near’	<i>puš</i> ‘inside (container)’	<i>pəl</i> ‘inside (material)’	<i>bord</i> ‘at’	<i>vĭl</i> ‘on’	<i>pal</i> ‘side’	<i>šer</i> ‘back’	<i>až</i> ‘front’
Pron-Gen	3.7	8	0.8	0.6	2.6	0.2	2.1	4
Pron-Nom	278	24	575	193	97	44	42	59
N-Gen	2.8	18	0.3	1.8	2.5	0	2.6	3.4
N-Nom	882	208	1254	821	2408	675	761	761

¹¹ Not including a handful of contexts where *dor* is used in the sense ‘country, homeland’.

Table 3 shows the numbers for two of these postpositions in Beserman Udmurt, for comparison. The zeroes in the “nominal, genitive” row are due to the small size of available corpora (with a total size of slightly over 100,000 words, even a single occurrence would translate into 9.7 ipm). The first two rows, however, clearly show that pronominal marking is different in Beserman.

Table 3: Relative frequencies (in ipm) of dependents in different cases with different postpositional stems (Corpora of spoken Beserman).

dependent	Postpositional stem	
	<i>dor</i>	<i>vâl (=vjl)</i>
Pron-Gen	107	107
Pron-Nom	417	243
N-Gen	0	0
N-Nom	3873	5437

It is interesting that, according to our preliminary observations, the literary Udmurt contexts that contain a pronominal dependent in the genitive tend to come predominantly from texts written by native speakers of the dialects that have a history of contact with the neighboring Turkic languages (Tatar and Bashkir). First of all this includes the translation of the New Testament by Mikhail Atamanov, who was born in a village very close to the border of Tatarstan. Importantly this translation contains a number of other dialectal features characteristic for the Udmurts living in Tatarstan and neighboring districts of Udmurtia, such as a hortative in *-ome*.

If a postposition has a dependent in the genitive, the phrase governed by it becomes more complex syntactically. Udmurt postpositional phrases with dependents in the genitive must bear a possessive suffix on the head. Postpositional phrases with dependents in the nominative resemble compounds in a certain way: nothing can be inserted between the head and the dependent, and the dependent-head order is always invariant. The corresponding construction with nominal heads has been described in these terms by Fejes (2005), and, from a different theoretical perspective, as pseudoincorporation by Simonenko and Leontyev (2012: 323). However, there are arguments against an analysis of that kind for phrases with nominative dependents headed by inflected postpositions. One of them is the fact that in postpositional phrases with dependents denoting inanimate objects, the

possessive suffix, which is semantically related to the dependent, can attach either to the head or to the dependent without any perceptible semantic difference:

- (37) *So tãb-i-z korka-je dor-e / korka dor-a-m*
 s/he go.up-PST-3SG house-P.1SG near-ILL house near-ILL-P.1SG
uža-nã.
 work-INF
 ‘He went up to my house to work.’
 Beserman Udmurt (field notes)

In corresponding phrases with nominal heads, however, no number or possessive marking on the dependent is possible, which is expected for a compound.

3.2.2 Morphological complexity

Thus, we argue that in Udmurt postpositional phrases of the type “noun/pronoun-GEN + inflected postposition” are more syntactically complex than the postpositional phrases with nominative dependents. This syntactic complexity correlates with morphological complexity. In both Beserman and literary Udmurt (as well as in Komi, see the Section 3.3), there are morphologically complex postpositions. They comprise a postpositional stem such as *pal* ‘side’ modified by either a nominal/postpositional stem like *ber* ‘rear’ or adjectives such as *valla* ‘upper’ or *ulla* ‘lower’, which are phonologically and syntactically inseparable from the modified stem. The type of the modifier does not influence the morphosyntactic behavior of the postposition, so we treat all these cases in the same way. All of these morphologically complex postpositions inflect for case. Unlike postpositions with morphologically simple stems, items with complex stems seem to have more uniform properties across Udmurt varieties, based on the available corpus data. Specifically, they only allow pronominal dependents in the genitive. If their dependent is nominal, the difference between literary Udmurt and Beserman can be seen, although it is less drastic than with items that have simple stems. In Beserman, these postpositions allow for both genitive and nominative nominal dependents (38–40), with a preference for the nominative. In literary Udmurt, however, only the genitive seems to be allowed (Table 4). Additionally, such postpositions seem to be less frequent in literary Udmurt.

- (38) *Sre so traktor valla.pal-an lu-e=uk kureg-ed.*
 then that tractor above-LOC become-PRS.3SG=FOC chicken-P.2SG
 ‘Then it is above the tractor, that chicken.’
 Beserman Udmurt (corpus)
- (39) *Žažeg-jos-len valla.pal-a-z kâk kureg pi-os.*
 goose-PL-GEN above-LOC-P.3SG two chicken kid-PL
 ‘There are two chickens above the geese.’
 Beserman Udmurt (corpus)
- (40) *So-len ulla.pal-a-z kâkt-eti [puš-ez].*
 that-GEN underneath-LOC-P.3SG two-ORD rod-P.3SG
 ‘There is a second rod under it.’
 Beserman Udmurt (corpus)

However, certain morphologically complex postpositions, such as *až.pal* ‘in front of’ always take dependents in the genitive, even in Beserman.

We can conclude that morphologically complex postpositions generally have a much stronger preference for the genitive marking of dependents in Udmurt. It means that they tend to head phrases which are syntactically more complex than the groups headed by morphologically simple postpositions. Table 3 indicates the number of occurrences for dependents with different morphological markers in phrases headed by morphologically complex inflected postpositions, based on the Corpus of Standard Udmurt:

Table 4: Relative frequencies of dependents (in ipm) in different cases with morphologically simple and morphologically complex postpositions (Corpus of Standard Udmurt).

dependent	Postpositional stem		
	<i>až.pal</i>	<i>ber.pal</i>	<i>šer.pal</i>
Pron-Gen	1.4	0.1	0
Pron-Nom	0	0	0
N-Gen	3.3	0.4	0
N-Nom	0	0	0

We propose the following continuum of postpositional complexity in Udmurt on the basis of data cited in this section (Figure 2).

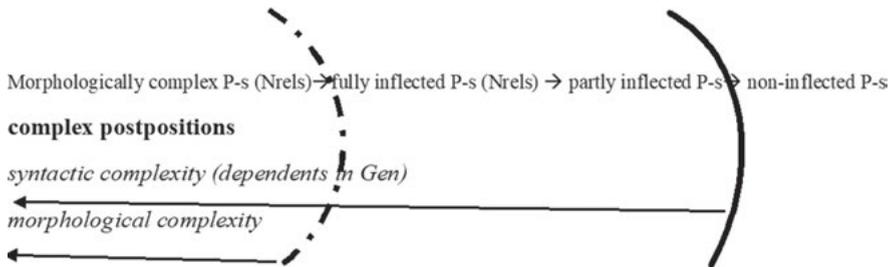


Figure 2: Continuum of Udmurt postpositions.

3.3 Syntax of postpositions in Komi

Postpositions in Komi do not take nominal dependents in the genitive. Both morphologically simple and morphologically complex postpositions can only have nominal dependents in the nominative. Compare examples with a morphologically simple postposition *dor* ‘near; edge’ and a morphologically complex postposition *vožla.dor* (front.place+edge) ‘in front of’ derived from it:

- (41) *Lun i voj sulal-əŋi paroxodnij kotol-jas dor-ŋn*
 day and night stand-PRS.3PL steamboat boiler-PL near-LOC
mašīnist-jas.
 captain-PL
 ‘Captains stay day and night near the steamboat boilers.’
 Standard Komi-Zyrian (corpus)

- (42) *Car' vožla.dor-yn taran žugəd-ə kar-lijš*
 tsar in.front.of-LOC battering.ram break-PRS.3SG town-GEN2
sštena-sə.
 wall-P.3.ACC
 ‘In front of the tsar, the battering ram is breaking the walls of the town.’
 Standard Komi-Zyrian (corpus)

As for pronominal dependents, they can appear both in the genitive and in the nominative. Compare:

- (43) *Mij te pricepit-č-i-n me dor-ə?*
 what you.NOM cling-DETR-PST-2SG I.NOM near-ILL
 ‘Why are you soliciting me (lit. clung to me)?’
 Pechora Komi-Zyrian (corpus)

- (44) *Viẓ̌əd-a-m, na-jas mijan dor-ə kut-i-snj lok-ŋi.*
 look-PRS-1PL s/he.NOM-PL we.GEN near-ILL start-PST-3PL come-INF
 ‘We see, they started moving on us.’
 Pechora Komi-Zyrian (corpus)

Nominative marking of pronominal dependents is generally more widespread in Komi, though it highly depends on the specific variety. As an example, our Yazva consultants told us that the main difference between varieties of Yazva spoken in two adjacent villages, Antipina and Vankova, lies in the field of case marking of pronominal dependents in postpositional phrases. People from Vankova use the genitive form of pronoun, whereas Antipina speakers choose the pronoun in nominative:

- (45) *ménam órd-a-m*
 I.NOM at.one’s.place-LOC-P.1SG
 ‘at my place’
 Yazva Komi (Vankova variety, field notes)

- (46) *me órd-a-m*
 I.NOM at.one’s.place-LOC-P.1SG
 ‘at our place’
 Yazva Komi (Antipina variety, field notes)

The general tendency in Komi is to avoid genitive marking of dependents in postpositional phrases, in other words, to make a syntactically simple structure. Additional way of keeping the syntactic structure simple is disallowing complex phrases as dependents. As an example, compare the postpositions with the meaning ‘between’ in the two languages: *kVst-/kVsp-* in Komi and *vis* in Udmurt.

The latter governs the nominative or the instrumental (in its comitative function) and thus can take a phrase containing both locations as a complement:

- (47) *I ot-ân kâk korka visk-ân sâl-e.*
 and that-LOC two house between-LOC stand-PRS.3SG
 ‘And [s/he] stands there between the two houses.’
 Beserman Udmurt (corpus)

- (48) *Giđ-en tuś-en visk-ân so-os.*
 pigsty-INS trough-INS between-LOC s/he-PL
 ‘They (the pigs) are between the pigsty and the trough.’
 Beserman Udmurt (corpus)

A corresponding postposition in Komi, however, only takes dependents in the nominative, and it cannot take a coordinated phrase as a dependent. So, if Komi speakers need to express the relation between two different locations, they have to use two postpositional phrases headed by *kVst-/kVsp-* ‘between’ with simple internal structure rather than one more complex postpositional phrase:

- (49) *úr-is kósp-λn kartá-λs kósp-λn sulál-λλs*
 trough-P.3SG between-LOC pigsty-P.3SG between-LOC stand-PRS.3PL
κλκ porś.
 two pig
 ‘Two pigs stand between the trough and the pigsty.’
 Yazva Komi (corpus)

Complex syntax in Komi is also avoided in groups headed by morphologically complex postpositions with non-spatial meanings. Thus, the postposition *voẓ̌. vjl* ‘before’ takes dependents in the nominative, as spatial inflected postpositions do:

- (50) *Kītənkə Iĺĺa lun-jas voẓ̌.vjl-ijn řit-ja*
 somewhere Elijah day-PL before-LOC evening-ATTR
dor-is nín Ĺena pond-i-s nuəẓ̌-ć-ijnj
 edge-P.3SG already PN begin-PST-3SG invite-DETR-INF
vetl-ijnj gob vot-nj.
 walk-INF mushroom gather.berries-INF
 ‘Before Elijah’s day (a holiday), towards the evening, Lena began to invite us to gather mushrooms.’
 Pechora Komi-Zyrian (corpus)

By contrast, morphologically complex temporal postpositions in Udmurt govern a case traditionally called ablative and labeled “second genitive” in this paper (51–52). This case, which exists in all Permian languages, occupies the periphery of the ablative domain and marks, among others, the stimuli of certain emotion verbs, material, or arguments of verbs such as ‘stop’ or ‘avoid’, while not marking the source of motion in more straightforward cases (which is done by the elative and the egressive instead). Its main syntactic function is marking the possessor of a nominal head in the accusative, which is otherwise marked in the (primary) genitive.

- (51) *Rodña-os-ti-zâ veľt-âl-i-zâ ož ši-isa ju-âsa*
 kin-PL-PROL-P.3PL go-ITER-PST-3PL so eat-CVB drink-CVB
armij-e pot-em-lâš aź.pal wallõ nunaľ-ľos-a-z.
 army-ILL come.out-NMZ-GEN2 before early day-PL-ILL-P.3SG
 ‘One would visit their relatives that way, eat and drink, before going to the
 army, in the past.’
 Beserman Udmurt (corpus)

- (52) *vojna-leš aź.vjľ*
 war-GEN2 before
 ‘before the war’
 Standart Udmurt (corpus)

4 Complex syntax inside postpositional phrases in Udmurt vs simple syntax in Komi: The consequences

We showed in the previous sections that postpositions in Komi can be morphologically complex (attach inflectional markers and/or consist of two stems), but the syntax of postpositional phrases is always simple. In Udmurt, morphological complexity correlates with syntactic complexity, and the internal syntactic structure of postpositional phrases is the first thing to be made complex. These strategies have certain consequences. In Komi, all complex relations expressed in the level of syntax in Udmurt are expressed in one of two possible ways. The first way is agglutination of morphemes inside the word. Thus, the sense ‘(to move) in the direction of a place near a given object’ is expressed by two case

markers in Komi, one for localization (approximative) and the other for orientation (illative):

- (53) *Ša davaj mun-a-m Šibir tuj-lañ-ə, gaškə pet-a-m.*
 EMPH let go-FUT-1PL Siberia road-APP-ILL maybe come.out-FUT-1PL
 ‘Let’s go towards the Siberian road, maybe we will come out [of the forest].’
 Pechora Komi-Zyrian (corpus)

In Udmurt, this sense is expressed by a postpositional phrase headed by *pal* ‘side’ inflected for case. In Beserman Udmurt, a new recessive case has grammaticalized to express the meaning ‘from the side of’ (Usacheva and Arkhangelskiy 2017):

- (54) *...i sält-e mašina dor-a-z, kud*
 ...and stand-PRS.3SG car near-LOC-P.3SG which
pal-ašen pukś-o šofor-jos, so pal-a-z.
 side-EGR sit-PRS.3PL driver-PL that side-LOC-P.3SG
 ‘...and stands near the car, at the side where the drivers sit.’
 Beserman Udmurt (corpus)
- (55) *Kud-lašen kâće mašina lâkt-e, mănâm ton*
 which-RCS which car come-PRS.3SG I.DAT you(SG).NOM
wállö vera-d?
 earlier tell-PST.2SG
 ‘From which side which car returns, you told me before?’
 Beserman Udmurt (corpus)

Morphologization of postpositions in Komi Permyak has reached its terminal stage. Postpositions *vil-* ‘surface’ and *din-* ‘vicinity’ have grammaticalized into suffixes *-l-* and *-d-* marking localizations AD ‘surface’ and APUD ‘neighborhood’, respectively:

- (56) *šojna-l-a-s tećən-y*
 grave-AD-ILL-P.3SG put-PRS.3PL
 ‘they put [it] on the grave’
 Komi Permyak (Batalova 1975: 132)
- (57) *vər-d-əž kotərt-i-sə*
 forest-APUD-TERM run-PST-3PL
 ‘they were running up to the forest’
 Komi Permyak (Batalova 1975: 138)

The process has reached different stages in different dialects. According to the results of a very detailed investigation made by Baker (1985: 175–196), the suffixes function only in Southern Permyak and are not attested in Northern Permyak. Moreover, the suffix *-d-* can be combined only with two of the local cases. Nevertheless, we think that the tendency towards agglutination of postpositions into localization suffixes is quite clear.

As for Udmurt, a new suffix *-ń-* has emerged to mark the DOMUS localization (at X's place), but it is attested only in certain groups of dialects (Lyukina 2008). So, expressing complex senses with a combination of morphemes, i.e. increasing morphological complexity, is typical for Komi, whereas in Udmurt it's usually the syntactic structure that becomes more complex.

5 The Permic split inside postpositional phrases: The reasons and the sources

As we have shown in the previous sections, complex postpositions in Udmurt sometimes allow for nominal dependents in the genitive (to what extent depends on the specific variety), and those in Komi strongly prefer the nominative. Komi and Udmurt are quite closely related, but have existed in different contact situations for quite some time. Therefore, it is possible that the difference has been partly induced by language contact. Theoretically, there are two possible scenarios of such contact-induced change: Russian influence on Komi and Turkic influence on Udmurt. The hypothesis of Russian influence is not supported by the structure of adpositional phrases in the source language. Apart from the fact that Russian does not have postpositions, prepositions in Russian, as a rule, govern cases other than the nominative, including the genitive. It has been also claimed that “postpositions in Finno-Ugric originally governed the nominative, although in the Western group (including Mordvinian) the genitive under government is also very old” (Bereczki 1993: 513). Bereczki considers the cases of the ablative (genitive 2 in our terminology) government of postpositions in Hungarian, Udmurt, Mari and Erzya to result from Turkic influence (Bereczki 1993: 514).

Having dependents in the genitive could also be a trait induced by contact with the neighboring Turkic languages. We have shown that first and second personal pronouns in several dialects of Udmurt strongly prefer genitive marking when used as dependents in postpositional phrases, and the third singular pronoun can also be used in the genitive in this function, although nominative marking is more frequent. In Tatar, the postpositions which arose from relational

nouns demonstrate a very similar split: they invariably attach first, second and third singular personal pronouns in the genitive, whereas other types of pronouns tend to be used in an unmarked form, and all types of nouns remain unmarked (Lyutikova 2017: 40). Genitive marking of dependents in postpositional phrases is also attested in another Turkic language of Volga-Kama region, Chuvash (Ivanova 2011: 68), and in Hill Mari, which has also undergone intensive Turkic influence (Pleshak and Davidyuk 2017). Nevertheless, our preliminary corpus-based research indicates that this possibility, however plausible in theory, is not supported by the data. We were able to find significant differences in dependent marking in postpositional phrases between Udmurt dialects, both for certain individual postpositions and for the system of complex postpositions as a whole. However, it turns out that it is precisely the texts written by people who live or were born far from the Turkic-speaking territories who were more likely to use Turkic-style genitive dependents. On the other hand, the texts written by residents of Tatarstan and Bashkortostan, or Udmurt districts close to the border with these regions, tend to use the nominative for all kinds of dependents. This, together with the fact that complex postpositions in Komi varieties, which have not experienced significant Turkic influence, also sometimes allow pronominal dependents in the genitive, allows us to conclude that this construction in Permic languages is unlikely to be the result of language contact, unlike the genitive 2 government.

6 Conclusion

In this paper, we compared postpositional systems in closely related Permic idioms, Udmurt and Komi. We have shown that most postpositions consist of a stem and a spatial case in both idioms, i.e. they are complex. But there are further differences in the syntax of these constructions, which make certain complex postpositions “more complex” than others. In Komi, newly formed morphologically complex postpositions only allow for dependents in the nominative; these new lexemes also show more noun-like properties than the older inflected postpositions do. In certain varieties of Udmurt, genitive marking of dependents in postpositional phrases (especially characteristic for first and second personal pronouns) is attested. In Udmurt we do not find any simplification of phrases headed by complex postpositions, which is found in Romance and Germanic languages. On the contrary, they participate in more complex syntactic constructions, which involve accepting genitive or “second genitive” dependents in certain cases and in certain dialects, alongside the nominative, which is the default option. This

closely resembles situation found in the Turkic languages spoken in adjacent regions, where having ablative dependents and the genitive/nominative split for complex postpositions is the norm. The fact that some postpositions govern the “second genitive”, or the ablative, case is likely to be the result of contact with Turkic languages. The genitive/nominative split, however, is unlikely to be the result of language contact and must have arisen independently in Permic languages. This development has no visible semantic grounds and probably could be explained by independent developments in the syntax of noun phrases, e.g. syntactic properties of the genitive. Differences in syntax of noun phrases between Permic languages and dialects are therefore probably responsible for the differences in syntactic complexity level of complex postpositions.

Abbreviations

ACC	accusative	ILL	illative
ACT	active	INF	infinitive
AD	adessive (localization ‘very close to, in contact with’)	INS	instrumental
ADD	additive	ITER	iterative
ADJ	adjective	LOC	locative
APP	approximative	NEG	negation
APUD	apudessive (localization ‘next to’)	NMZ	nominalizer
ATTR	attributive	NOM	nominative
CAR	caritive	ORD	ordinal
CAUS	causative	P	possessive
CIT	citative	PL	plural
CVB	converb	PN	proper noun
DAT	dative	PROL	prolative
DEB	debitive	PRS	present
DETR	detransitive	PST	past
DIM	diminutive	PTCP	participle
EGR	egressive	Q	question clitic
EL	elative	RCS	recessive
EMPH	emphatic	REFL	reflexive
EXIST	existential	RES	resultative
FOC	focus	RUS	Russian
FUT	future	SG	singular
GEN	genitive	TERM	terminative

References

- Arkhangelskiy, Timofey, & Ekaterina Georgieva. 2018. Sound-aligned corpus of Udmurt dialectal texts. *Proceedings of the Fourth International Workshop on Computational Linguistics of Uralic Languages*, 26–38. Helsinki, Finland: Association for Computational Linguistics. doi:10.18653/v1/W18-0203. <http://aclweb.org/anthology/W18-0203> (28 September, 2018).
- 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.
- Aurnague, Michel, Kader Boulanouar, Jean-Luc Nespoulous, Andrée Borillo, & Mario Borillo. 2000. Spatial semantics: the processing of Internal Localization Nouns. *Cahiers de Psychologie Cognitive – Current Psychology of Cognition* 19(1). 69–110.
- Baker, Robin. 1985. *The Development of the Komi Case System. A Dialectological Investigation* (Mémoires de La Société Finno-Ougrienne 189). Helsinki: Suomalais-Ugrilainen Seura.
- Batalova, Raisa M. 1975. *Komi-permyackaya dialektologiya* [Komi-Permyak dialectology]. Moscow: Nauka.
- Batalova, Raisa M. 1982. *Arealnye issledovaniya po vostochnym finno-ugorskim yazykam (komi yazyki)* [The areal studies of Eastern Finno-Ugric languages (Komi languages)]. Moscow: Nauka.
- Bereczki, Gábor. 1993. The character and the scale of Turkic influence on the structure of Finno-Ugric languages. In Bela Brogyanyi & Reiner Lipp (eds.), *Comparative-historical Linguistics: Indo-European and Finno-Ugric* (Current Issues in Linguistic Theory 97), 509–520. Amsterdam & Philadelphia: John Benjamins.
- Biryuk, Olga, & Maria Usacheva. 2012. Diskursivnye faktory, vlijajuschie na vybor mezhdru posledozhnoj i posledozhno-padezhnoj formoj v besermjanskom dialekte udmurtskogo jazyka [Discourse factors that influence the choice between postpositional and case-postpositional forms in Beserman Udmurt]. In Ariadna Kuznetsova (ed.), *Finno-ugorskie jazyki: Fragmenty grammaticheskogo opisanija* [Finno-Ugric languages: Fragments of grammatical description], 607–646. Moscow: Jazyki slavjanskix kul'tur.
- DeLancey, Scott. 1997. Grammaticalization and the Gradience of Categories: Relator Nouns and Postpositions in Tibetan and Burmese. In Joan L. Bybee, John Haiman, & Sandra A. Thompson (eds.), *Essays on Language Function and Language Type*, 51–70. Amsterdam: John Benjamins. doi:10.1075/z.82.07del. <https://benjamins.com/catalog/z.82.07del> (28 February, 2020).
- Fejes, László. 2005. *Összetett szavak finnugor nyelvekben* [Compound words in Finno-Ugric languages]. Budapest: Eötvös Loránd Tudományegyetem Dissertation.
- Fraurud, Kari. 2001. Possessives with extensive use: A source of definite articles? In Irène Baron, Michael Herslund, & Finn Sørensen (eds.), *Dimensions of Possession* (Typological Studies in Language 47), 243–268. Amsterdam & Philadelphia: John Benjamins. doi:10.1075/tsl.47.14fra. <https://benjamins.com/catalog/tsl.47.14fra> (28 February, 2020).
- Grenoble, Lenore A. 2014. Spatial semantics, case and relator nouns in Evenki. In Pirkko Suihkonen & Lindsay J. Whaley (eds.), *On Diversity and Complexity of Languages Spoken in Europe and North and Central Asia* (Studies in Language Companion Series 164), 109–132. Amsterdam & Philadelphia: John Benjamins. doi:10.1075/slcs.164.04gre. <https://benjamins.com/catalog/slcs.164.04gre> (28 February, 2020).

- Ivanova, Alyona M. 2011. Morfosintaksis posleloga v chuvashskom yazyke [Morphosyntax of postpositions in Chuvash]. *Vestnik ChGPU im. I. Ya. Yakovleva* 1(69). 65–69.
- Lytkin, Vasilij. 1962. *Komi-Jaz'vinskij dialekt* [Yazva Komi dialect]. Moscow: Izdatel'stvo Akademii nauk SSSR.
- Lyutikina, Nadezhda. 2008. *Osobennosti jazyka balezinskix i jukamenskix besermjan* [The peculiarities of the language of Balezino and Yukamenskoye Besermans]. Izhevsk: Udmurt State University Dissertation.
- Lyutikova, Ekaterina. 2017. Agreement, case and licensing: Evidence from Tatar. *Uralo-Altajskie issledovanija* [Ural-Altai Studies] 25(2). 26–45.
- Maytinskaya, Klara E. 1979. *Istoriko-sopostavitelnaya morfologija finno-ugorskikh jazykov* [Historical and comparative morphology of Finno-Ugric languages]. Moscow: Nauka.
- Nasibullin, Rif Sh. 2000. Udmurtskie dialekty – ideal'nyj object dlya lingvogeograficheskikh issledovanij [Udmurt dialects as an ideal object for investigations of linguistic geography]. In Vilja Oja & Seppo Suhonen (eds.), *Studia ad geographiam linguarum pertinentia* (Eesti Keele Instituudi Toimetised 6). Tallinn: Eesti Keele Sihtasutus.
- Pleshak, Polina, & Tatiana Davidyuk. 2017. Struktura imennoj i poslelozhnoy grupp v gornomarijskom yazyke [NP and PP structure in Hill Mari]. Presented at the Mal'je yazyki v bol'shoj lingvistike, Moscow.
- Schachter, Paul. 1985. Parts-of-speech systems. In Timothy Shopen (ed.), *Language Typology and Syntactic Description*, vol. 1, 3–61. Cambridge, London, New York, New Rochelle, Melbourne & Sydney: Cambridge University Press.
- Serdobolskaya, Natalia, Maria Usacheva, & Timofey Arkhangelskiy. 2018. Grammaticalization of possessive markers in the Beserman dialect of Udmurt. In Lars Johanson, Irina Nevskaya, & L.F. Mazzitelli (eds.), *Linguistic possession. New insights from the languages of Europe and North and Central Asia* (Studies in Language Companion), 291–311. John Benjamins.
- Simonenko, Aleksandra, & Alexey Leontyev. 2012. Morfosintaksis imennogo kompleksa v finno-permskix jazykax: analiz v ramkax gipotezy minimalizma [Morphosyntax of the NP complex in Finno-Permic languages: a minimalist analysis]. In Ariadna Kuznetsova (ed.), *Finno-ugorskie jazyki: Fragmenty grammaticheskogo opisanija* [Finno-Ugric languages: Fragments of grammatical description], 262–340. Moscow: Jazyki slavjanskix kul'tur.
- Usacheva, Maria, & Timofey Arkhangelskiy. 2017. Grammaticalization of New Cases in Beserman Udmurt. *Linguistica Uralica* 53(1). 27–48. doi:10.3176/lu.2017.1.03

Gerd Jendraschek

9 Case marking and complex adpositions in Basque

Abstract: In addition to a paradigm of eleven cases, the Basque benefactive and motivative are halfway between case suffix and complex postposition. One new postposition expressing ‘after’ is the result of degrammaticalization. The omission or optionality of overt adverbial marking leads to the establishment of a set of lexical items with flexible word class membership. Certain adpositions may trigger differential argument marking; this means that their complement can receive different cases, or no case, depending on semantic factors. Differential case marking can correlate with a difference in meaning of the adposition. We further observe differential marking conditioned by the Nominal Hierarchy; for yet another class of adpositions, the determining factor is referential status. While Basque has phrase-level inflection, nouns in certain adpositional constructions have a privileged relation to their head. This stronger bondedness may be labelled “morphological affinity”. Such an analysis only pertains to the morphological boundary between the adposition and its immediate neighbour, but leaves the syntactic constituency unaffected. Adverbial phrases are routinely turned into adnominal modifiers by adding a suffix used to derive complex modifiers. One major insight from the Basque data is that the distinction between a simple and a complex adposition is not very robust. It is a common development for the additional relator that makes an adposition complex to be omitted as the adposition becomes autonomous from its etymon. Such omissibility of the inner and outer relators suggests a continuum of grammaticalization.

Keywords: degrammaticalization, differential marking, grammaticalization, juxtaposition, morphological affinity

Gerd Jendraschek, Gyeongsang National University, Jinju, South Korea. gerd@gnu.ac.kr

<https://doi.org/10.1515/9783110686647-010>

1 Introduction

This chapter¹ looks at complex adpositions in Basque and explores their relations to simple adpositions and case markers. Section 2 gives basic sociolinguistic and typological information on Basque. Section 3 then gives an overview of morphologically complex case markers, as well as simple and complex adpositions and their uses. The etymology of relators, their reanalysis, including a case of degrammaticalization, deverbal adpositional expressions, and structural attrition are presented in the section on diachronic variation (section 4). The section on synchronic variation presents a corpus study confirming that the pluricentricity of Basque is visible in the field of adpositions (section 5). Section 6 explores topics particularly relevant to Basque, such as differential case marking, the adnominal use of adpositional phrases, and contains a brief note on the distinction between location nouns and common nouns in genitive constructions. The major insights are summarized in section 7.

2 The Basque language

Basque is a language isolate spoken natively by about 489,000 people in Spain and 54,000 people in France. Adding a net gain (L2-acquisition minus L1-attrition) of 209,000 second-language speakers, the total number of Basque speakers is approximately 750,000 (not taking into account the diaspora outside of the Basque Country; data are from Basque Government et al. 2016). The language is surrounded by and sociolinguistically overlapping with Romance languages. Depending on time and place these are Castilian Spanish, French, and the Gascon variety of Occitan. Before the standardization of the 20th century, Basque was a collection of six major dialects (plus two or three minor ones, cf. Trask 1997: 5–6), and the little writing that was done in Basque would be in a literary form of the local dialect. The extent of diatopic variation within Basque was comparable to that between the contiguous Romance languages of Western Europe, so that the easternmost dialect Zuberoan was roughly as different from the westernmost dialect Bizkaian as Italian is from Portuguese. This gives us an idea of the challenges during the creation of a standardized form of Basque called *euskara batua* ‘unified Basque’. The standardization efforts of the 20th century have considerably reduced, but not eliminated diatopic variation, the result being a pluricentric

¹ I would like to thank Michel Aurnague for his many helpful comments on a previous version of this chapter.

language whose subdivisions now mirror political, rather than former dialectal borders (cf. Jendraschek 2007: 110).

Despite a long tradition of extensive calquing (cf. Jendraschek 2008), the morphology and syntax of Basque remain quite different from Romance. Relative clauses are typically preposed to their head, auxiliaries and modals follow the lexical verb. Unlike Romance, Basque has no nominal gender, not even in personal pronouns. It does however have gender agreement in allocutive verb forms, i.e. indexing of the gender of the addressee where the latter is not an argument of the predicate. As this typological rarity is restricted to intimate addressees, we will not be concerned with it here. For the purposes of this volume, the overwhelmingly head-final nature of intra-clausal syntax, the widespread use of postpositional nouns, and the elaborate inventory of case marking suffixes are particularly relevant. Example (1) illustrates the allative case suffix *-ra* corresponding to a simple preposition in English or French.

- (1) *Ume-a-k ikastola-ra Ø-d-oa-z.*
 child-DET-PL school-ALL S.3-PRS-go-S.PL
 ‘The children go to school.’

(Zubiri 1994: 114)

Case marking applies to the level of the phrase. This means that the case suffix appears only once, on the final word of the noun phrase, which can be a noun, an attributive adjective, or a determiner. This does not mean that it is a clitic: case marking follows number marking (which is also phrase-final), with some fusion between the two. Example (2) illustrates the fusional nature of nominal inflection.

- (2) *Donostia-ko hiru hondartza eder hau-e-ta-n*
 San.Sebastian-MR three beach beautiful D1-PL-LOC-LOC
 ‘on these three beautiful beaches of San Sebastian’

(adapted from Jendraschek 2007: 128)

The glossing as a “double locative” following a plural morpheme is based on an etymologizing segmentation, and originated as a form of plurality marking restricted to locative forms. Although largely relying on reconstruction, it is insightful for a deeper understanding of Basque morphology, and adopted here for consistency. From a synchronic point of view, however, *-etan* could be analyzed as a single morpheme to be glossed as LOC.PL (as a cumulative morpheme) or PL:LOC (as a fused morpheme sequence).

Following Trask (1997: 92–96), there are nine cases marked by morphologically simple suffixes, including an (unmarked) absolutive, an ergative, and a dative. These three cases correspond to cross-reference markers on the finite predicate. These person-number indices agree with or represent the corresponding noun phrases. Example (3) is therefore a complete sentence with or without the noun phrases in square brackets.

- (3) *[ni-k zu-ri liburu-a-k] ema-ten Ø-d-i-z-ki-zu-t.*
 1SG-ERG 2-DAT book-DET(ABS)-PL give-1PFV O.3-PRS-IO-O.PL-IO-2-A.1SG
 ‘I give them [the books] to you.’

(Jendraschek 2012: 8)

The remaining cases are usually not governed, but typically mark adjuncts (or in the case of the genitive, attributes) modifying another part of the clause. The paradigm of morphologically simple case markers includes three local cases for locative (position at or in a landmark), ablative (movement away from a landmark), and allative (movement to a landmark). Two more local cases are diachronically complex, as they contain the allative marker followed by a second suffix to form a directional (‘towards’), and a terminative (‘as far as, up to’); it has been suggested that these markers are the result of fusion between the allative and an erstwhile postposition (Hualde 2003a: 186). Table 1 illustrates these eleven cases on the root *mendi* ‘mountain’. The nominal forms are shown with the default determiner *-a*, which surfaces before most case suffixes and is also used in citation forms. While traditionally labelled “definite article”, this suffix functions as an individualizer (cf. Haase 1992: 54).

Table 1: Basque case paradigm.

case type		label	form (singular, definite)	
governed cases		absolutive	<i>mendia</i>	
		ergative	<i>mendiak</i>	
		dative	<i>mendiari</i>	
modifying cases	non-local	instrumental	<i>mendiaz</i>	
		comitative	<i>mendiarekin</i>	
		genitive	<i>mendiaren</i>	
	local	simple	locative	<i>mendian</i>
			ablative	<i>menditik</i>
			allative	<i>mendirara</i>
		complex	directional	<i>mendirantz</i>
			terminative	<i>mendiraino</i>

Note that Basque also has a partitive suffix *-(r)ik* in paradigmatic opposition with the absolutive. Most of the time, it functions like an indefinite negative determiner, similar to English *any* in *ez daukagu dirurik* ‘we don’t have **any** money’.

3 Complex relators in Basque

In this section, I describe in detail the system of relators beyond the core system of cases, beginning with compound cases, i.e. case markers that are themselves morphologically complex. These include markers whose second element has sometimes been analyzed as a postposition rather than a case suffix. After surveying simple postpositions and their governed cases, I will then introduce the main adpositional construction consisting of a relational noun governing a genitive-marked nominal complement. Finally, we will see complex postpositional constructions governing other cases. Note that Basque has no prepositions or circumpositions. As Trask (1998: 320) emphasizes, even the Romance preposition *contra* ‘against’ appears in Basque as a postposition, as in *Jonen kontra* ‘against John’, likely as a result of transiting through a stage as a borrowed noun; cf. Hualde (2002: 330).

3.1 Complex case markers

In addition to the eleven cases listed in Table 1, we find markers that are halfway between case and complex postposition. Let us begin with the benefactive *-entzat*, included among the “compound cases” in Trask (1997: 93), as it consists of the genitive ending followed by *-tzat*. We find *-tzat*, this time without determiners or the genitive case, also used as a “prolative” or “essive/translative” (Trask 1997: 94). The benefactive and the prolative are illustrated in (4) vs. (5), respectively.

- (4) *Pastel hori ume-a-**rentzat** Ø-d-Ø-a.*
 cake D2.SG(ABS) child-DET-BEN S.3-PRS-S.SG-IND.INTR
 ‘That cake is for the child.’

(Zubiri 1994: 96)

- (5) *italiano-**tzat** har-tu n-in-du-te-n*
 Italian-PROL take-PFV O.1SG-PST-IND.TR-A.3PL-PST
 ‘They took me for an Italian.’

(Hualde 2003a: 185)

While the English preposition *for* and its French counterpart *pour* cover both functions, Spanish has differentiated them into *para* vs. *por*. It is intriguing that the Spanish benefactive preposition *para* goes back to a “compound” preposition *por a* (Penny 2002: 243), mirroring the compound benefactive case in Basque. While an etymological link between the benefactive and the prolativ is likely within Basque, the question whether the Spanish-Basque parallelism is the result of pattern replication goes beyond the scope of this chapter. Basque differs from Spanish however in that it distinguishes between a benefactive marker attaching to animate hosts on the one hand, and a destinative case marker attaching to inanimate entities, typically places, but also time expressions, on the other. Both uses are rendered by *para* in Spanish.

- (6) a. *Ispilu hori komun-erako Ø-d-Ø-a.*
 mirror D2.SG(ABS) toilet-DEST S.3-PRS-S.SG-IND.INTR
 ‘That mirror is for the toilet.’
 (Zubiri 1994: 130)
- b. *Ispilu hori komun-eko eros-i Ø-d-u-t.*
 mirror D2.SG(ABS) toilet-DEST buy-PFV 0.3-PRS-IND.TR-A.1SG
 ‘I bought that mirror for the toilet.’
 (Michel Aurnague p.c.)

This destinative marker *-(e)rako* is formed by adding *-ko* (in some northern varieties *-kotez*) to the allative case suffix *-(e)ra*, as in (6a). In the North, we also find constructions like (6b), where destinative marking does not contain an allative morpheme.

A better candidate for the scenario of a complex adposition grammaticalizing into a case marker is *-(r)engatik* ‘because of’, which is reconstructed as *-(r)en gai-tik* GEN matter-ABL ‘from the matter of’ (Hualde 2002: 332; Hualde 2003b: 188). Synchronically, it is sometimes labelled ‘motivative case’ (e.g. in Zubiri and Zubiri 2000: 394) or ‘motivational’ (de Rijk 2008: 33). Others consider it an invariable postposition governing a complement in the genitive case (Trask 1997: 94; Hualde 2003b: 188).

Although already included among the simple case markers in Table 1, the comitative also has its diachronic origin in a complex phrase. While the case suffix *-rekin* ‘with’ is considered synchronically monomorphemic, it is a contraction of *-re kide-(a)n* GEN fellow-LOC ‘in the company of’. Accordingly, a phrase such as *gizonare(n) kide(a)n* ‘in the man’s company’ would have come to be simplified, both structurally and semantically, to *gizonarekin* ‘with the man’ (Trask 1997: 201–202). Even more speculative is Trask’s (1997: 203) reconstructed origin of the locative case suffix *-(e)an* from the noun *ga(i)n* ‘top’, as in *etxe-an* house-LOC

‘in the house’ evolving from the noun sequence **etxe gan* house top ‘(on) top of the house’. Such an evolution would exemplify the grammaticalization cline noun > adposition > case (cf. Lehmann 2015: 81–92). For an alternative reconstruction that traces the locative ending *-ean* to the demonstrative *han* (<**ha-n* D3-LOC) preceded by epenthetic *-e-* to break up consonant clusters where necessary (**lan-e-ha-n* work-Ø-D3-LOC ‘in that work’ becoming present-day *lan-ean* work-LOC ‘at work’), see Manterola (2015: 255).

3.2 Simple adpositions

Table 2 includes only simple, invariable, postpositions; for complex postpositions see section 3.3. Also not included are lexemes commonly classified as adverbs such as *gertu/hurbil* ‘close (to)’ or *urruti* ‘far (from)’; these items are inflectable (*gertuago* ‘closer’, *urrutien* ‘farthest’) and are often used independently (*urruti dago* ‘it’s far’). A preceding nominal is marked with ablative case, as in *Euskal Herritik urruti* ‘far away from the Basque Country’. A similar case is the temporal construction *baino lehen* ‘before’, as in *astelehena baino lehen* ‘before Monday’, made up of *baino* ‘than’ and *lehen* ‘first’ (cf. Trask 1997: 208).

Table 2: Basque simple adpositions.

case governed	form	meaning	Basque example	example translation
∅	<i>arte</i>	‘until’	<i>goiz arte</i>	‘until the morning’
∅	<i>barik</i>	‘without’	<i>betaurreko barik</i>	‘without glasses’
∅	<i>bezala</i>	‘like’	<i>beti bezala</i>	‘as usual’
∅	<i>eske</i>	‘asking for’	<i>diru eske</i>	‘asking for money’
∅	<i>gisa</i>	‘as’	<i>laburpen gisa</i>	‘as a summary’
∅	<i>inguru</i>	‘approximately’	<i>duela bost urte inguru</i>	‘about five years ago’
∅	<i>ondoren</i>	‘after’	<i>gerra ondoren</i>	‘after the war’
PRTV	<i>gabe</i>	‘without’	<i>dirurik gabe</i>	‘without money’
GEN	<i>alde</i>	‘on somebody’s side’	<i>zure alde</i>	‘on your side’
GEN	<i>arabera</i>	‘according to’	<i>legearen arabera</i>	‘according to the law’
GEN	<i>begira</i>	‘waiting for’	<i>erantzunaren begira</i>	‘waiting for an answer’
GEN	<i>bila</i>	‘looking for’	<i>lagun baten bila</i>	‘looking for a friend’

Table 2 (continued)

case governed	form	meaning	Basque example	example translation
GEN	<i>bizkar</i>	'at the expense of'	<i>senitartekoen bizkar</i>	'at the expense of relatives'
GEN	<i>esku</i>	'in somebody's hands'	<i>zure esku</i>	'in your hands' (i.e. you are responsible)
GEN	<i>kontra/aurka</i>	'against'	<i>hormaren kontra</i>	'against the wall'
GEN	<i>pare</i>	'like'	<i>tximistaren pare</i>	'like a flash of lightning'
GEN	<i>truke</i>	'(in exchange) for'	<i>auto zaharraren truke</i>	'in exchange for the old car'
GEN	<i>zain</i>	'waiting for'	<i>autobusaren zain</i>	'waiting for the bus'
DAT	<i>begira</i>	'looking at'	<i>zuri begira</i>	'looking at you'
DAT	<i>esker</i>	'thanks to'	<i>lagun bati esker</i>	'thanks to a friend'
LOC	<i>gaindi</i>	'passing through'	<i>Tolosan gaindi</i>	'passing through Toulouse'
LOC	<i>zehar</i>	'through'	<i>basamortuan zehar</i>	'through the desert'
ABL	<i>at</i>	'outside'	<i>herritik at</i>	'outside of the village'
ABL	<i>kanpo</i>	'outside'	<i>etxetik kanpo</i>	'outside of the house'
ABL	<i>landa</i>	'outside'	<i>hiritik landa</i>	'outside the city'
ALL	<i>begira</i>	'looking at (space or time)'	<i>zerura begira</i>	'looking up at the sky'
INS	<i>gain</i>	'in addition to'	<i>zutaz gain</i>	'in addition to you'
INS	<i>gero</i>	'after'	<i>barbaroen inbasioaz gero</i>	'since the invasion of the barbarians'
INS	<i>landa</i>	'apart from'	<i>hiriaz landa</i>	'apart from the city'

Lack of case-marking, i.e. juxtaposition, indicates advanced grammaticalization; see section 7. The genitive is the default marking for the complement of a relational noun, which explains its predominance in Table 1. While the dative and the local cases have a semantic origin, the function of the instrumental is more abstract. Certain adpositions may trigger differential argument marking: this means that their complement can receive different cases, or no case, depending on semantic factors. These topics will be discussed in more detail in section 6.1.

Note that while the string of a case marker and a postposition is technically an instance of a complex relator, the adpositions in Table 2 are all simple in the same way the German preposition *für* 'for' is simple despite governing a complement in accusative case: agglutinative vs. fusional case marking on the complement should have no bearing on the synchronic analysis of the adposition.

3.3 Complex adpositions

Complex postpositions are typically made up of nouns followed by a case marker. That case is typically, but not always, one of the three simple local cases locative, ablative, or allative. Complex adpositions may or may not trigger case-marking on their complement. Let us begin our tour with a pair of adpositions whose complexity is merely diachronic: *gora* ‘up’ and *behera* ‘down’ are listed in dictionaries as (lexicalized) postpositions (or adverbs, if used without a complement), but they go back to the nouns *goi* ‘top, high place’ and *behe* ‘bottom’, followed by allative case. We see an example in (7), where the complement of *gora* is marked with locative case.

- (7) *Mendi-an go-ra joan Ø-d-ir-a.*
 mountain-LOC top-ALL go S.3-PRS-S.PL-IND.INTR
 ‘They have gone up the mountain.’

(Zubiri 1994: 189)

In contrast, the canonical postpositional phrase in the sense of this volume has a straightforwardly transparent internal structure. The construction can be schematized as [N-GEN [Nrel-LOC]], where Nrel stands for a relational noun such as *aurre* ‘front’ or *barru* ‘inside’, followed by one of the local cases. This construction is the mirror image of the English *in front of*, where English uses the preposition *of* as the functional equivalent of the genitive, and another preposition *in* corresponding to the locative.

- (8) *Iñaki zu-re aurre-an Ø-da-go-Ø.*
 Iñaki 2(SG)-GEN front-LOC S.3-PRS-be-S.SG
 ‘Iñaki is in front of you.’

(Zubiri 1994: 180)

The genitive case marker is a priori not part of the postposition. The relational noun *aurre* ‘front’ governs a possessor complement (whose front?) just like a transitive verb governs a direct object (cf. Lehmann 1983: 373). The genitive case affix on the adpositional complement is then the equivalent of an accusative case suffix on a verbal complement; case alternations within the adpositional phrase are comparable to different valency frames of a verb. However, while the case affix on the complement is a priori governed by the adposition, regardless of the latter’s being simple or complex, the linear chunk of case and adposition (or whichever morphemes constitute the construction) may very well coalesce to something more compact. We will see some extreme instances of this happening in Basque in section 4.2. More trivial illustrations of the disappearance of the syntactic boundary

between an adposition and its complement nominal are French examples such as *au* (<*à* ‘to’+*le* ‘DEF.SG.M’) or *des* (*de* ‘of’+*les* ‘DEF.PL’), or German examples such as *zur* (<*zu der* ‘to DEF.SG.F.DAT’) and *am* (<*an dem* ‘at DEF.SG.M/N.DAT’), where prepositions fuse with (in the case of German, case-marked) articles.

In addition to the case marking on the complement, the case marking on the complex adposition itself may vary. Thereby, the same relational noun can yield several complex adpositions: compare the a. and b. versions in (9) and (10).

- (9) a. *gela txiki-a-ren barren-ean*
 room small-DET-GEN inside-LOC
 ‘inside the small room’
 b. *mundu-an barren-a*
 world-LOC inside-ALL
 ‘throughout the world’

(de Rijk 2008: 75)

- (10) a. *mahai-a-ren gain-ean*
 table-DET-GEN top-LOC
 ‘on the table’
 b. *mahai-a-z gain-era*
 table-DET-INS top-ALL
 ‘besides the table’

While *-ren barrenean* ‘inside’ and *-ren gainean* ‘on (top of)’ follow the transparent canonical construction, the locative and allative of the expression *-n barrena* ‘throughout’, as well as the instrumental and allative of *-z gainera* ‘besides, in addition to’, are morphological fossils without synchronic transparency. This difference in transparency has no obvious bearing on the frequency of the constructions: The phrase *munduan barrena* ‘throughout the world’ comes up 33000 times in a Google search.

Complex postpositions are not necessarily derived from nouns. In *ezean* ‘for lack of’, we have the general negator *ez*, which transcends word classes and can be used as the pro-clause ‘no’, the predicate negator ‘not’, or a noun meaning ‘lack, absence’. Because of its negative polarity, the complement is marked with the partitive.

- (11) *Ur-ik ez-ean, esne-a edan-go Ø-d-u-t.*
 water-PTV NEG-LOC milk-DET drink-PROSP 0.3-PRS-IND.TR-A.1SG
 ‘For lack of water, I am going to drink milk.’

(Zubiri and Zubiri 2000: 242)

As for recent additions to the inventory of complex adpositions, we may mention here *mailan* ‘at the level of’, ‘in terms of’, and *arloan* ‘in the field of’, derived respectively from *maila* ‘level’ and *arlo* ‘field’ via addition of the locative case suffix. These emergent postpositions typically do not trigger case marking if following a single noun, effectively forming a compound with it, as in *ekonomia mailan* ‘at the level of the economy, at the economic level’ (cf. Zubiri and Zubiri 2000: 253). The use as adposition entails the rebracketing of *ekonomia mailan* from (12a) to (12b) (cf. Lehmann 2015: 83). This reanalysis shifts the status of *maila* from head of the (case-marked) noun phrase to head of the adpositional phrase.

- (12) a. $[[ekonomia\ maila]_{N-n}]_{AdvP}$
 b. $[[ekonomia]_N [maila-n]_{Adp}]_{AdpP(=AdvP)}$

While *maila* and *arlo* survive as regular nouns that can be used in all kinds of constructions, one indication of the reanalysis suggested in (12) is the high frequency of certain collocations with the locative suffix in comparison to their counterparts with absolutive case-marking. A search on the site of the Basque-language newspaper *berria* turned up 174 examples of *politika arloan* ‘in the field of politics’ compared to only 2 of *politika arloa*, as well as 700 hits for *mundu mailan* ‘on the global level’ as compared to 4 for *mundu maila*.

Table 3 lists complex postpositions with spatial semantics that follow the canonical pattern. These are singled out because of their productive and predictable structure and their central role in this functional domain; cf. also Aurnague (1996b) for an in-depth semantic description of location nouns.

Table 3: Basque canonical complex spatial adpositions.

stem form	meaning	Basque example	example translation
<i>albo/saihets</i>	‘side’	<i>nire alboan</i>	‘beside me’
<i>aldamen</i>	‘vicinity’	<i>etxearen aldamenean</i>	‘near the house’
<i>alde</i>	‘side’	<i>oherearen aldean</i>	‘next to the bed’
<i>arte</i>	‘gap’	<i>liburuen artean</i>	‘between the books’
<i>atze/gibel/oste</i>	‘backside’	<i>eliza handiaren atzean</i>	‘behind the big church’
<i>aurre/ai(n)tzin</i>	‘front(side)’	<i>etxe zuriaren aurrean</i>	‘in front of the white house’
<i>azal</i>	‘surface’	<i>lurraren azalean</i>	‘on the surface of the earth’
<i>azpi/pe</i>	‘lower part’	<i>eguzkiaren azpian</i>	‘under the sun’
<i>barren/barne/barru</i>	‘inside’	<i>gela txikiaren barrenean</i>	‘inside the small room’
<i>bazter</i>	‘corner, edge’	<i>bidearen bazterrean</i>	‘at the edge of the road’

Table 3 (continued)

stem form	meaning	Basque example	example translation
<i>behe(re)</i>	'lower part'	<i>Larrun mendiaren behean</i>	'at the foothill of the Rhune'
<i>buru</i>	'head'	<i>bidearen buruan</i>	'at the end of the road'
<i>erdi</i>	'middle'	<i>hiriaren erdian</i>	'in the middle of the city'
<i>ertz</i>	'edge, margin'	<i>gona gorriaren ertzean</i>	'on the hem of the red skirt'
<i>eskuin</i>	'right side'	<i>Jainkoaren eskuinean</i>	'at the right hand of God'
<i>ezker</i>	'left side'	<i>nire ezkerrean</i>	'on my left'
<i>gain</i>	'upper part'	<i>hiriaren gainean</i>	'above the city'
<i>hegi</i>	'edge'	<i>errekaren hegian</i>	'on the shore of the creek'
<i>inguru</i>	'circumference; vicinity'	<i>gatzeluaren inguruan</i>	'around the castle'
<i>izkin/zoko/kantoi(n)</i>	'angle, corner'	<i>etxearen izkinean</i>	'at the corner of the house'
<i>mutur</i>	'snout, muzzle'	<i>sokaren muturra</i>	'the end of the rope'
<i>ondo</i>	'vicinity'	<i>amaren ondoan</i>	'at mother's side'
<i>pare</i>	'pair'	<i>zure parean</i>	'opposite you'
<i>punta/moko</i>	'extremity'	<i>lantzaren mokoan</i>	'at the end of the lance'
<i>zola</i>	'bottom'	<i>putzuaren zolan</i>	'at the bottom of the well'

These spatial postpositions are also homogeneous in their internal structure, which is [N-GEN [Nrel-LOC]]; for the option to omit the genitive suffix on inanimate nouns, see section 6.1. Note that the local case on the postpositional noun can vary depending on the relationship to the predicate, as illustrated in (13), from de Rijk (2008: 71).

- (13) a. *mendi atze-an*
 mountain behind-LOC
 'behind the mountain'
- b. *mendi atze-tik*
 mountain behind-ABL
 'from behind the mountain'
- c. *mendi atze-ra*
 mountain behind-ALL
 '(to) behind the mountain'

Table 4 gives a list of complex postpositions following different patterns. These complex adpositions are labelled idiosyncratic for the lower predictability of the case markers on both the complement and the head.

Table 4: Basque idiosyncratic complex adpositions.

form	gloss	meaning	case governed	Basque example	example translation
<i>antzean</i>	resemblance-LOC	'in the manner of'	∅	<i>handiki antzean</i>	'like a big shot'
<i>antzera</i>	resemblance-ALL	'in the manner of'	GEN	<i>zure antzera</i>	'(in) your way'
<i>barrena</i>	inside-ALL	'along, through(out)'	LOC	<i>kalean barrena</i>	'along the street'
<i>bestalde(an)</i>	other-side(-LOC)	'on the other side'	INS	<i>mugaz bestaldean</i>	'on the other side of the frontier'
<i>bidez</i>	path-INS	'by means of'	GEN	<i>sistema honen bidez</i>	'by using this system'
<i>bitartez</i>	mediator-INS	'by means of'	GEN	<i>lagun baten bitartez</i>	'through a friend'
<i>buruz</i>	head-INS	'about'	DAT	<i>Joni buruz</i>	'about John'
<i>ezean</i>	NEG-LOC	'in the absence of'	PRTV	<i>urik ezean</i>	'in the absence of water'
<i>gainera</i>	top-ALL	'besides'	INS	<i>diruaz gainera</i>	'in addition to the money'
<i>geroztik</i>	after-INS-ABL	'since'	INS	<i>herenegunez geroztik</i>	'since the day before yesterday'
<i>gisara</i>	manner-ALL	'in the manner of'	GEN	<i>neure gisara</i>	'(in) my way'
<i>kontura</i>	count-ALL	'at the expense of'	GEN	<i>zure kontura</i>	'at your expense'
<i>lepotik</i>	neck-ABL	'at the expense of'	GEN	<i>gurasoen lepotik</i>	'at the parents' expense'
<i>menpean</i>	authority-LOC	'under the authority of'	GEN	<i>zuzendariaren menpean</i>	'under the director's authority'
<i>ordez</i>	place-INS	'instead of'	GEN	<i>aitaren ordez</i>	'instead of the father'
<i>partez</i>	part-INS	'on behalf of'	GEN	<i>nire partez</i>	'on my behalf'

Etymologically, these forms are also nouns followed by a case-marker; *gero* 'later (time)' can be used both as an adverb and a noun. Deverbal adpositions will be discussed in section 4.3.

Table 5 gives an overview of the patterns found in selected semantic domains. The emerging motivative case marker *gatik* ‘because of’ is included among the simple adpositions; *ondoren* ‘after’ seems to be etymologically derived from *ondo* ‘rear; bottom, side’ (cf. Aurnague 1996b: 28).

Table 5: Form-function mapping.

domain	meaning	Basque	simple adpos.	compl. adpos.	adverbial	deverbal
space	in front of	<i>aurrean</i>		x		
	behind	<i>atzean</i>		x		
	near	<i>hurbil</i>				x
	far	<i>urruti</i>				x
	through	<i>zehar</i>	x			
	among	<i>artean</i>			x	
	along	<i>barren</i>			x	
time	before	<i>baino lehen</i>				x
	after	<i>ondoren</i>	x			
	during	<i>bitartean</i>		x		
cause	in view of	<i>ikusirik</i>				x
	considering	<i>kontuan hartuz</i>				x
	because of	<i>gatik</i>	x			
	due to	<i>dela eta</i>				x
concession	instead of	<i>ordez</i>			x	
	despite, although	<i>arren</i>	x			
topic	concerning	<i>dagokionez</i>				x
	about	<i>buruz</i>			x	
purpose	with a view to	<i>asmoz</i>			x	

3.4 Adpositions vs conjunctions

The difference between an adposition and a subordinating conjunction is that the latter governs a verb. Adposition and semantically equivalent conjunction may be homonymous (compare *after your departure* with *after you left*) or suppletive (*despite the rain* vs. *although it rains*); see Jendraschek (2016). The distinction between adpositions and conjunctions is neutralized when verbal forms adopt nominal properties, and a few examples of this happening in Basque shall be presented here. A clear case of deverbal nominalization is *ikusteagatik* ‘because

of seeing’, where *gatic* ‘because of’ is attached to the verbal noun *ikustea* ‘the seeing’, derived from the stem *ikus* ‘see’. The adposition *landa* ‘outside; apart from’ governs a verbal noun in instrumental case: *izateaz landa* ‘besides being’. Examples of adpositions combining with the perfective participle are *ikusi ondoren* ‘after seeing’, *heldu aitzin* ‘before arriving’, *gauzak konpondu ordez* ‘instead of fixing things’, and *ezer esan gabe* ‘without saying anything’. All this is not too different from what the English gerund does.

Things get interesting, and the distinction between adposition and conjunction becomes relevant again, where the meaning changes depending on the syntactic class of the complement. This is what happens with *gero*, whose lexical meaning is ‘later (time)’. After a noun in instrumental case, *gero* expresses ‘after’ (14a), yet it marks a conditional nexus when attached to a participle (14b).

- (14) a. *barbaro-en* *inbasio-a-z* ***gero***
 Barbarian-GEN(PL) invasion-DET-INS later
 ‘after the invasion of the Barbarians’
- b. *Bide* *horr-etan* *abia-tu-z* ***gero***,
 path D2-LOC head-PFV-INS later
laster *aurki-tu-ko* *Ø-d-Ø-u-gu*.
 soon find-PFV-PROSP 0.3-PRS-O.SG-IND.TR-A.1PL
 ‘If we take that road, we’ll find it soon.’

(Elhuyar s. v. *-z gero*)

This is just one instance of the astonishingly versatile role of *gero*, whose tale will be continued in section 4. Certain adpositions may occur without a complement, in which case they function as adverbials. This is particularly common with location nouns, e.g. *segi aurrera* ‘go straight’, where the relational noun *aurre* ‘front’ is followed by allative case. Halfway between adpositional use and independent use is *-tik aurrera* ‘from ... forward’, as in *bihartik aurrera* ‘from tomorrow’, with ablative marking on *bihar* ‘tomorrow’; see section 6.2 for a discussion of the very similar *kanpora* ‘outwards’.

4 Diachronic variation

4.1 Etymologies

Basque adpositions can be divided into those with Latin etymologies on the one hand, and those that underwent semantic change within Basque on the other.

Examples of both classes are listed in Tables 6 and 7, respectively, together with their suggested etymologies, adapted from sources such as Trask (1997; 2008) and de Rijk (2008). Note that Latin forms may be given in a non-classical form reflecting the most likely morphological and phonological form at the time of borrowing. There are more recent borrowings from Romance that do not go as far back as Latin, e.g. *truke* ‘(in) exchange (for)’, whose origin is Spanish *trueque* ‘barter’, ‘exchange’. The etymology of *ondo* ‘side’, ‘bottom’, ‘rear’ appears to be a case of phono-semantic conflation of a Latin etymon with a pre-existing Basque noun (*h*)*ondo* ‘foot’ (Aurnague 1996b: 27–29).

Table 6: Basque adpositions of Latin origin.

form	meaning	origin	meaning in Latin
<i>gertu</i>	‘near’	CERTU	‘fixed, firm’
<i>inguru</i>	‘vicinity’	INGYRU	‘around, in a circle’
<i>kanpo</i>	‘outside’	CAMPO	‘field’
<i>ondo</i>	‘bottom, side’	FUNDU	‘bottom’

Table 7: Basque adpositions with Basque etymology.

form	meaning	etymon	meaning of etymon	notes
<i>aitzin</i>	‘frontside’	<i>aitzi</i>	‘against’	<i>-n</i> is a fused locative
<i>at</i>	‘except’	<i>ate</i>	‘door’; ‘outside’	
<i>atze</i>	‘back’	(<i>h</i>) <i>atz</i>	‘trace, track’; ‘paw’	<i>-e</i> through reanalysis from locative <i>-ean</i>
<i>aurka</i>	‘against’	<i>aurre</i>	‘front’	see entry for <i>aurre</i> ; <i>-ka</i> is an adverbializer
<i>aurre</i>	‘front’	<i>aur~ahur</i>	‘palm of the hand’; ‘face’	
<i>barrena</i>	‘along’	<i>barru</i>	‘inside’	<i>-en</i> probably fused superlative; <i>-a</i> allative
<i>bizkar</i>	‘at the expense of’	<i>bizkar</i>	‘(a person’s) back’	ablative optional
<i>buruz</i>	‘about’	<i>buru</i>	‘head’	instrumental <i>-z</i>
<i>gaindi</i>	‘passing through’	<i>gain</i>	‘top’	<i>-di</i> may be an old ablative or directional
<i>gibel</i>	‘behind’	<i>gibel</i>	‘liver’	putative <i>gi</i> ‘meat’ and <i>bel</i> ‘dark’

Noteworthy is the form *aitzinean* ‘in front’, which can be reconstructed as a reinforced locative. If the etymon *aitzi*, still used in the sense of ‘against’, is correct, then the *-n* in *aitzin* is the locative case. The form *aitzinean* with a second locative

suffix *-ean* would then be the consequence of the fusion of the erstwhile locative suffix *-n* with its host *aitzi*, leading to the reanalysis of *aitzin* as a single morpheme (cf. de Rijk 2008: 73).

4.2 Chunking

The term “chunking” describes the mechanism whereby a syntagmatic string of structural units (i.e. morphemes and words) comes to be stored as a more complex set expression. It explains the emergence of multi-word expressions such as proverbs, idiomatic phrases, lexicalized compounds, complex predicates, and complex adpositions. Chunking is an example of conventionalization (cf. Bybee 2010: 34–37).

An interesting case is the relator *eta gero* ‘after’, made up of the conjunction *eta* ‘and’ followed by *gero* ‘later (time)’. Since *eta gero* can follow both nouns (15a) and verbs (15b), it is a complex adposition which in combination with the perfective participle can be used as a subordinator.

- (15) a. *Kantaldi-a eta gero, afal-tze-ra joan g-in-Ø-en.*
 concert-DET and later dine-NR-ALL go S.1PL-PST-IND.INTR-PST
 ‘We went to have dinner after the concert.’
- b. *Kantaldi-a amai-tu eta gero,*
 concert-DET end-PFV and later
afal-tze-ra joan g-in-Ø-en.
 dine-NR-ALL go S.1PL-PST-IND.INTR-PST
 ‘We went to have dinner after the concert was over.’ (Artiagoitia 2003: 748)

This is surely a case of reanalysis as schematized in Figure 1.

Stage 1: [NP] [*eta gero* [S]] ‘NP, and then S’

Stage 2: [[NP] *eta gero*] [S] ‘After NP, S’

Figure 1: Reanalysis from coordinator to adposition.

In stage 1, the initial conjoint, which can be a non-finite clause or just a noun phrase, is followed by the second conjoint introduced by a conjunction and, typically but not necessarily, a temporal adverbial. Any time expression will do here, but *gero* ‘later’ is the default (see de Rijk 2008: 536). In stage 2, the connecting *eta gero* is no longer construed as being preposed to the second conjoint but as being postposed to the first. Prosodically too, it has come to be attached

to the first element. The conjunction has thereby been reanalysed as a complex adposition.

This *gero* also shows up in a chunk involving double case marking. In (16), *gero* is first marked with instrumental case, followed by the ablative case suffix.

- (16) *herenegun-ez* ***gero-z-tik***
 day.before.yesterday-INS later-INS-ABL
 ‘since two days ago’ (Elhuyar s. v. *-z/-e)nez geroztik*)

Simpler *geroz* would already express the meaning of ‘thenceforth’; the additional presence of the ablative can be explained by postulating an omitted goal adverbial, as in the phrase *geroztik hona* ‘since then’, literally ‘from a later time hither’, *hona* being the allative form of *hemen* ‘here’. Semantically, the spatio-temporal ablative reinforces the more abstract instrumental; see also section 6.1 on alternations involving these two markers. The resulting form *geroztik* functions both as an adverb and as a postposition governing a complement in instrumental case. The instrumental-ablative string seen in *geroztik* may fall under the concept of ‘compound case’ (see Blake 2001: 106), although it should be noted that the morphological make-up of *-ztik* remains transparent, and should therefore be distinguished from that of the directional *-rantz* and terminative *-raino* (see Table 1). The latter are diachronically derived from the allative case suffix *-ra*, but this should be seen as historical reconstruction, not synchronic segmentation.

The story of *gero* is not finished, as it also provides the context for a rather unique shift from case suffix to (part of a) postposition (Trask 1997: 241; Zubiri and Zubiri 2000: 81, 728; Artiagoitia 2003: 718; de Rijk 2008: 511). Compare (17a–d).

- (17) a. *gabon-e-z* ***gero***
 Christmas-DET.PL-INS later/after
 b. *gabon-ez* **=*kerro***
 Christmas-INS after
 c. *gabon-ezkerro*
 Christmas-after
 d. *gabon-a-k* ***ezkerro***
 Christmas-DET.PL after
 ‘after/since Christmas’
 (Artiagoitia 2003: 718)

In (17a), *gero* is used as a postposition governing a complement in the instrumental case *-z*. The complement here is a plural form, as marked by the suffix *-e*

preceding the case marker. This represents the origin of the construction, and is still the orthographic standard in modern Basque. As a next step, *gero* cliticizes to its complement, as evidenced by voicing assimilation of /g/ to the preceding segment /z/, giving rise to the cluster [zk] in *-z=kerō* (§#g>§=k in IPA notation, where # represents a word boundary and = a clitic boundary). This is the stage depicted in (17b), which represents the current spoken form of (a). Now, the form *-ez* not only marks the instrumental plural, but also occurs as an allomorph of *-z* after a consonant in contexts where the host does not carry overt number marking. This homonymy explains the pervasiveness of *-ez*, which has come to be processed as a single morpheme regardless of its morphological genesis. The enclitic *=kerō* then fuses with this *-ez* to invariable *-ezkerō*, as illustrated in (17c). The writing and analysis as a single word is inappropriate with lexical nouns, but has been adopted into the standard in lexicalized forms such as *honezkerō* ‘from now on’, where *ezkerō* has fused with the oblique demonstrative stem *hon-*. The end result, however, is not **gabon ezkerō*, with a mere word boundary shift, but the construction shown in (17d), where the emergence of a new postposition *ezkerō* coincides with the reappearance of individuation and number marking on the complement. This restoration of inflectability of the host is a clear symptom of syntagmatic severance.

From (17a) to (17c), we have a classic case of grammaticalization, with *-ezkerō* following the same cline as the benefactive, the comitative, the directional, and the terminative. Yet except for the possible analysis of *honezkerō* ‘from now on’ and *horrezkerō~harrezkerō* ‘from then on’ as bimorphemic with a case marker *-ezkerō* to which we might give the fancy label ‘subsequentive’, the new *ezkerō* is no longer bound, and therefore not a case marker, but a – diachronically complex but synchronically simple – postposition. Following the definition in Norde (2009: 186), this debonding of *ezkerō* is an instance of degrammaticalization.

4.3 Deverbal adpositional strategies

Complex adpositions can further be based on verbal forms. An English example would be *concerning* NP, S. The syntactic relations correspond as follows: NP is the object complement of the verb *concern*, which then functions as the complement of the new adposition *concerning*. The subject slot of the verb is deleted. The verb form is marked as dependent to S by its non-finite morphology (here a gerund), and constitutes the predicate of an adverbial modifying clause. The new adpositional phrase maintains this function as an adverbial modifier of S, the only difference being that its status has shifted from clausal to nominal.

Where the adpositional expression contains both a verb and a noun as in *taking into account* NP, the clausal nature makes it less amenable to go down the paths of attrition, obligatorification, and coalescence (cf. Lehmann 2015: 174). Its internal structure remains transparent and motivated, which justifies the label “adpositional strategy”, i.e. the use of an expression that functions like an adposition without being one. While being a conventionalized chunk, it remains at the stage of a periphrastic expression rather than joining the lexical class of (complex) adpositions.

Some examples of deverbal adpositional expressions from Basque are presented in (18a–c).

- (18) a. *eguraldi-a ikus-i-rik*
 weather-DET see-PFV-PTV
 ‘in view of the weather’
- b. *egoera kontu-an har-tu-rik*
 situation account-LOC take-PFV-PTV
 ‘considering the circumstances’
- c. *gero-a kontu-an har-tu-z*
 later-DET account-LOC take-PFV-INS
 ‘considering the future’

(18a) consists of a verbal stem followed by the partitive suffix, which functions here as a converb marker. (18b) also features the partitive, but the whole adpositional periphrasis *kontuan harturik* contains another noun. We see the same construction again in (18c), the only difference being that the verb is now marked as dependent by the instrumental suffix.

Another deverbal adposition is *delata* ‘due to’ in (19).

- (19) *euri-a Ø-d-Ø-Ø-ela eta*
 rain-DET S.3-PRS-S.SG-IND.INTR-SR and
 ‘due to rain’

This time, the adpositional complement corresponds to the subject of the verb. The verb, a form of the copula *izan* ‘be’, is followed by the subordinating *-ela*, whose main function is to embed declarative speech reports. The periphrasis ends in *eta* ‘and’, which has shifted again from coordinating to subordinating (cf. example 15 and Figure 1).

The final member in this section will be what might be described as an agreeing postposition, as its form indexes the person and number of its complement; see (20a–c).

- (20) a. *euskara-ri* *∅-da-go-∅-ki-**o**-n-ez*
 Basque.language-DAT S.3-PRS-be-S.SG-IO-**IO.3SG**-REL-INS
 ‘as regards the Basque language’
- b. *ondorio-e-i* *∅-da-go-∅-ki-**e**-n-ez*
 consequence-PL-DAT S.3-PRS-be-S.SG-IO-**IO.3PL**-REL-INS
 ‘as for the consequences’
- c. *gu-ri* *∅-da-go-∅-ki-**gu**-n-ez*
 1PL-DAT S.3-PRS-be-S.SG-IO-**IO.1PL**-REL-INS
 ‘as far as we are concerned’

These are forms of the verb *egon* ‘to be, stay, exist’, whose valency has been increased to include an indirect object corresponding to the adpositional complement. In (20a–c), this complement is third person singular, third person plural, and first person plural, respectively. This is reflected in the verb form, where the corresponding cross-reference suffixes are *-o*, *-e*, and *-gu*. The agreeing verb forms are then subordinated by adding a nominalizing relativizer followed by instrumental case.

4.4 Morphological reduction and flexible word class membership

The default construction of a complex adposition is [NP-case1 [Nrel-case2]], where NP stands for the adpositional complement, case1 is the case governed by the adposition (whose nucleus is Nrel), and case2 marks the whole construction as adverbial within a clause. Note that the term “default construction” is a hyperonym to the “canonical construction”, which is [N-GEN [Nrel-LOC]] and in which case1 and case2 are represented by their most expected terms, cf. section 3.3.

The default construction can be reduced in several ways, most notably by losing case1 and/or case2 (cf. Lehmann 2015: 85). For the omission of case1, see the discussion on differential case marking in section 6.1. The omission of case2, called “inflection dropping” in Hualde (2002: 326), leaves a bare noun as the – henceforth simple – adposition. While the a-versions of (21) and (22) are case-marked as adverbials, the forms in the b-versions are morphologically nouns.

- (21) a. *zu-ta-z* ***gain-era***
 2(SG)-∅-INS top-ALL
- b. *zu-ta-z* ***gain***
 2(SG)-∅-INS besides
 ‘in addition to you’

(Zubiri and Zubiri 2000: 256)

- (22) a. *zuzendari-a-ren* ***menpe-an*** *n-a-go*
 director-DET-GEN domination-LOC S.1SG-PRS-be
 b. *zuzendari-a-ren* ***menpe*** *n-a-go*
 director-DET-GEN domination S.1SG-PRS-be
 ‘I am under the authority of the director’

(Zubiri and Zubiri 2000: 250)

The omission or optionality of overt adverbial marking leads to the establishment of a set of lexical items with flexible word class membership. Table 8 presents a preliminary inventory.

Table 8: Basque adpositions with flexible word class membership.

form	meaning as noun	meaning as adposition
<i>alde</i>	‘side’	‘in favour of’
<i>bizkar</i>	‘back’ (body part)	‘at the expense of’
<i>eske</i>	‘demand’	‘asking for’
<i>esku</i>	‘hand’	‘in the hands of’, ‘under the control of’
<i>gain</i>	‘top’, ‘upper part’	‘besides’, ‘in addition to’
<i>kanpo</i>	‘land’, ‘exterior’	‘outside (of)’
<i>landa</i>	‘field’	‘outside’, ‘besides’, ‘apart from’
<i>menpe</i>	‘dominion’, ‘domination’	‘under the authority of’
<i>truke</i>	‘exchange’	‘in exchange for’

In an adpositional slot, these nouns are always used without the individualizer *-a*. This is a clear indication that syntactically these forms are not nouns, i.e. not heading a noun phrase. We will get back to the use of bare nouns as adpositions in section 7.

5 Synchronic variation

Basque being a pluricentric language, diatopic variation remains important. Table 9 calculates the ratio of exponents for selected location nouns in two newspapers, *Berria* and *Herria*, archived in the digital corpus *Ereduzko Prosa Gaur* (‘Exemplary prose today’). The two newspapers originated in the Southern and Northern Basque Country, respectively, and predominantly use the respective regional standards. The southern *Berria* corpus consists of 10 million words, the northern *Herria* corpus of 2 million.

Table 9: Diatopic distribution of Basque location nouns in locative case.

form	meaning	occurrence per million words in <i>Herria</i> (North)	occurrence per million words in <i>Berria</i> (South)	ratio <i>Herria/Berria</i> (“index of Northernness”)	
<i>aintzinean</i>	‘in front’	401.0	0.1	4010.0	
<i>aitzinean</i>	‘in front’	331.0	42.2	7.8	
<i>gibelean</i>	‘behind’	92.5	16.7	5.5	
<i>ondoan</i>	‘nearby’; ‘after’	748.5	158.7	4.7	rather Northern
<i>barnean</i>	‘inside’	178	41.4	4.3	
<i>aldean</i>	‘on the side’	299.0	84.1	3.6	
<i>artean</i>	‘between’	1294.5	1343.1	1.0	
<i>azpian</i>	‘under’	48.0	51.8	0.9	neutral
<i>inguruan</i>	‘around’	480	611.7	0.8	
<i>aldamenean</i>	‘in the vicinity’	3.5	14.7	0.2	
<i>aurrean</i>	‘in front’	14.0	587.6	0.0	
<i>ostean</i>	‘behind’; ‘after’	2.5	476.9	0.0	rather Southern
<i>atzean</i>	‘behind’	1.0	188.1	0.0	
<i>barruan</i>	‘inside’	1	313.3	0.0	
<i>alboan</i>	‘on the side’	0.5	34.7	0.0	
<i>barrean</i>	‘inside’	0.0	4.2	0.0	

Table 9 confirms the common characterization of location nouns such as *ai(n)tzinean* ‘in front’ and *gibelean* ‘behind’ as Northern Basque, whose Southern counterparts are *aurrean* and *atzean/ostean*. It further shows that the Northern use oscillates between the forms *aintzinean* and *aitzinean*, the latter resulting from cluster simplification (Trask 1997: 142). What the table does not show is the proportion of adverbial vs. adpositional use (i.e. without vs. with a complement) and the ratio of verbal vs. nominal complements, which typically correlates with temporal vs. spatial meanings: while *ostean* frequently follows the participle to express ‘after’, *atzean* ‘behind’ is purely spatial.

In addition to preferring different lexical items, the regional varieties also display grammatical differences. The borrowed adposition *kontra* ‘against’ governs the genitive in Southern Basque (23a), but (at least used to govern) the dative in the North (23b), when used with a spatial sense (de Rijk 2008: 81).

- (23) a. *horma-ren kontra*
 wall-GEN against
 b. *horma-ri kontra*
 wall-DAT against
 ‘against the wall’

Note further certain differences in the form of case-markers: the allative is *-ra* in the West, but *-rat* in the East. The ablative too has various allomorphs in diachronic and synchronic variation such as *-ik, -rik, -ti, -tik, -tika, -tikan*, which hints at the historical affinity between the (presumably older) partitive and the (likely innovative) ablative (cf. Trask 1997: 205). Finally, the locative with animate entities is *-gan* vs. *baitan*, also roughly following a West-East division (cf. Fernández et al. 2016). This variation is evidence of the conventionalized status of these forms. The conventionalization has brought about a reduction of the paradigm; the semiotically ideal outcome would be a single member per domain and variety, so that form-meaning isomorphism will be reestablished.

6 Further grammatical issues

6.1 Differential case marking on the complement

As presented in section 3.3, the canonical construction for adpositional phrases is [N-GEN [Nrel-LOC]]. Basque displays however a fairly regular alternation whereby the genitive on the complement can be omitted. Compare (24a) and (24b) (Zubiri and Zubiri 2000: 243).

- (24) a. *ume-a-ren gain-ean*
 child-DET-GEN top-LOC
 ‘on top of the child’
 b. *sukalde-ko mahai gain-ean*
 kitchen-MR table top-LOC
 ‘on the kitchen table’

This is an instance of differential marking conditioned by the Nominal Hierarchy (cf. Dixon 1994: 85), with a split according to the feature [+/-animate]. While both animate and inanimate nouns can be genitive-marked before an adposition such as *gain* ‘top’, it is more common, especially in the spoken language, to omit the genitive on complements with inanimate referents (cf. also Trask 1997: 203;

Hualde 2002: 327; de Rijk 2008: 70–71). In addition to this semantic conditioning, there is the following structural constraint. The genitive can only be omitted if the possessor noun is singular and immediately precedes the adposition. Therefore, differential case marking does not apply when the possessor noun is followed by an attributive adjective, a demonstrative, or the indefinite article *bat*. Basque having phrase-level case marking, the genitive must be attached to whichever element comes last in such noun phrases. In (25), this is the adjective *txiki* ‘small’; the b-version without the genitive is grammatically ill-formed (Zubiri and Zubiri 2000: 243). While postposed modifiers neutralize the split, preposed modifiers such as the relative clause in (26) have no effect on it (de Rijk 2008: 70).

- (25) a. *Zapata-k ohe txiki-a-ren azpi-an Ø-da-u-de.*
 shoe-PL bed small-DET-GEN under-LOC S.3-PRS-be-S.PL
- b. * *Zapatak ohe txiki azpi-an Ø-da-u-de.*
 shoe-PL bed small under-LOC S.3-PRS-be-S.PL
 ‘The shoes are under the small bed.’
- (26) *Feldman bizi Ø-d-Ø-Ø-en etxe aurre-an*
 Feldman live S.3-PRS-S.SG-IND.INTR-REL house front-LOC
 ‘before the house where Feldman lives’

That, despite these structural constraints, the case alternation is primarily conditioned by semantic factors is corroborated by examples such as (27), where zero-marking is possible on an animate complement, because the animal is perceived here as a simple functional landmark, which overrides its animacy (Zubiri and Zubiri 2000: 243).

- (27) *Jarri zaku-a asto gain-ean.*
 put bag-DET donkey top-LOC
 ‘Put the bag on the donkey.’

What is the functional explanation for this split? “It is plainly most natural and economical to ‘mark’ a participant when it is in an unaccustomed role”, writes Dixon (1994: 85). In this context, Aurnague (1996a) mentions that toponyms cannot be case-marked at all when possessing a location noun, as in *Uharte gaina* ‘the heights of Uharte’ (**Uharteren gaina*). Based on whether formal marking of the relation is obligatory, optional, or impossible, we can thus postulate a continuum animate > inanimate object > non-specified location/place (common noun of geographical entity) > specified location/place (proper noun of geographical entity).

For yet another class of adpositions (*arte* ‘among’, *bila* ‘looking for’, *eske* ‘asking for’), the determining factor for differential case marking is referential status: if the complement has generic reference, it is not case-marked, as in (28d) (Zubiri 1994: 184; Zubiri and Zubiri 2000: 249–250).

- (28) a. *lagun-a-ren* *bila*
 friend-DET-GEN looking.for
 ‘looking for the friend’
- b. *lagun bat-en* *bila*
 friend one-GEN looking.for
 ‘looking for a friend’
- c. *lagun-en* *bila*
 friend-GEN(PL) looking.for
 ‘looking for the friends’
- d. *lagun bila*
 friend looking.for
 ‘looking for (any) friends’

The claim that the constructions without the genitive as in (24b) and (28d) are compounds (cf. e.g. Villasante 1978: 48; Hualde 2002: 329) requires further examination. Firstly, the compound hypothesis does not address the animacy split. Secondly, it would create a mismatch between structure and meaning, for *Feldman bizi den etxe aurrean* in (26) means ‘before the house where Feldman lives’ rather than ‘at the house front where Feldman lives’; Feldman lives in the house, not in front of it. Incorporating the modifier into the compound, yielding a structure ‘at the house-where-Feldman-lives-front’, is not convincing either. Thirdly, postulating a compound *diru eske*, perhaps translatable as ‘money request’, in a sentence such as *diru eske zetorren* ‘he came asking for money’ (Zubiri and Zubiri 2000: 252) would require a second nominal argument of the verb *etorri* ‘come’. The sentence would then correspond to an impossible ‘he came money request’. Juxtaposition therefore seems to be a more appropriate term than compounding to describe these constructions.

These observations lead to a more general statement about the phrase-level inflection of Basque, namely that it is not impervious to the internal structure of the noun phrase. Trask’s (1997: 89) categorical statement that “it is noun phrases, and only noun phrases, which are inflected in Basque” thus requires some refinement, for it is in fact nouns, which have a privileged relation to their head. This stronger bondedness may be labelled “morphological affinity” and is another instance of markedness iconicity. While adjectives and determiners can take over the hosting – or, in this context, the omission – of inflection marking, such a

syntactic arrangement is functionally and therefore structurally marked in the contexts surveyed here. As argued above with regards to prenominal modifiers, a compounding analysis has syntactic implications incompatible with the scope of the adposition over the whole complement phrase. In contrast, the affinity analysis put forward here only pertains to the morphological boundary between the adposition and its immediate neighbour, but leaves the syntactic constituency unaffected.

Where the outer relator is omitted as well, as in *diru eske*, the erstwhile noun acquires a secondary function as adposition. Constructions with location nouns such as *aurre* ‘front’ or noun-adposition hybrids such as *eske* ‘request(ing)’ thus have to be distinguished from those with *maila* ‘level’ and *arlo* ‘field’ discussed in section 3.3, where an analysis as compounds is less problematic.

As a consequence of functional differentiation, differential case marking can also correlate with a difference in meaning of the adposition. When *alde* ‘side’ means ‘on N’s side’, i.e. ‘in favour of’, it governs the genitive (29); when it means ‘in the area of N’, as in (30), the complement does not carry any case-marker (Zubiri and Zubiri 2000: 247).

(29) *amnistia-ren alde-ko manifestaldi-a*
 amnesty-GEN side-MR demonstration-DET
 ‘the demonstration for an amnesty’

(30) *Nafarroa alde-ko zainzuri-a-k*
 Navarre side-MR asparagus-DET-PL
 ‘asparagus from the Navarre area’

Differential case marking is not limited to the alternation between the genitive and zero; *begira* ‘looking’ can govern genitive, dative, or allative case. With a genitive-marked complement, the postposition corresponds to ‘waiting for’; with a dative-marked complement, it translates as ‘watching’; and with the allative, it expresses ‘looking (in a certain direction)’ (Zubiri and Zubiri 2000: 254; Elhuyar s.v. *begira*); see Table 2 for examples. Also interesting is the ablative-instrumental alternation observed among others with *landa* ‘outside, beyond’, as it is indicative of the more concrete meaning of the ablative case marker *-tik*, contrasting with the relatively desemanticized function of instrumental *-z*. Phrases such as *hiritik landa* ‘outside the city’ and *Euskal Herritik landa* ‘outside the Basque Country’ have a spatial meaning (de Rijk 2008: 76). In contrast, the instrumental case in *hiriaz landa* ‘apart from the city’ has a more abstract function, and can therefore also attach to nominalized verb forms, as in [...] *izateaz landa* ‘besides being [...]’. Further illustrations of this opposition are *herritik kanpo* ‘outside the village’,

with an ablative expressing a spatial relation, contrasting with *legez kanpo* ‘outside the law’, *epez kanpo* ‘after the deadline’, *lekuz kanpo* ‘out of place’, or *jokoz kanpo* ‘offside’, where the instrumental is preferred on complements which semantically are not locations.

6.2 Differential case marking on the adposition

Section 4.4 has presented lack of case marking on the erstwhile complex adposition as a departure from the default construction. Yet, it may be insufficient to analyze such lack as merely structural attrition; rather, the presence or absence of a marker can be used to distinguish different meanings, as in (31a) vs. (31b) (de Rijk 2008: 77).

- (31) a. *ama-ren alde-an*
 mother-GEN side-LOC
 ‘next to mother’
 b. *ijito-en alde*
 gypsy-GEN(PL) side
 ‘in favour of the gypsies’

The noun *alde* ‘side’ has given rise to two separate synchronic reflexes, one being a complex postposition *aldean* with a predictable spatial meaning ‘on/to the side of’, and the other one being the simple postposition *alde* with a more figurative meaning of ‘side’, as in *take sides* (cf. Villasante 1978: 52). Another noun with two such reflexes is *arte* ‘interval’, which in adpositional position means predictably ‘between’ when followed by a local case, but has come to be used in its bare form to express ‘until’ (cf. Villasante 1978: 56–57; de Rijk 2008: 533).

The pattern of locative marking preserving the spatial meaning on the one hand, and zero-marking expressing figurative relations, on the other, is also visible with *bestalde*, a compound of *beste* ‘other’ and the already familiar *alde* ‘side’ (32). It tends to be in locative case when meaning ‘on the other side of’, but is zero-marked when having meanings such as ‘in addition to’ (Zubiri and Zubiri 2000: 257; Elhuyar s.v. *bestalde*).

- (32) a. *muga-z best-alde-an*
 border-INS other-side-LOC
 ‘on the other side of the border’
 b. *igande-a-z best-alde*
 Sunday-DET-INS other-side
 ‘in addition to Sunday’

While locative-marking on an adpositional noun correlates with spatial semantics in examples such as (31) and (32), the locative is also the case that can be most readily inferred. In other words, the locative is the functionally unmarked case among the spatial cases, and this functional unmarkedness can translate into structural unmarkedness. This explains the lack of case-marking in (33a), but its presence in (33b) (Zubiri 1994: 181).

- (33) a. *Etxe-tik kanpo bazkal-du Ø-d-Ø-u-t gaur.*
 house-ABL outside have.lunch-PFV 0.3-PRS-O.SG-IND.TR-A.1SG today
 ‘I’ve eaten out for lunch today.’ (lit.: ‘outside the house’)
- b. *Ezin Ø-d-Ø-a gela hon-eta-tik kanpo-ra irten.*
 NPOT S.3-PRS-S.SG-IND.INTR room D1-Ø-ABL outside-ALL exit
 ‘One cannot go outside from this room.’

Apart from the issue of markedness, the two constructions are fundamentally different. In (33a), we have a (simple) postposition *kanpo* governing a complement in ablative case. In (33b), in contrast, the ablative phrase *gela honetatik* ‘from this room’ and the allative-marked *kanpora* ‘outside’ are merely juxtaposed adverbials. The whole is no different from *Bilbotik Donostiara* ‘from Bilboa to San Sebastian’. Both structurally and semantically, then, there is no reason to include *kanpora* among the complex adpositions. Another indication of the grammatical split between postpositional and adverbial use of *kanpo* is the fact that in independent (i.e. adverbial) use, it cannot be used in its bare form, but must carry the locative case-marker to produce the adverbial expression *kanpoan* ‘outside’ (de Rijk 2008: 76). On the functional motivation of this kind of ‘economic case marking’, see section 7.

6.3 Adnominalization of adpositional phrases

Adverbial phrases are routinely turned into adnominal modifiers by adding *-ko*, as in *etxe aurreko parkea* ‘the park in front of the house’ (without genitive marking of the complement *etxe*) or *zure aldameneko mutila* ‘the boy next to you’ (with the genitive marker *-re* on the possessive pronoun *zure* ‘your’). This is an important typological difference between languages. In the European type, an adverbial adpositional phrase can be adnominalized by simply omitting the predicate linking it to the noun phrase to be modified. In many languages, this is not possible. In Basque, **etxe aurrean parkea* is ill-formed, and the modifier suffix *-ko* must be added. As a result of structural simplification, the *-an* in *aurrean* is dropped, yielding the form *aurreko* ‘(which is) in front’ (on the underlying locative, cf. Azkarate and Altuna

2001: 100; de Rijk 2008: 91). The suffix *-ko* has often been described as a distinct ‘locative genitive’ case (e.g. in Lafitte 2001 [1944]: 54), but as convincingly argued in Trask (1997: 94; 100–102) such an analysis is unnecessary and misleading. It is a suffix used to derive complex modifiers, and its hosts are not limited to locative adverbials. It can attach to already case-marked noun phrases, as in *mendirako bidea* ‘the road to the mountain’, derived from *mendi-ra* ‘mountain-ALL’. On the distribution of *-ko*, see also Aurnague (2004: 87–89) and de Rijk (2008: 89–105).

6.4 Adpositions vs. common nouns

Just as there are intermediate cases between case suffix and postposition, there are possessive constructions that are not typically listed among the postpositions even though they fulfil a similar function, such as *ozeanoaren hondoan* ‘at the bottom of the ocean’. One criterion to distinguish complex postpositions from mere possessive constructions is modifiability of the head, as in *ozeanoaren hondo sakonean* ‘at the deep bottom of the ocean’. Another are the selectional restrictions with regards to possible possessors, which are fewer for location nouns. They also differ in their referential autonomy, as evidenced by the observation that location nouns display less syntactic manipulability than nouns referring to component parts of the possessor such as ‘roof’ or ‘door’; Aurnague (2004: 97–105). In summary, location nouns can be described as a semi-closed class with fuzzy boundaries; see Aurnague (1996a) for an in-depth discussion, Aurnague (1996b) for an inventory of localization strategies in Basque, and Aurnague (2004: 105–109) on the continuum between location nouns and common (i.e. component) nouns.

7 Conclusion and outlook

One major insight from the Basque data, in particular those presented in Table 8 and section 6.2, is that the distinction between a simple and a complex adposition is not very robust. It is a common development for the additional relator that makes an adposition complex to be omitted as the adposition becomes autonomous from its etymon; cf. Hualde (2002: 326–333). The omission of the outer relator further brings about a split between adpositional and adverbial use, as the relator persists when there is no adpositional complement. Compare (34) and (35) from German and French, where the a-version exemplifies the simplified adposi-

tional use of an erstwhile noun, the b-version shows the ill-formedness of independent use, and the c-version provides the correction of (b).

- (34) a. *Wir gehen Richtung Bahnhof.*
 we go direction station.
 ‘We are going towards the station.’
 b. * *Wir gehen Richtung.*
 we go direction
 c. *Wir gehen in die Richtung.*
 we go in the direction
 ‘We are going in that direction.’
- (35) a. *Côté températures, pas de changement.*
 side temperatures NEG PTV change
 ‘As for the temperatures, no change.’
 b. * *Côté, pas de changement.*
 side NEG PTV change
 c. *De ce côté, pas de changement.*
 from this side NEG PTV change
 ‘In that regard, no change.’

The loss of the outer relators *in* and *de*, and of determiners such as *die* and *ce*, is a corollary of the adpositionalization of the erstwhile nouns *Richtung* and *côté*. This is the same process as in Basque, where the locative case *-an* and, where applicable, the determiner *-a* undergo attrition if and only if the noun is used in an adpositional slot. This “slotting” shaves the filler of redundant formal material; the filler adapts to the slot. The same semiotic phenomenon is also found in the following types of utterances.

- (36) a. *Caution: no smoking. Violators will be fined.*
 b. *Danger: Flammable gas. Keep fire or flame away.*
 c. *Destination: Frankfurt. Go to gate.*

Expressions like *caution*, *danger*, and *destination* are highly predictable in these contexts and do not vary much, that is, their paradigm is restricted. They merely introduce what follows, which is from an open set and can hardly be reduced. The signalling element is akin to a focus marker, the element from the open set is put into focus. Together they constitute a pre-clausal topic, and what follows is a comment on that topic. Instead of *côté températures*, we could just write *Températures: pas de changement*.

This structure explains why independent use is less prone to reduction: From the perspective of its syntagmatic relations, it is like a focus marker without a focused element, like a colon or an exclamation mark with nothing preceding. The explanation that a bare noun phrase can function as an adverbial if it is akin to a topic is also applicable to the use of *bestalde*, a contraction of *beste* ‘other’ and *alde* ‘side’, as a sentence-initial adverb(ial) meaning ‘besides’. This construction is pragmatically comparable to the weather report *London: sunny with occasional showers*. There is no need to make the relation between the topic *London* and the comment *sunny with occasional showers* explicit by saying *in London* or *as for London*. Structural attrition is thus a consequence of improved inferability; the latter is a consequence of routinization.

The omissibility of the inner and outer relators suggests a continuum of grammaticalization as shown in Figure 2. The stages represent merely degrees rather than paths of grammaticalization, so that adjacent stages may or may not correspond to actual diachronic developments. Lrel is a relational lexeme such as a relational noun or a bivalent verb, but as discussed at several occasions, word class membership may be transient. The case in stage 4 fills all the postnominal slots at once.

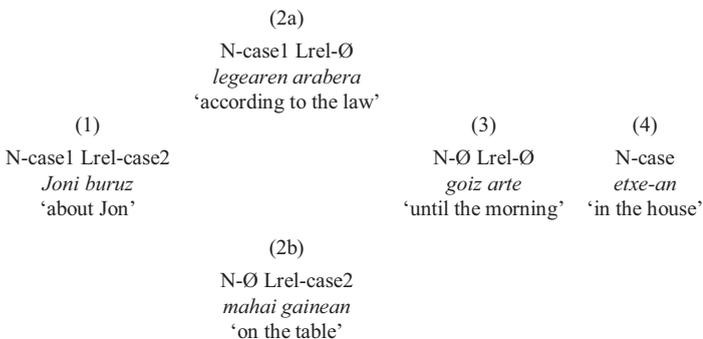


Figure 2: Basque relator constructions on a continuum of grammaticalization.

Adpositions in stage 1 are emerging and usually identifiable as such only within a relevant construction. The intermediate stage 2 comes in two forms, depending on which relator is omitted. In stage 2a, the adposition is no longer complex, but the inner relator reveals that the construction is still within the realm of syntax. In stage 2b, the adposition remains complex, but juxtaposition ushers in morphological coalescence; this is where most spatial adpositions are found in Basque. Stage 3 combines the monomorphemic status of the adposition of stage 2a with the juxtaposition of stage 2b. In stage 4, grammaticalization has reached its endpoint: the adposition has become bound, in other words, has turned into a case suffix.

Abbreviations

A	subject index in transitive inflection	MR	modifier
ABL	ablative	NEG	negator
ABS	absolutive	NPOT	impossibility
ALL	allative	NR	nominalizer
BEN	benefactive	O	(direct) object index
D	deictic	PFV	perfective
DAT	dative	PL	plural
DEST	destinative	PROL	prolative
DET	determiner	PROSP	prospective
ERG	ergative	PRS	present
GEN	genitive	PST	past
IND	indicative	PTV	partitive
INS	instrumental	REL	relativizer
INTR	intransitive	S	subject index in intransitive inflection
IO	indirect object (index/flag)	SG	singular
IPFV	imperfective	SR	subordinator
LOC	locative	TR	transitive

References

- Artiagoitia, Xabier. 2003. Adjunct subordination. In José Ignacio Hualde & Jon Ortiz de Urbina (eds.), *A grammar of Basque*, 710–762. Berlin & New York: Mouton de Gruyter.
- Aurnague, Michel. 1996a. Les noms de localisation interne: tentative de caractérisation sémantique à partir de données du basque et du français. *Cahiers de lexicologie* 69. 159–192.
- Aurnague, Michel. 1996b. Petit dictionnaire raisonné des noms de localisation interne du basque. *Cahiers de Grammaire* 21. 1–44.
- Aurnague, Michel. 2004. *Les structures de l'espace linguistique: Regards croisés sur quelques constructions spatiales du basque et du français*. Louvain & Paris: Peeters.
- Azkarate, Miren and Patxi Altuna. 2001. *Euskal morfologiaren historia* [The history of Basque morphology]. Donostia: Elkarlanean.
- Basque Government, Government of Navarra & Office Public de la Langue Basque. 2016. *VI Encuesta Sociolingüística del conjunto del territorio del euskera*. http://www.euskara.euskadi.eus/r59-738/es/contenidos/informacion/argitalpenak/es_6092/encuestas_sociolingüísticas.html. (accessed 3 March 2020)
- Blake, Barry J. 2001 [1994]. *Case*. 2nd edn. Cambridge etc.: Cambridge University Press.
- Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge etc.: Cambridge University Press.
- de Rijk, Rudolf P. G. 2008. *Standard Basque. A progressive grammar*. Cambridge, Massachusetts & London: MIT Press.
- Dixon, R. M. W. 1994. *Ergativity*. Cambridge etc.: Cambridge University Press.

- Fernández, Beatriz, Ane Berro, Itziar Orbegozo, Alazne Arriortua & Josu Landa. 2016. *Euskara Bariazioan. Basque in Variation (BiV)*. UPV/EHU. <http://basdisyn.net/Bas&Be/biv/> (accessed 3 March 2020).
- Haase, Martin. 1992. *Sprachkontakt und Sprachwandel im Baskenland. Die Einflüsse des Gaskognischen und Französischen auf das Baskische*. Hamburg: Helmut Buske.
- Hualde, José Ignacio. 2002. Regarding Basque postpositions and related matters. In Xabier Artiagoitia, Patxi Goenaga and Joseba Lakarra (eds.), *Erramu Boneta. A Festschrift for Rudolf P.G. de Rijk. Supplements of ASJU 44*, 325–339. Bilbao: Universidad del País Vasco.
- Hualde, José Ignacio. 2003a. Case and number inflection of noun phrases. In José Ignacio Hualde & Jon Ortiz de Urbina (eds.), *A grammar of Basque*, 171–186. Berlin & New York: Mouton de Gruyter.
- Hualde, José Ignacio. 2003b. Postpositions. In José Ignacio Hualde & Jon Ortiz de Urbina (eds.), *A grammar of Basque*, 187–190. Berlin & New York: Mouton de Gruyter.
- Jendraschek, Gerd. 2007. *La notion modale de possibilité en basque. Morphologie, syntaxe, sémantique, variations diachronique et sociolinguistique*. Paris: Harmattan.
- Jendraschek, Gerd. 2008. *Basque in contact with Romance languages*. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *Grammars in Contact: a cross-linguistic typology*, 143–162. Oxford: Oxford University Press.
- Jendraschek, Gerd. 2012. From polyfusional to post-fusional: Obsolescence and innovation in Basque predicate morphosyntax and its typological implications. *Rice Working Papers in Linguistics* 3. 1–33.
- Jendraschek, Gerd. 2016. Nebensätze im Sprachvergleich. In Jianhua Zhu, Jin Zhao & Michael Szurawitzki (eds.), *Akten des XIII. Internationalen Germanistenkongresses Shanghai 2015 – Germanistik zwischen Tradition und Innovation: Band 2* (Publikationen der Internationalen Vereinigung für Germanistik [IVG]), 137–142. Frankfurt am Main: Peter Lang.
- Lafitte, Pierre. 2001 [1944]. *Grammaire basque (Navarro-Labourdin littéraire). Edition revue et corrigée*. 3rd edn. Donostia & Baiona: Elkarlanean.
- Lehmann, Christian. 1983. Rektion und syntaktische Relationen. *Folia Linguistica* 17. 339–378.
- Lehmann, Christian. 2015 [1995]. *Thoughts on grammaticalization*. 3rd edn. Berlin: Language Science Press.
- Manterola, Julen. 2015. *Euskararen morfologia historikorako: artikulua eta erakusleak* [Towards a historical morphology of Basque: articles and demonstratives]. Vitoria-Gasteiz: University of the Basque Country dissertation. <http://hdl.handle.net/10810/15848> (accessed 3 March 2020).
- Norde, Muriel. 2009. *Degrammaticalization*. Oxford: Oxford University Press.
- Penny, Ralph. 2002. *A history of the Spanish language*. Second edition. Cambridge: Cambridge University Press.
- Trask, Robert L. 1997. *The history of Basque*. London & New York: Routledge.
- Trask, Robert L. 1998. The typological position of Basque: then and now. *Language Sciences* 20 (3). 313–324.
- Trask, Robert L. 2008. *Etymological dictionary of Basque*. Edited for web publication by Max W. Wheeler. University of Sussex. www.sussex.ac.uk/english/documents/lxwp23-08-edb.pdf (accessed 3 March 2020).
- Villasante, Luis. 1978. *Estudios de sintaxis vasca*. Oñate: Editorial Franciscana Aranzazu.
- Zubiri, Ilari. 1994. *Gramática didáctica del euskera*. Bilbao: Didaktiker.
- Zubiri, Ilari & Entzi Zubiri. 2000. *Euskal gramatika osoa* [Complete Basque grammar]. Bilbo: Didaktiker.

Corpora

Berria. www.berria.eus (accessed 3 March 2020).

Elhuyar hiztegiak. <https://hiztegiak.elhuyar.eus/eu/> (accessed 3 March 2020).

Ereduzko Prosa Gaur. <http://www.ehu.eus/euskara-orria/euskara/ereduzkoa/> (accessed 3 March 2020).

Gerd Jendraschek

10 Structure and origin of Turkish adpositions

Abstract: Of the 117 Turkish adpositions in our corpus, 85 are morphologically complex. Of these, 60 contain a possessive suffix representing the complement. Within the group of possessed adpositions, the main distinction is between those with fixed outer case, representing abstract relations, and those with variable case, expressing spatial relations. A particularity of Turkish are the case-alternating postpositions governing the genitive forms of personal pronouns, demonstrative pronouns, and the interrogative pronoun for ‘who’, but the unmarked nominative elsewhere. Of the 117 adpositions surveyed, 51 only occur with nominal complements, 56 occur with both nominal and deverbal complements, and 10 only occur with deverbal complements (“secondary conjunctions”). As regards language contact and corpus planning, Turkish adpositions are a moderately conservative word class, with identifiable neologisms only constituting 9 percent of adpositions. It is common for certain postposition-case combinations to become lexicalized. This process is a correlate of desemantization, as the spatial noun adopts a metaphorical meaning. Turkish rarely permits the use of adpositional phrases as adnominal attributes, so that a modifier suffix must be attached to the adpositional phrase to derive an attribute. However, such adnominalization of adpositional phrases is restricted to those expressing location in time, as well as relational nouns with locative case marking. Postpositional nouns marked by other suffixes require attributive clauses to function as adnominals. Finally, it could be demonstrated that differential argument marking in Turkish does not only apply to the accusative, but also to the genitive.

Keywords: desententialization, differential marking, genitive-attracting pronouns, metaphorical shift, secondary conjunctions

1 Introduction

This chapter presents the adpositions of Turkish, with a special focus on the characteristics of the various kinds of complex adpositions. Section 2 gives some background information on the genetic affiliation of Turkish and presents an overview of its typological profile, including its case system. Section 3 examines relevant criteria for the description of Turkish adpositions, such as possessive

Gerd Jendraschek, Gyeongsang National University, Jinju, South Korea. gerd@gnu.ac.kr

<https://doi.org/10.1515/9783110686647-011>

marking, local cases used as outer relator, their morphological structure, differential complement marking depending on the word class of the complement, and the role of adpositions in clause linkage. Section 4 then deals with diachronic and synchronic variation, in particular the competition between adpositions of Turkic vs. Arabic origin, lexicalization, and semantic change. Section 5 will look once more at differential complement marking, this time conditioned by metaphorical shift and the referential status of the complement. It will also show that adnominal use of adpositional phrases is restricted and emphasizes once more the central role of the locative among the local cases. Section 6 summarizes the most important findings.

2 The Turkish language

With close to 80 million native speakers, Turkish is the main representative of the Turkic language family, whose geographic distribution reaches from East Asia (Northeastern Russia, Western China) to the Balkans (see Boeschoten 1998: 4). In addition to these older areas of settlement, migrations in the second half of the 20th century have produced important Turkish-speaking diasporas, most notably in Germany, where 2.8 million people have a Turkish background (Bundeszentrale für politische Bildung 2019). There is increasing evidence that the Turkic languages are related to other languages in North-East Asia such as Mongolian, Korean, and Japanese, with which they are often grouped together as an Altaic or Transeurasian macro-family (Robbeets 2005: 423). While such a genetic connection remains controversial (cf. Robbeets 2015: 6), these languages share many typological features such as head-final syntax, agglutinative morphology, and vowel harmony (cf. Starostin 2016).

Turkish features a system of six cases, if including zero-marking used for subjects as well as direct objects with a low degree of individuation; see Göksel and Kerslake (2005: 175–176; 371–384) for the details of differential object marking, i.e. the functional motivation for overt accusative marking. The case paradigm includes three local cases, which are a conflated allative-dative (henceforth “dative”), a locative, and an ablative case. Vowel harmony reduces vocalic contrasts in (most) suffixes to the opposition of closed (/i~y~u~u/) vs. open (/ɛ~ɑ/). For open vowels, the rounding opposition is neutralized in favour of the unrounded set. The resulting archiphonemes are represented by the capitals *I* and *A*. The same notational convention applies to consonants undergoing voicing assimilation. Table 1 lists the cases and their allomorphs; the consonants in round brackets are added when the preceding segment is a vowel.

Table 1: Turkish case paradigm.

label	form
nominative	–∅
accusative	–(y)l
genitive	–(n)In
dative	–(y)A
locative	–DA
ablative	–DAN

In plural forms, a suffix *-lar* precedes the case markers; first and second person pronouns have idiosyncratic plural forms in *-z* (*ben* ‘I’, *biz* ‘we’; *sen* ‘you (sg)’, *siz* ‘you (pl)’). Possession is typically marked on both the head and the dependent. On the head, person and number of the possessor are indexed by suffixes following number marking, but preceding case marking; see (1). The possessor is marked by the genitive *-(n)In*, which has the allomorph *-Im* on first person pronouns.

- (1) *dünya-nın önemli şehir-ler-in-de*
 world-GEN important city-PL-POSS.3-LOC
 ‘in the important cities of the world’

Case markers also attach to nominalized forms of verbs (see e.g. Göksel and Kerslake 2005: 92–93). The resulting construction functions as a complement (2) or adverbial (3) clause; for easier readability, zero-morphemes for present tense and third person are omitted from examples.

- (2) *git-me-niz-i bekl-iyor*
 go-NR-POSS.2PL-ACC wait-PROG
 ‘(s)he expects you to leave’

- (3) *yürü-düğ-üm-de*
 walk-NFUT.NR-POSS.1SG-LOC
 ‘when I walk’

What holds for case markers, also applies to adpositions. When describing the adpositions of a language like Turkish, we have to take into account that dependent clauses are typically desententialized to various degrees and that nominalized verb forms can then become the complement of an adposition (cf. Lehmann 1988:

198). The boundary between adpositions and conjunctions is thus fuzzy, and true conjunctions are marginal; this topic will be discussed further in section 3.7.

3 Complex relators in Turkish

The first thing to be noticed is that Turkish has no prepositions. The terms adposition and postposition can therefore be used interchangeably. For the classification of Turkish adpositions, the following eight criteria can be identified:

1. Is the adposition morphologically simple or complex? (section 3.1)
2. Can it be cliticized to its host? (3.2)
3. If morphologically complex, does it contain possessive marking? (3.3)
4. If possessed, can it be followed by all of the three local cases, or is its outer relator fixed? (3.4)
5. If complex, is it derived from a noun or a verb? (3.5)
6. Does it (always, or under certain conditions) require overt case-marking on its complement? (3.6)
7. Does it (also, predominantly, or exclusively) take nominalized verbal complements? (3.7)
8. Is it borrowed, or of Turkish origin? If the latter, is it in competition with a loan? (section 4)

Each section will explain and illustrate the criterion. In order to keep sections concise, more readable, and minimally overlapping, the full list of adpositions, their morphosyntactic characteristics, and their frequency will be relegated to Table 10 in the appendix. For frequency counts and some of the data collection, a 35 million word corpus of Turkish Wikipedia entries has been used, primarily because it represents 21st-century standard Turkish, but also because it contains useful tools for collocation analysis (see Sezer and Sezer 2013). For the sake of brevity, Turkish adpositions will be referred to in the text without the case they govern on their complement, so the Turkish equivalent of ‘against’ will be given as *karşı* rather than *-A karşı*. Finally, verbs will be given in their root or stem form, leaving out the infinitive suffix *-mAk*.

3.1 Simple vs complex adpositions

Simple adpositions are monomorphemic and cannot be synchronically segmented; typical representatives are *için* ‘for’ and *gibi* ‘like’. Lehmann (2015: 85)

claims that *için* is “partly analyzable” as containing *iç-i* interior-POSS.3 ‘its interior’, but Kabak (2006: 61) gives its etymology as **üçün* < **uçun*, an (old) instrumental form of the noun *uç* ‘reason’ (i.e., *uç-un* ‘with the reason’). It is also listed as “one of the four primeval Turkish postpositions”, with attestations predating the thirteenth century (Kabak 2006: 48).

The adjectives *başka* ‘other’ and *doğru* ‘straight, correct’ are also used as adpositions meaning ‘apart from’ and ‘towards’, respectively. The noun *beri* ‘the near side’ can be used as an adposition meaning ‘since’; a similar conversion occurs with the Arabic loan *zaman* ‘time’ in adverbial clauses, where it translates as ‘when’. For *ile* ‘with’ see section 3.2; for *kadar* ‘until’; ‘as much as’, section 3.6.

Arabic loans such as *nazaran* ‘compared to’ and *takiben* ‘subsequent to’ are derived within Arabic (cf. the borrowed nouns *nazar* ‘regard’ and *takip* ‘pursuit’), but can be considered unsegmentable in (at least 21st-century) Turkish. There is furthermore a group of five ‘semi-complex’ adpositions. This label designates adpositions that look complex in the sense that part of their structure can be identified with other morphemes, but their remaining material could not be identified. These are *sonra* ‘after’ (*son* ‘end’), *önce* ‘before’ (*ön* ‘front’), *göre* ‘according to’ (possibly a converb form of *gör* ‘see’), *dolayı* ‘due to’ (*dolay* ‘surroundings’), and *değin* ‘until’ (alternating with shorter *dek*). They are attested in pre-20th century texts and therefore not classified as neologisms; see Turan (1998) and Kabak (2006) for details. For diachronic considerations, they shall be grouped with the complex adpositions; where the focus is on their synchronic status, they will be considered simple.

Of the 117 adpositions listed in Table 10, 85 are morphologically complex. They can be subdivided into those with and those without possessive marking. The first group will be dealt with in section 3.3. Those without possessive marking fall into the following groups, for each of which a few examples will be given:

- (a) Case-marked nouns: *üzere* ‘in order to’; ‘about to’ from the bound noun root *üzer-* ‘surface’, followed by dative case’; *birlikte* ‘together’, literally ‘in oneness’, *yana* ‘in favour of’, literally ‘to the side’.
- (b) Derivations: *ilgili* ‘with regard to’, from *ilgi* ‘interest’, which is in turn derived from the verb root *il* ‘to tie loosely’; *aşkın* ‘surpassing’ from the verb root *aş* ‘surpass’, *yönelik* ‘aimed at’, which looks like the dative form of *yön* ‘direction’, followed by the nominalizer *-lik* seen in *birlik* ‘union’ (one-NR).
- (c) Multi-word forms such as *bu yana* ‘since’, literally ‘to this side’, *bir yana* ‘apart from’, literally ‘to one side’; see also section 3.4.
- (d) Converb forms, such as *olarak* ‘being’, corresponding to ‘as’, or *gelince* ‘as regards’, ‘concerning’, from *gel* ‘come’ (see section 3.5).
- (e) Many of the “secondary conjunctions” fall in this group. This term is used to refer to adpositions that only take nominalized verbal complements (e.g. *git-*

tiğim halde ‘although I went’), and are therefore functionally conjunctions. It is not to be confused with the term “secondary postpositions” as used e.g. in Lewis (1967: 85) for complex postpositions with possessive marking; see section 3.3. Secondary conjunctions differ from primary conjunctions such as *çünkü* ‘because’, which link finite clauses; more on secondary conjunctions in section 3.7;

- (f) Neologisms: These are almost inevitably complex, as they have been created by modifying pre-existing material, often by attaching ad-hoc suffixes; an example is *karşın* ‘despite’, related to *karşı* ‘against’. While such neologisms are a special type of derivation, they are mentioned separately here for their sociolinguistic importance in shaping present-day Turkish; see section 4.1 for a discussion.

3.2 Clitics

In terms of the parameters of integrity and bondedness, clitics follow – in the sense of a potential diachronic path – the complex and simple unbound adpositions, but precede case affixes. Consequently, cliticization is a relevant issue for an exhaustive review of the criteria used in categorizing relators. Moreover, the enclitic presented in this section is frequently used in the formation of complex postpositions (cf. section 3.4), and therefore deserves to be fully understood before we proceed. We are talking about the comitative-instrumental postposition *ile* ‘with’. When it cliticizes to its host, the initial vowel is reduced to a glide after vowels, and dropped after consonants; the final vowel falls under the scope of vowel harmony, changing to *a* when following a syllable containing a back vowel. The resulting clitic form can be represented as $=(y)lA$. While Göksel and Kerslake (2005: 70) assert that it “only rarely” appears as *ile*, the separate form has a corpus frequency of 6149 per million words, making it the second most frequent adposition after *olarak* ‘as’. A corpus study presented in Kabak (2006: 47) also shows that the free form is one of the most frequent postpositions, but also that the bound form is about four times more frequent. In terms of distribution, the free form is indeed very rare after the so-called “genitive-attracting pronouns” (Göksel and Kerslake 2005: 71; cf. section 3.6): The Wikipedia corpus has only 10 examples of *bunun ile* ‘with this’ and 7 of *onun ile* ‘with that/him/her’, compared with 3926 tokens of *bununla* and 1821 of *onunla*. The most frequent collocation for the free form is *adı ile* name-POSS.3 COM ‘under the name of’ (1519 tokens), but here too, the bound form *adıyla* is far more frequent (11347 tokens). Note that *için* ‘for’ also had the bound allomorphs $=(y)çin/= (y)çün$ in earlier stages of Turkish, which are now obsolete (cf. Kabak 2006: 43).

3.3 Complex adpositions with possessive marking

60 of the 85 complex adpositions contain a possessive suffix representing the complement. Adpositions that for semantic reasons are restricted to third person complements can be considered fossilized combinations of N-POSS-CASE: *sırasında* ‘during’, for example, does not occur with first or second person complements. Within the group of possessed adpositions, the main distinction is between those with fixed case, representing abstract relations, and those with variable case, expressing spatial relations; see section 3.4.

The 13 spatial adpositions are based on a relational noun referring to a region in space (Table 2). Note that a relational noun such as *arka* ‘back’ should not only be understood as referring to a part of the complement referent, but to a space that extends from there.

Table 2: Turkish canonical complex spatial adpositions.

stem form	meaning of noun	meaning as adposition
<i>alt</i>	‘bottom’	‘under(neath), below’
<i>ara</i>	‘space’	‘between, among’
<i>arka</i>	‘back’	‘behind’
<i>çevre</i>	‘surroundings’	‘around’
<i>dış</i>	‘exterior’	‘outside’
<i>etraf</i>	‘surroundings’	‘around’
<i>iç</i>	‘interior’	‘in(side)’
<i>karşı</i>	‘opposite side’	‘opposite’
<i>ön</i>	‘front’	‘in front of’
<i>öte</i>	‘far side’	‘beyond’
<i>peş</i>	‘back’	‘behind’, ‘following’
<i>üst</i>	‘top’	‘on top of, above on’
<i>yan</i>	‘side’	‘beside, next to’

Since the postposition shows agreement with its complement, the attached person marker alone can represent the complement (example 4).

- (4) *ön-üm-de dur-ma*
 front-POSS.1SG-LOC stand-NEG
 ‘don’t stand in front of **me**’

3.4 Outer relator variation

Spatial adpositions can combine with any of the three local cases to express direction, position, or origin, respectively, as illustrated in examples 5a–c (Çakır 2009: 96).

- (5) a. *masa-nın üst-ün-e*
 table-GEN top-POSS.3-DAT
 ‘onto the table’
- b. *masa-nın üst-ün-de*
 table-GEN top-POSS.3-LOC
 ‘on the table’
- c. *masa-nın üst-ün-den*
 table-GEN top-POSS.3-ABL
 ‘from the (top of the) table’

While spatial adpositions have a predictable morphological make-up (always possessed, marked by local cases), those expressing abstract relations are more variable. They may be possessed (47 types) or not (17 types); and their outer relator, i.e. the morpheme establishing the syntactic function within the higher syntactic unit, is from a semi-closed class of eight morphemes (semi-closed in the sense that other morphemes are imaginable but not attested in our sample). We will leave adpositions derived from verb stems to section 3.5 and consider only those with a nominal base here. Table 3 gives the type frequencies of the various morphological options, Table 4 exemplifies each category.

Table 3: Frequencies of non-spatial denominal adpositions.

outer relator	possessed	non-possessed	total
locative	14	7	21
comitative	12	3	15
dative	9	4	13
adverbializer –CA	5	1	6
ablative	5	0	5
–∅	2	0	2
derivation –ll	0	1	1
derivation –llk	0	1	1
total	47	17	64

Table 4: Examples of non-spatial denominal adpositions.

outer relator	possessed	non-possessed
locative	<i>sirasında</i> ‘at the time of’	<i>birlikte</i> ‘together with’
comitative	<i>nedeniyle</i> ‘because of’	<i>oranla</i> ‘in comparison to’
dative	<i>yerine</i> ‘instead of’	<i>yana</i> ‘in favour of’
adverbializer –CA	<i>boyunca</i> ‘throughout’	<i>sürece</i> ‘as long as’
ablative	<i>ardından</i> ‘after’	–
–Ø	<i>sonucu</i> ‘as a result of’	–
derivation – <i>il</i>	–	<i>ilgili</i> ‘with regard to’
derivation – <i>lik</i>	–	<i>yönelik</i> ‘aimed at’

The possessed nouns *sonucu* ‘its result’ > ‘as a result of’ and *gereği* ‘its necessity’ > ‘according to’ lack an outer relator in their adpositional use. An example is *anlaşma sonucu*, whose original noun-phrase meaning is ‘the result of the agreement’. This construction has been reanalysed as an adverbial adjunct meaning ‘as a result of the agreement’, turning the head noun into a postposition. Another example is *kanun gereği* ‘according to the law’. We further find zero-derivation from noun to (simple) adposition in *zaman* ‘time’ > ‘when’, and in the obsolete *inat* ‘obstinacy’ > ‘in defiance of’; both nouns are Arabic loans. With *beri* ‘since’/‘the near side’, there is no obvious directionality, and the double use is best considered an example of synchronic homonymy. Finally, the nominal uses of the adverbs *önce(sinde)* ‘before’ and *sonra(sında)* ‘after’ are examples of adverb-to-noun conversion, imitable in English as ‘in the before/after (of)’.

3.5 Deverbal adpositions

Table 10 contains seven adpositions derived from verbs. Morphologically, they can be divided into three groups. The first group consists of converb forms with the suffix *-(y)ArAk*. These converb forms express the manner in which the event described in the superordinate clause is performed. By far the most frequent adposition with this morphological make-up is *olarak* ‘as’, derived from the copula *ol* ‘be’. Example (6) has been taken from Göksel and Kerslake (2005: 217).

- (6) *Şimdi-lik bu sandığ-ı masa ol-arak kullan-ıyor=uz.*
 now-NR D1 chest-ACC table be-CONV use-PROG=1PL
 ‘For the moment we’re using this chest **as** a table.’

Example (6) shows that one of the consequences of lexicalization as adposition is the neutralization of co-reference implications. Converbial use of forms in *-(y)ArAk* would express that the subject referent of the finite verb (here ‘we are using’) is also the subject of the subordinate proposition (‘be a table’). Further adpositions in this group are *dayanarak* ‘on the basis of’ (containing the verbal stem *dayan* ‘lean on’) and *bakarak* ‘considering’ (from *bak* ‘look’).

The second group of deverbal adpositions ends in *-(y)IncA*, another converb suffix, which allows subject discontinuity. In its clause-linking use, it expresses a sequential relation. Adpositions containing this suffix are *gelince* ‘coming (to)’ > ‘as for’ and *uyarınca* ‘advising’ > ‘in accordance with’. As already mentioned in section 3.1, a potential addition to the inventory of deverbal adpositions is *göre* ‘according to’, which might have been derived from *gör* ‘see’ by adding the converb suffix *-e*.

The final group contains the adjective-forming derivational suffix *-KIn*: *ilişkin* ‘relating to’, from the root *il* ‘tie loosely’ followed by the reciprocal suffix *-İş*; and *aşkın* ‘over, surpassing’, transparently derived from *aş* ‘surpass, exceed’.

3.6 Inner relator variation

A particularity of Turkish are the case-alternating postpositions *için* ‘for’, *gibi* ‘like’, *kadar* ‘as ... as’, and *ile* ‘with’. These govern the unmarked nominative on lexical nouns, but the genitive forms of personal pronouns, demonstrative pronouns, and the interrogative pronoun *kim* ‘who’ (example 7). Note that the demonstrative pronouns are also used as third person personal pronouns (example 8). The nominative-genitive alternation also applies to the clitic *=(y)lA*; see section 3.2 for examples. The alternation is usually neutralized in favour of the nominative when the plural suffix *-lAr* is added to any of the genitive-attracting pronouns (9a), but (rare) exceptions to this “rule” are attested, as in (9b) from the TS Corpus v2. For details on the use of these postpositions, cf. Göksel and Kerslake (2005: 242–245).

- (7) a. *baba-m* *için*
 father-POSS.1SG for
 ‘for my father’
 b. *ben-im* *için*
 1SG-GEN for
 ‘for me’

- (8) *o-nun kadar güzel bir kadın*
 D3-GEN until beautiful one woman
 ‘a woman as beautiful as her’
- (9) a. *biz-ler gibi*
 1PL-PL like
 ‘like all of us’
- b. *hem biz-im için hem de on-lar-in için*
 alike 1PL-GEN for alike also D3-PL-GEN for
 ‘both for us and for them’

The nominative-genitive alternation can also be observed with the complex possessive-marked adpositions *sayesinde* ‘thanks to’, *üzerine* ‘upon; about’, *vasıtasıyla* and *aracılığıyla* ‘through (the mediation of)’, *lehin(d)e* ‘in favour of’, and *aleyhin(d)e* ‘against’. While lexical complements of these adpositions are not case-marked, demonstratives carry the genitive suffix, as exemplified in (10). The examples listed in (11) constitute the two most frequent collocations in the Wikipedia corpus.

- (10) a. *bu sistem saye-sin-de*
 D1 system shadow-POSS.3-LOC
 ‘thanks to this system’
- b. *o-nun saye-sin-de*
 D3-GEN shadow-POSS.3-LOC
 ‘thanks to him/her/that’
- (11) a. *ölüm-ü üzer-in-e*
 death-POSS.3 surface-POSS.3-DAT
 ‘upon his/her death’
- b. *bu-nun üzer-in-e*
 D1-GEN surface-POSS.3-DAT
 ‘thereupon’

Finally, the simple postposition *kadar* has two different meanings depending on the case it governs. When following the nominative (or genitive), it expresses comparison with regards to some property expressed by an adjective following the postposition, as in (12a); when governing the dative, it has a terminative meaning; see (12b).

- (12) a. *bu-gün kadar güzel bir gün*
 D1-day until beautiful one day
 ‘a day as beautiful as today’
- b. *bu-gün-e kadar*
 D1-day-DAT until
 ‘to this day’

3.7 Adpositions with verbal complements

Since Turkish rarely links finite clauses of equal status (cf. Jendraschek 2016: 246–247), subordinate clauses are typically desententialized (cf. Lehmann 1988: 193). They can be divided into those where the verb takes on the form of a converb (cf. section 3.6), and those with nominalized predicates. A consequence and symptom of nominalization is that the subject of the subordinate clause, where overtly expressed at all, is no longer shown by finite person-marking, but by possessive suffixes. The interpropositional relation to the superordinate clause is shown by case markers, or, if the semantic relation is more precise, adpositions.

Two classes of clause-linking adpositions can be distinguished, namely those which primarily have nominal (read: non-deverbal) complements and whose distribution is extended to include nominalized verbs as possible complements; and those which are found only in clause-linkage. As desententialization is a gradual phenomenon, there is no clear cut-off point between conjunction and adposition, which is why these relators are included in this overview.

Of the 117 adpositions in our corpus, 51 only occur with nominal complements, 56 occur with both nominal and deverbal complements, and 10 only occur with deverbal complements (“secondary conjunctions”). These are listed in Table 5. With the exception of *zaman* ‘time’ (see sections 3.1 and 3.4), they are all complex.

Table 5: Secondary conjunctions.

form	meaning	form of verb
<i>derecede</i>	‘to such an extent that’; ‘so much ... that’	various types of attributive clause
<i>halde</i>	‘although’	attributive clause (– <i>DIK</i>)
<i>ölçüde</i>	‘to the extent that’	various types of attributive clause
<i>oranda</i>	‘to the extent that’	mostly with –(y) <i>An</i> attributive clause

Table 5 (continued)

form	meaning	form of verb
<i>şartıyla</i>	‘on condition of’	infinitive (- <i>mA</i>) and - <i>MA</i> s/
<i>sırada</i>	‘while’	attributive clause
<i>sürece</i>	‘as long as’	attributive clause
<i>takdirde</i>	‘in the event that’	attributive clause
<i>üzere</i>	‘in order to; on the understanding that; on the verge of’	infinitive
<i>zaman</i>	‘when’	attributive clause

Secondary conjunctions (13a) may have post-nominal counterparts (13b). Both examples are from the Wikipedia corpus.

- (13) a. *Almanya-ya git-me plan-lar-ı yap-tığ-ı sıra-da*
 Germany-DAT go-NR plan-PL-POSS.3 make-NFUT.NR-POSS.3 ROW-LOC
 ‘**while** he was making plans to go to Germany’
- b. *Bu katliam sıra-sın-da bir-çok Jedi öl-dü.*
 D1 massacre row-POSS.3-LOC one-many Jedi die-PST.
 ‘Many Jedi died **during** this massacre.’

The secondary conjunction *sırada* ‘while’ follows a desententialized attributive clause, so *yaptığı sırada* in (13a) literally translates as something like ‘at the turn of his making’. It is in complementary distribution with its possessed equivalent *sırasında* ‘during’ in (13b), which follows nominal expressions in nominative case. There are exceptions to the semantic equivalence between possessed and non-possessed, as witnessed in (14). Semantic divergence can also occur if post-nominal and post-verbal use are not morphologically distinguished, as in (15).

- (14) a. *yurt-dış-ın-da bul-un-duğ-u hal-de*
 homeland-exterior-POSS.3-LOC find-REFL-NFUT.NR-POSS.3 state-LOC
 ‘**although** he was abroad’
- b. *kan-da aşırı miktar-da kolesterol ol-ma-sı hal-in-de*
 blood-LOC extreme amount-LOC cholesterol
 be-NR-POSS.3 state-POSS.3-LOC
 ‘**in case** there is an extreme amount of cholesterol in the blood’

- (15) a. *Gün-ümüz-de en çok tanın-an ve*
 day-POSS.1PL-LOC SUP much know-PASS-SBJ.NR and
en çok konuş-an-ı bul-un-an yapay dil
 SUP much speak-SBJ.NR-ACC find-REFL-SBJ.NR artificial language
ol-mak=la birlik-te ulus-lar-ara-sı iletişim
 be-INF=COM union-LOC nation-PL-space-POSS.3 communication
dil-i ol-ma amac-ın-a ulaş-ama-mış=tr.
 language-POSS.3 be-NR goal-POSS.3-DAT reach-NPOT-PFV=EPIS
 ‘**Although** it is today the artificial language that is best known and
 has the most speakers, it did not attain the objective of becoming an
 international language of communication.’
- b. *Engels=le birlik-te Brüksel-e geç-er=ler.*
 Engels=COM union-LOC Brussels-DAT pass-DISP=PL
 ‘**Together with** Engels, they move to Brussels.’

4 Diachronic and synchronic variation

4.1 Etymologies and puristic neologization

The official language of the Ottoman Empire (1299–1922) was Ottoman Turkish. While genetically a Turkic language, it was written in an Arabo-Persian script, and its lexicon was overwhelmingly made up of Arabic and Persian loans, many of which are now obsolete. The establishment of the Republic of Turkey in 1923 ushered in the Turkish Language Reform, whose most visible achievement was a new Latin-based alphabet adopted in 1928. During the following decades, many Arabic and Persian loanwords were replaced by deliberately coined neologisms. A considerable part of present-day Turkish vocabulary is the result of that reform movement, so that 21st-century Turkish and Ottoman Turkish are like mutually unintelligible languages. The language reform also affected some postpositions, where e.g. *rağmen* ‘despite’ of Arabic origin now coexists with *karşın*, based on *karşı* ‘against’ (Lewis 1999: 80). Nevertheless, not all Arabic and Persian loans have been replaced. Sometimes, they remain the only option, sometimes they coexist with an old or new Turkic form. Table 6 provides a statistic of the origin of Turkish adpositions; Turkish neologisms have been identified with the help of Lewis (1999). As the focus is on diachrony, the five semi-complex items with unclear etymology have been added to the complex adpositions.

Table 6 confirms that the lexical make-up of Turkish is still hybrid, with Turkish and Arabic adpositions constituting the two largest classes. It also shows that adpositions are a moderately conservative word class, for the native adpositions resisted the pre-20th century Arabicisation; the Arabic adpositions then resisted the 20th century purism, with identifiable neologisms only constituting 9 percent of adpositions. However, type frequency does not necessarily translate into token frequency; whereas the former indicates trends in lexicography and grammaticography, the latter reflects the usefulness for the language user and learner. In the case of synonyms, corpus frequencies are a good indication as to whether the imported adposition is still dominant or has become obsolete. Such synonyms are compared in Table 7. The “Turkic” equivalent may be a conscious attempt to replace the loan, or it may be a vague and old synonymy leading to overlapping usage. The most frequent item in the Wikipedia corpus is coded for origin, its synonyms or near-synonyms are classified as less frequent (but still common) alternatives vs. obsolete and/or marginal forms, based on absolute and relative frequencies.

Table 6: Etymological origins of adpositions.

	simple			complex			total	percentage
	N	N/V	V	N	N/V	V		
Turkish, apparently old	2	7	0	21	23	5	58	50
Turkish, neologism	0	0	0	1	10	0	11	9
Arabic	9	7	1	14	7	4	42	36
Persian	0	1	0	3	1	0	5	4
French	0	0	0	1	0	0	1	1
Total	11	15	1	40	41	9	117	100

Out of the twenty-two meanings, only four show an item of Arabic origin at the top. Table 7 thus shows more clearly than Table 6 the dominance of native adpositions. This impression is corroborated by the average frequencies and median values of adpositions of different origins, shown in Table 8. As the focus is on the synchronic relevance of different classes of adpositions, the simple adpositions include the five semi-complex items.

Both the average and the median values show the same trend; the average values are much higher because of some very frequent outliers. The simple adpositions of Turkic origin clearly dominate in terms of token frequency, followed by complex adpositions of Turkish origin. As for the role of complex adpositions in the language system, no evolutionary direction can be detected, as the Turkish

Table 7: Turkish adpositions in diachronic and synchronic variation.

translation	most frequent	alternatives	marginal or obsolete	origin of winner
1. 'about'	<i>üzerine</i>	<i>dair</i> <i>hakkında</i>		T
2. 'after'	<i>sonra</i>		<i>müteakip</i>	T
3. 'around'	<i>çevresinde</i>	<i>etrafında</i>		T
4. 'as for'; 'in respect of'	<i>açısından</i>	<i>bakımından</i>		T
5. 'because of'	<i>nedeniyle</i>	<i>sebebiyle</i>		T
6. 'before'	<i>önce</i>		<i>evvel</i>	T
7. 'by means of'	<i>yoluyla</i>	<i>suretiyle</i>		T
8. 'compared to'	<i>oranla</i>	<i>kıyasla</i> <i>nispeten</i>		T
9. 'considering'	<i>göre</i>		<i>nazaran</i> <i>bakarak</i>	T
10. 'despite'	<i>rağmen</i>	<i>karşın</i>		A
11. 'during'	<i>sırasında</i>		<i>esnasında</i>	T
12. 'for the duration of'	<i>süresince</i>		<i>devamınca</i> <i>müddetince</i>	T
13. 'from (time)'	<i>itibaren</i>	<i>beri</i>		A
14. 'on behalf of'	<i>adına</i>		<i>hesabına</i> <i>namına</i>	T
15. 'on the basis of'	<i>dayanarak</i>		<i>istinaden</i>	T
16. 'on the subject of'	<i>konusunda</i>		<i>hususunda</i>	T
17. 'outside of'	<i>dışında</i>	<i>hariç</i>		T
18. 'subsequent to'	<i>ardından</i>		<i>takiben</i>	T
19. 'through (the mediation of)'	<i>aracılığıyla</i>	<i>vasıtasıyla</i>	<i>kanalıyla</i>	T
20. 'to such an extent that'; 'so much ... that'	<i>derecede</i>	<i>ölçüde</i>	<i>oranda</i>	A
21. 'together'	<i>birlikte</i>	<i>beraber</i>		T
22. 'until'	<i>kadar</i>		<i>dek</i> <i>değün</i>	A

neologisms usually perpetuate patterns already found in Arabic complex adpositions, such as possessive suffixes, case-marking, and adverbialized forms. There is one change in the morphology, however, which is the decline of the Arabic adverbializer *-en/-an*.

Table 8: Frequencies per million words according to origin.

	average	median
Turkish, simple	1437	592
Turkish, complex	333	93
Arabic, simple	217	36
Arabic, complex	153	16
Total	399	67

4.2 Lexicalization and semantic change

While spatial adpositions can be followed by one of the three local cases, this only holds as long as the construction remains compositional. It is however common for certain postposition-case combinations to become lexicalized. The same relational noun can form different complex adpositions. Example (16) shows the four adpositional uses of the noun *yan* ‘side’.

- (16) a. *ağac-ın yan-ın-da*
 tree-GEN side-POSS.3-LOC
 ‘**next** to the tree’
- b. *demokrasi-den yan-a*
 democracy-ABL side-DAT
 ‘**in favour** of democracy’
- c. *dün-den bu yan-a*
 yesterday-ABL D1 side-DAT
 ‘**since** yesterday’
- d. *şirin ol-ma-sı bir yan-a*
 cute be-NR-POSS.3 one side-DAT
 ‘**apart** from being cute’

This process is a correlate of desemanticization, as the spatial noun adopts a metaphorical meaning. Therefore, *yan* does not always mean ‘side’ in the sense of strict laterality, but also ‘presence in the same (social) space’, as expressed in (17).

- (17) a. *baba-lar-ı-nın yan-ın-da*
 father-PL-POSS.3-GEN side-POSS.3-LOC
 ‘in their father’s presence’

(Göksel and Kerslake 2005:253)

- b. *Çocuk daha çok anne-anne-si-nin yan-in-da kal-ıyor.*
 child more much mother-mother-POSS.3-GEN side-POSS.3-LOC
 stay-PROG
 ‘The child lives mainly with his grandmother.’

(Göksel and Kerslake 2005:252)

Another example is *üzer-*, a bound stem that was originally a synonym of *üst* ‘top’, but is now mostly confined to non-spatial meanings such as ‘about (a topic)’, ‘upon (an event)’. In such non-spatial contexts, the complement has no genitive marking, as seen in (18). When the postposition contains the ablative case suffix, it means ‘via’, as in (19).

- (18) a. *uyuşturucu kullan-ım-ı üzer-in-e*
 drug use-NR-POSS.3 surface-POSS.3-DAT
 ‘on/about the use of drugs’

(Göksel and Kerslake 2005:253)

- b. *koca-sı-nın öl-üm-ü üzer-in-e*
 husband-POSS.3-GEN die-NR-POSS.3 surface-POSS.3-DAT
 ‘on her husbands death’

(Göksel and Kerslake 2005:253)

- (19) *Stuttgart üzer-in-den Münih-e git-ti-k.*
 Stuttgart surface-POSS.3-ABL Munich-DAT go-PST-1PL
 ‘We went to Munich via Stuttgart.’

(Ersen-Rasch 2004: 106)

5 Further grammatical issues

5.1 Differential case marking on the complement

In addition to the differential marking for certain classes of pronouns (the “genitive-attracting pronouns”, see section 3.6), there is differential marking on lexical complements. Here, the functional motivation is quite different. While location nouns such as *iç* ‘interior’ govern a complement in the genitive (20a) when referring to a space portion of the complement (the “interior” is part of the forest), the genitive is lacking when the location noun is used in a non-spatial “metaphorical” sense, as in (20b).

- (20) a. *orman-in iç-in-de*
 forest-GEN interior-POSS.3-LOC
 ‘in the forest’
- b. *ter iç-in-de kal-muş=ti-m*
 sweat interior-LOC stay-PFV=PST-1SG
 ‘I was covered in sweat’

(Göksel and Kerslake 2005:251)

In (20b), *iç* is not a spatial region of *ter* ‘sweat’; and in the expression *korku içinde* ‘in a state of fear’ no reference is made to a hard to imagine ‘inside’ or ‘interior’ of fear, since fear is an abstract concept without spatial dimensions. The same logic applies in (21).

- (21) *Hep baba-sı-nın etki-si alt-in-da yaşa-dı.*
 always father-POSS.3-GEN influence-POSS.3 bottom-POSS.3-LOC live-PST
 ‘He lived constantly **under** the influence of his father.’

(Göksel and Kerslake 2005:252)

However, such differential case marking does not require that the complement be an abstract entity. We can produce a contrast of spatial vs. metaphorical meaning while keeping the nature of the complement constant, as exemplified by (22). Here, it is the nature of the trajector which reveals the non-spatial relationship.

- (22) a. *Ülkü ile Ali-nin ara-sın-da Erol otur-uyor.*
 Ülkü with Ali-GEN space-POSS.3-LOC Erol sit-PROG
 ‘Erol is sitting **between** Ülkü and Ali.’
- b. *Ülkü ile Ali ara-sın-da bir sorun var.*
 Ülkü with Ali space-POSS.3-LOC one problem EXIST
 ‘There is a problem **between** Ülkü and Ali.’

(Ersen-Rasch 2004: 105)

In (22a), Erol is occupying a space located between Ülkü and Ali, whereas in (22b), *sorun* ‘problem’ has no physical existence and thus cannot occupy any physical space. While in these examples, the spatial relationship is still recoverable if taking into account the metaphorical shift, other cases of semantic change are more idiosyncratic. In (23), genitive-marking correlates with the literal meaning (23a), whereas zero-marking is associated with semantic shift (i.e. usually, bleaching). Examples such as (23b) and (24b) thus illustrate the correlation between semantic and structural attrition.

- (23) a. *Ali-nin hesab-in-a bira iç-ti-m.*
 Ali-GEN bill-POSS.3-DAT beer drink-PST-1SG
 ‘I drank beer on Ali’s charge.’
- b. *Ali hesab-in-a konuş-tu-m.*
 Ali bill-POSS.3-DAT speak-PST-1SG
 ‘I spoke on behalf of Ali.’ (Ersen-Rasch 2004: 107)
- (24) a. *Koca-m-in saye-sin-de*
 husband-POSS.1SG-GEN support-POSS.3-LOC
Türkçe-m-i geliş-tir-di-m.
 Turkish-POSS.1SG-ACC develop-CAUS-PST-1SG
 ‘With my husband’s support, I developed my Turkish.’
- b. *Koca-m saye-sin-de*
 husband-POSS.1SG support-POSS.3-LOC
Türkçe-m-i geliş-tir-di-m.
 Turkish-POSS.1SG-ACC develop-CAUS-PST-1SG
 ‘Thanks to my husband, I developed my Turkish.’ (Ersen-Rasch 2004: 107)

In the a.-examples, *hesap* and *saye* have their literal meanings of ‘bill’ and ‘support’, respectively. Therefore, the a.-versions are not examples of adpositional use, whereas the b.-versions are. This distribution is comparable to the contrast in English between (25a) and (25b).

- (25) a. *They sat in the front of the bus.*
 b. *They sat in front of the bus.*

The construction *in the front of* is not a complex adposition, whereas *in front of* is. The two have different meanings and only the latter is lexicalized; it is interesting to note that American English has also lexicalized its antonym *in back of*. What the comparison of (23) and (24) with (25) reveals is that in terms of referential status the Turkish genitive functionally corresponds to the definite article in English. The conclusion is that in such constructions the genitive fulfils the function of a determiner. That the genitive case plays the role of a determiner is corroborated by the behaviour of spatial adpositions. While the presence of the genitive suffix on the complement does not affect their status as postpositions, it marks the complement as definite. In (26a), the complement is non-referential, whereas in (26b), it is interpreted as definite.

- (26) a. *yargıç ön-ün-e*
 judge front-POSS.3-DAT
 ‘before a judge’
 b. *yargıç-ın ön-ün-e*
 judge-GEN front-POSS.3-DAT
 ‘before **the** judge’

Ersen-Rasch (2004: 105) also provides contrastive examples where genitive-marking on the complement signals a higher degree of individuation. In (27a), the month is individuated by the demonstrative *bu*; in (27b), the lower degree of individuation of *ay* ‘month’ is reflected in its coalescence with *orta* ‘middle’.

- (27) a. *Bu ay-ın orta-sın-da para-m bit-ti.*
 D1 month-GEN middle-POSS.3-LOC money-POSS.1SG end-PST
 ‘In the middle of this month, I ran out of money.’
 b. *Ay orta-sın-da para-m bit-ti.*
 month middle-POSS.3-LOC money-POSS.1SG end-PST
 ‘In the middle of the month, I ran out of money.’

5.2 Adnominalization of adpositional phrases

Turkish rarely permits the use of adpositional phrases as adnominal attributes. Compare example (28) with its English translation.

- (28) *bahçe-nin yağmur-dan önce-ki hal-i*
 garden-GEN rain-ABL before-MR state-POSS.3
 ‘the state of the garden before the rain’

(Göksel and Kerslake 2005:259)

In the English version, the adpositional phrase *before the rain* is simply juxtaposed to the nominal *state of the garden*. In the Turkish version, however, the modifier suffix *-ki* must be attached to the adpositional phrase to derive an attribute. The adnominalization of adpositional phrases is restricted to those expressing location in time, as well as relational nouns with locative case marking, as in (29).

- (29) *sendika-cı-lık tarih-i konu-sun-da-ki*
 trade.union-AN-NR history-POSS.3 subject-POSS.3-LOC-MR
araştır-ma-lar-ınız
 research-NR-PL-POSS.2PL
 ‘your research on (the subject of) the history of trade unionism’
 (Göksel and Kerslake 2005:259)

Postpositional nouns marked by other suffixes require attributive clauses to function as adnominals. In (30), the non-finite predicate *meydana gelen* ‘arising, happening’ allows the adpositional phrase *yolun kaygan olması nedeniyle* ‘due to the road’s being slippery’ to modify *kaza* ‘accident’.

- (30) *Yol-un kaygan ol-ma-sı neden-i=yle*
 way-GEN slippery be-NR-POSS.3 cause-POSS.3=COM
meydan-a gel-en kaza,
 square-DAT come-SBJ.NR accident
petrol istasyon-u-nun güvenlik kamera-lar-ınca
 petrol station-POSS.3-GEN security camera-PL-ADVR
saniye saniye kayded-il-di.
 second second record-PASS-PST
 ‘The accident, which was caused by the slippery road,
 was recorded second by second by the petrol station’s security cameras.’
 (www.cnnturk.com/turkiye/buyuk-bir-facia-kil-payi-atlatildi?page=3)

In this regard, complex adpositions behave like nouns referring to locations, when these are marked by local cases other than the locative. Of the three local cases in (31), only the locative can host the modifier suffix *-ki*. The allative and ablative, in contrast, require the insertion of verbs to be used as attributes.

- (31) a. *Ankara-da-ki ev-i-nin ön-ün-de*
 Ankara-LOC-MR house-POSS.3-GEN front-POSS.3-LOC
 ‘in front of his house in Ankara’
 b. *Ankara-ya gid-en otobüs-ler (*Ankara-ya-ki)*
 Ankara-DAT go-SBJ.NR bus-PL
 ‘the buses (going) to Ankara’
 c. *Ankara-dan gel-en otobüs-ler (*Ankara-dan-ki)*
 Ankara-ABL come-SBJ.NR bus-PL
 ‘the buses (coming) from Ankara’

The frequency study in Table 9 from the Wikipedia corpus compares different forms of the location nouns *alt* ‘under’ and *ön* ‘front’. It confirms that the locative is the most frequent case, followed by the allative, the ablative being the least frequent. The distribution of *-ki* could then be explained as a defective paradigm conditioned by markedness, the locative being the unmarked local case as diagnosed, for example, by relative frequency.

Table 9: Relative frequencies per million words of adnominalized adpositions.

case	adposition	adverbial	adnominal <i>-ki</i>
locative	<i>altında</i>	412.85	64.01
	<i>önünde</i>	93.10	12.37
allative	<i>altına</i>	146.68	0.00
	<i>önüne</i>	64.15	0.00
ablative	<i>altından</i>	17.87	0.00
	<i>önünden</i>	6.10	0.00

6 Summary

Of all the Turkish adpositions listed in our corpus, 72 percent are complex. Of these, 70 percent contain a possessive marker. The locative case is the preferred outer relator, followed by the comitative and dative. The central status of the locative is also shown by the distribution of the modifier suffix *-ki*, which can derive adnominal modifiers from locative phrases, but not from allative or ablative phrases.

Some adpositions govern the genitive on pronouns, but not on common nouns. This is similar to the distinction in English whereby oblique forms such as *me*, *him*, *her* etc. are limited to pronouns. Based on this analogy, we can conclude that the genitive in constructions such as *benim için* ‘for me’ represents the older pattern that was lost on common nouns. This is corroborated by the behaviour of complex adpositions displaying the same alternation, where the possessive suffix on the adposition should be matched by a genitive on the complement. The fact that it isn’t suggests that bare complements are an innovation that pronouns have resisted. Those constructions where the genitive and the nominative are in synchronic opposition are instances of differential case (complement, argument) marking, correlating with the referential status of the complement. It is an interesting phenomenon that Turkish has differential argument marking for both the accusative and the genitive.

While complex adpositions are primarily additions to the lexical inventory and therefore subject to lexicalization, the constructions in which they are used (and in certain cases the previously lexicalized adpositions themselves) may be added to the grammar of the language, i.e. grammaticalized. Lexicalization of adpositions can bring about semantic change from the original meaning of their components, and may lead to fossilized combinations of nouns and affixes. On the relationship between the lexicalization and grammaticalization of adpositions see Lehmann (2002: 8–16).

As nominalization is frequent in clause-linkage, adpositions are used as conjunctions. These are however a functionally separate class, as shown by cases of semantic divergence leading to a split between adpositions with nominal vs. nominalized complements. Although in terms of type frequency, 36 percent of adpositions are loans from Arabic, they account for only 16 percent of the tokens, even though certain adpositions with Arabic roots such as *tarafından* ‘by’ or *kadar* ‘until’ are among the most frequently used adpositions.

Appendix

Spatial postpositions based on relational location nouns are given in their default form with third person singular possessive marking and locative case; note that these may change. For technical reasons, frequencies are usually given for word forms and may occasionally include non-postpositional uses and coincidentally homonymous forms. Where sensible and feasible, divergent uses have been discarded by means of collocation analysis; see the notes for each item.

case:	N/G genitive only with pronouns
compl:	+ synchronically or diachronically segmentable; given as +/- for items containing an identifiable root but without independent confirmation of complex origin (e.g. <i>göre, dolayı, önce, sonra</i>)
poss:	+ contains possessive marking
spatial:	+ outer relator can be any of the local cases
denom:	– etymology either unknown or word class other than noun
vc:	+ can take desententialized verbal complements (DVC); ++ can only take DVC
etym:	T Turkish; Tn 20th century Turkish neologism; A Arabic; P Persian; F French

Table 10: Turkish adpositions and their characteristics.

form	internal structure	meaning	case governed	compl	poss	spatial	denom	vc	etym	freq. per million words	notes
1. <i>açısından</i>	angle-POSS-ABL	'in terms of'	NOM	+	+	-	+	+	Tn	137	
2. <i>adına</i>	name-POSS-DAT	'on behalf of'	NOM	+	+	-	+	+	T	147	with DVC meaning 'in order to'
3. <i>ait</i>	belonging	'belonging to'	DAT	-	-	-	-	-	A	380	
4. <i>aleyhinde</i>	opposition-POSS-LOC	'against'	N/G	+	+	-	+	-	A	10	alternating with <i>aleyhine</i>
5. <i>aleyhine</i>	opposition-POSS-DAT	'against'	N/G	+	+	-	+	-	A	16	
6. <i>altında</i>	bottom-POSS-LOC	'under'	GEN	+	+	+	+	-	T	412	
7. <i>aracılığıyla</i>	mediation-POSS-COM	'through (the mediation of)'	N/G	+	+	-	+	-	Tn	65	probably 20th-century replacement of <i>vasitasıyla</i>
8. <i>arasında</i>	space-POSS-LOC	'between'	GEN	+	+	+	+	-	T	1691	
9. <i>ardından</i>	back-POSS-ABL	'after'	GEN	+	+	-	+	+	T	603	mostly with ablative to express temporal posteriority; spatial senses supplanted by <i>arka-</i>
10. <i>arkasında</i>	back-POSS-LOC	'back'	GEN	+	+	+	+	-	T	38	
11. <i>aşkın</i>	surpass-ADJR	'over, surpassing'	ACC	+	-	-	-	-	T	48	
12. <i>bakarak</i>	look-CONV	'considering'	DAT	+	-	-	-	+	T	8	
13. <i>bakımından</i>	aspect-POSS-ABL	'in terms of'	NOM	+	+	-	+	+	T	88	with DVC meaning 'for ... ing'

form	internal structure	meaning	case governed	compl	poss	spatial	denom	vc	etym	freq. per million words	notes
21. <i>bu yana</i>	D1 SIDE-DAT	'since'	ABL	+	-	-	+	+	T	106	while this could simply mean 'to this side', a collocation analysis shows that the most frequent collocations are all postpositional
22. <i>çerçevesinde</i>	frame-POSS-LOC	'in the context of'	NOM	+	-	+	-	-	P	42	
23. <i>çevresinde</i>	surroundings-POSS-LOC	'around'	GEN	+	+	+	+	-	T	74	
24. <i>dâhil</i>	interior	'within, included in'	NOM	-	-	-	-	-	A	247	often as <i>dâhil olmak üzere</i> 'including'; frequency count includes <i>dâhil etmek</i> 'include', <i>dâhil olmak</i> 'be included' with dative complements
25. <i>dair</i>	about	'about'	DAT	-	-	-	-	-	A	119	
26. <i>dayanarak</i>	lean-CONV	'on the basis of'	DAT	+	-	-	-	+	T	26	
27. <i>değün</i>	until	'until'	DAT	+/-	-	-	-	+	T	48	alternating with <i>dek</i>
28. <i>dek</i>	until	'until'	DAT	-	-	-	-	+	T	84	
29. <i>derecede</i>	degree-LOC	'to such an extent that'; 'so much... that'	NOM	+	-	-	+	++	A	9	frequency count restricted to post-verbal use; occurs with various types of attributive clause

(continued)

Table 10 (continued)

30. <i>devamınca</i>	duration-POSS-ADVR	'for the duration of'	NOM	+	+	-	+	-	A	0
31. <i>dışında</i>	exterior-POSS-LOC	'outside'	GEN	+	+	+	+	-	T	323
32. <i>doğru</i>	straight	'towards'	DAT	-	-	+	-	-	T	420
33. <i>dolayı</i>	due	'due to'	ABL	+/-	-	-	+	+	T	308
34. <i>dolayısıyla</i>	surroundings-POSS-COM	'due to'	NOM	+	+	-	+	+	T	139
35. <i>esnasında</i>	duration-POSS-LOC	'during'	NOM	+	+	-	+	+	A	70
36. <i>etrafında</i>	surroundings-POSS-LOC	'around'	GEN	+	+	+	+	-	A	62
37. <i>evvel</i>	before	'before'	ABL	-	-	-	-	+	A	14
38. <i>gelince</i>	come-CONV	'as regards'	DAT	+	-	-	-	-	T	20
39. <i>gereği</i>	necessity-POSS	'according to'	NOM	+	+	-	+	-	T	43
40. <i>gereğince</i>	requirement-POSS-ADVR	'in accordance with'	NOM	+	+	-	+	-	T	17
41. <i>gereğinceyle</i>	justification-POSS-COM	'on grounds of'	NOM	+	+	-	+	+	Tn	41
										typically preceded by attributive clause; use with nominal complement rare

form	internal structure	meaning	case governed	compl	poss	spatial	denom	vc	etym	freq. per million words	notes
42. <i>gibi</i>	like	'like'	N/G	-	-	-	-	+	T	2054	
43. <i>göre</i>	according	'according to'	DAT	+/-	-	-	-	+	T	1272	presumably from verb stem <i>gör</i> 'see' + converb suffix <i>-e</i>
44. <i>hakkında</i>	right-POSS-LOC	'about'	NOM	+	+	-	+	-	A	288	
45. <i>halde</i>	state-LOC	'although'	NOM	+	-	-	+	++	A	46	frequency count restricted to post-verbal use; occurs with attributive clause
46. <i>halinde</i>	state-POSS-LOC	'in case of', 'in the event of'	NOM	+	+	-	+	+	A	152	the most frequent collocation is <i>gruplar halinde</i> 'in groups'; after DVC <i>-mAsI</i>
47. <i>hariç</i>	exterior	'outside of'	NOM	-	-	-	-	-	A	53	
48. <i>hesabına</i>	account-POSS-DAT	'on behalf of'	NOM	+	+	-	+	-	A	4	
49. <i>hususunda</i>	subject-POSS-LOC	'on the subject of'	NOM	+	+	-	+	+	A	4	supplanted by <i>konusunda</i>
50. <i>için</i>	for	'for'	N/G	-	-	-	-	+	T	4109	
51. <i>içinde</i>	interior-POSS-LOC	'inside'	GEN	+	+	+	+	-	T	845	
52. <i>ile</i>	COM	'with'	N/G	-	-	-	-	+	T	6149	frequency does not include cliticized use
53. <i>ilgili</i>	interest-ADJR	'with regard to'	COM	+	-	-	+	+	Tn	388	
54. <i>ilışkin</i>	tie.loosely-REC-ADJR	'relating to'	DAT	+	-	-	-	+	Tn	86	

(continued)

form	internal structure	meaning	case governed	compl	poss	spatial	denom	vc	etym	freq. per million words	notes
73. miteakip	following	'subsequent to'	ACC	-	-	-	-	-	A	6	
74. namına	name-POSS-DAT	'on behalf of'	NOM	+	+	+	+	-	P	1	supplanted by <i>adına</i>
75. nazaran	regard:ADVR	'compared to'; considering'	DAT	-	-	-	-	-	A	18	complex within Arabic
76. nedeniyle	reason-POSS-COM	'because of'	NOM	+	+	+	+	+	Tn	409	
77. nispeten	relation:ADVR	'in comparison to'	DAT	-	-	-	-	-	A	20	frequency count includes uses as adverb 'relatively'
78. nispetle	relation-COM	'in comparison to'	DAT	+	-	-	+	-	A	0	
79. olarak	be-CONV	'as' (in the role of)	NOM	+	-	-	-	-	T	6338	
80. ölçüde	measure-LOC	'to the extent that'	NOM	+	-	-	+	++	T	5	frequency count restricted to post-verbal use; occurs with various types of attributive clause
81. önce	before	'before'	ABL	+/-	-	-	-	+	T	749	presumably noun <i>ön</i> 'front' + -ce 'ADVR'
82. öncesinde	before-POSS-LOC	'ahead of'	NOM	+	+	-	+	-	T	55	rarely with genitive complement
83. önünde	front-POSS-LOC	'front'	GEN	+	+	+	+	-	T	93	
84. oranda	proportion-LOC	'to the extent that'	NOM	+	-	-	+	++	T	2	frequency count restricted to post-verbal use; occurs mostly with -(ü)An attributive clause

(continued)

Table 10 (continued)

85. oranla	proportion-COM	'in comparison to'	DAT	+	-	-	+	-	T	30	
86. ötesinde	far.side-POSS-LOC	'beyond'	GEN	+	+		+	+	T	15	with DVC meaning 'beyond ...ing'; typical verb <i>ol-</i> 'be'
87. ötürü	because	'because of'	ABL	-	-	-	-	+	T	63	
88. peşinde	back-POSS-LOC	'behind'	GEN	+	+		+	-	P	41	'behind' in the sense of 'following sb'
89. rağmen	despite	'despite'	DAT	-	-	-	-	+	A	444	
90. şartıyla	condition-POSS-COM	'on condition of'	NOM	+	+		+	++	A	8	frequency count restricted to post-verbal use; occurs with infinitive and <i>-mAs/</i>
91. sayesinde	shadow-POSS-LOC	'thanks to'	N/G	+	+		+	+	P	181	
92. sebebiyle	cause-POSS-COM	'because of'	NOM	+	+		+	+	A	108	
93. sırada	row-LOC	'while'	NOM	+	-	-	+	++	T	40	frequency count restricted to post-verbal use; occurs with attributive clause
94. sırasında	time-POSS-LOC	'at the time of'	NOM	+	+		+	+	T	519	
95. sonra	after	'after'	ABL	+/-	-	-	-	+	T	3256	presumably noun <i>son</i> 'end' + <i>-ra</i> (unknown origin and not otherwise attested)
96. sonrasında	after-POSS-LOC	'following'	NOM	+	+		+	+	T	172	
97. sonucu	result-POSS	'as a result of'	NOM	+	+		+	+	Tn	321	includes the frequent collocation <i>burun sonucu olarak</i> 'as a result of this'

form	internal structure	meaning	case governed	compl	poss	spatial	denom	vc	etym	freq. per million words	notes
98. sürece	duration-ADVR	'as long as'	NOM	+	-	-	+	++	T	15	frequency count restricted to post-verbal use; occurs with attributive clause
99. süresince	period-POSS-ADVR	'during the period of'	NOM	+	+	-	+	-	T	29	
100. suretyle	shape-POSS-COM	'by means of'	NOM	+	+	-	+	+	A	29	93% of collocations with DVC; occurs with infinitive
101. takdirde	appreciation-LOC	'in the event that'	NOM	+	-	-	+	++	A	16	frequency count restricted to post-verbal use; occurs with attributive clause
102. takiben	pursuit-ADVR	'subsequent to'	ACC	-	-	-	-	+	A	18	complex within Arabic
103. tarafından	side-POSS-ABL	'by' (an agent)	NOM	+	+	-	+	-	A	2915	
104. uğruna	luck-POSS-DAT	'for the sake of'	NOM	+	+	-	+	+	T	8	with DVC meaning 'in order to'
105. üstünde	top-POSS-LOC	'on'	GEN	+	+	+	+	-	T	59	
106. uyarınca	advise-CONV	'in accordance with'	NOM	+	-	-	-	-	T	20	
107. üzere	surface-DAT	'in order to; on the understanding that; on the verge of'	NOM	+	-	-	+	++	T	641	frequency count restricted to post-verbal use; occurs with infinitive
108. üzere	surface-POSS-DAT	'upon'; 'as soon as'; 'about'	N/G	+	+	-	+	+	T	799	

(continued)

Abbreviations

ABL	ablative	MR	modifier
ACC	accusative	NEG	negation
ADVR	adverbializer	NFUT	non-future
AN	agent noun	NPOT	impossibility
CAUS	causative	NR	nominalizer
COM	comitative	PASS	passive
CONV	converb	PFV	perfective
D	demonstrative	PL	plural
DAT	dative	POSS	possessive
DISP	dispositive aspect	PROG	progressive
EPIS	epistemic marker	PST	past
EXIST	existential predicate	REFL	reflexive
GEN	genitive	SBJ	subject
INF	infinitive	SG	singular
LOC	locative	SUP	superlative

References

- Boeschoten, Hendrik. 1998. The speakers of Turkic languages. In Lars Johanson & Éva Ágnes Csató (eds.), *The Turkic languages*, 1–15. Abingdon & New York: Routledge.
- Bundeszentrale für politische Bildung. 2019. *Statistisches Bundesamt: Mikrozensus – Bevölkerung mit Migrationshintergrund*. <http://www.bpb.de/nachschlagen/zahlen-und-fakten/soziale-situation-in-deutschland/61646/migrationshintergrund-i> (accessed 3 March 2020)
- Çakır, Hasan. 2009. *Grammatik kurz & bündig: Türkisch*. Stuttgart: Pons.
- Ersen-Rasch, Margarete. 2004. *Türkische Grammatik für Anfänger und Fortgeschrittene*. Ismaning: Max Hueber.
- Göksel, Asli, & Celia Kerslake. 2005. *Turkish. A comprehensive grammar*. London & New York: Routledge.
- Jendraschek, Gerd. 2016. The zero-marked switch-reference system of the Papuan language latmul. In Rik van Gijn & Jeremy Hammond (eds.), *Switch reference 2.0*, 231–252. Amsterdam & Philadelphia: John Benjamins.
- Kabak, Barış. 2006. An obstacle to the morphologization of postpositions. *Studies in Language* 30 (1). 33–68.
- Lehmann, Christian. 1988. Towards a typology of clause linkage. In John Haiman & Sandra A. Thompson (eds.), *Clause combining in grammar and discourse*, 181–225. Amsterdam & Philadelphia: John Benjamins.
- Lehmann, Christian. 2002. New reflections on grammaticalization and lexicalization. In Ilse Wischer & Gabriele Diewald (eds.), *New reflections on grammaticalization*, 1–18. Amsterdam & Philadelphia: John Benjamins.
- Lehmann, Christian. 2015 [1995]. *Thoughts on grammaticalization*, 3rd edn. Berlin: Language Science Press.

- Lewis, G. L. 1967. *Turkish Grammar*. Oxford and New York: Oxford University Press.
- Lewis, Geoffrey. 1999. *The Turkish language reform. A catastrophic success*. Oxford & New York: Oxford University Press.
- Robbeets, Martine. 2005. *Is Japanese Related to Korean, Tungusic, Mongolic and Turkic?* Wiesbaden: Harrassowitz.
- Robbeets, Martine. 2015. *Diachrony of Verb Morphology: Japanese and the Transeurasian Languages*. Berlin and Boston: Walter de Gruyter.
- Sezer, Bengü Sever, & Taner Sezer. 2013. TS Corpus: Herkes için Türkçe derlem [Turkish corpus for everyone]. *Proceedings 27th National Linguistics Conference*. 3–4 May 2013. Antalya, Kemer: Hacettepe University, English Linguistics Department. 217–225.
- Starostin, George. 2016. Altaic Languages. In Mark Aronoff (ed.), *Oxford Research Encyclopedia of Linguistics*. Oxford University Press. <http://linguistics.oxfordre.com/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-35>. DOI: 10.1093/acrefore/9780199384655.013.35 (accessed 3 March 2020)
- Turan, Fikret. 1998. Syntactic and semantic aspects of postpositions in Old Anatolian Turkish. *Turcica* 30. 297–308.

Corpora

- TS Corpus Wikipedia*. http://cqpweb.tscorpus.com/cqpweb/ts_wikipedia/ (accessed 3 March 2020)
- TS Corpus v2*. http://cqpweb.tscorpus.com/cqpweb/ts_corpus_ver_2/ (accessed 3 March 2020)

Dictionaries

- Büyük Türkçe Sözlük* [Grand Turkish dictionary]. http://www.tdk.gov.tr/index.php?option=com_bts (accessed 3 March 2020)
- English Turkish online dictionary Tureng*. <http://tureng.com/en/turkish-english> (accessed 3 March 2020)
- Langenscheidt Turkish-German dictionary*. <https://en.langenscheidt.com/turkish-german/> (accessed 3 March 2020)

Thomas Stolz

11 A chapter on compound prepositions in Maltese: PREP-PREP combinations and related issues

Abstract: The Afro-Asiatic language Maltese displays an intricately organised system of prepositions whose internal structure ranges from simple monosyllabic prepositions to fully fledged multiword expressions. The paper focuses on PREP-PREP combinations and similar phenomena. The inventory of PREP-PREP compounds is presented. Their functions and morphosyntactic properties are discussed on the basis of original data taken from contemporary Maltese prose. The systematic aspects of the Maltese facts are situated in the wider context of typological studies devoted to adpositions in general. Moreover, special attention is paid to the Italian impact on the Maltese system via language contact and borrowing. Additionally, diachronic aspects such as the univerbation of former sequences of prepositions and other syntagms are taken account of.

Keywords: Maltese, PREP-PREP compounds, language contact, univerbation

1 Introduction

The subject addressed in this paper concerns the formal and functional properties of certain subtypes of compound prepositions (henceforth COMPREPS) as attested in contemporary (mostly written) Maltese. The reasons for limiting this study to the presentation and discussion of only a selection of especially prominent categories and phenomena are disclosed below. The goal of this article is to demonstrate that one has to distinguish several different types of COMPREPS synchronically, some of which, however, may be connected to each other diachronically. The diachronic perspective is excluded from this study mainly for lack of data from earlier stages of Maltese. The analysis of the available texts from the late 18th and 19th century has to be relegated to a follow-up study.

The paper is organized as follows. In Section 2, the necessary background information as to the genetic affiliation and typologically most interesting structural properties of Maltese are provided. Section 3 gives insights into the theory, methodology, and empirical sources which are important for this study. The pre-

Thomas Stolz, University of Bremen, Bremen, Germany. stolz@uni-bremen.de

<https://doi.org/10.1515/9783110686647-012>

vious research on Maltese COMPREPS is reviewed in Section 4. The canonical type of Maltese COMPREPS is illustrated and discussed in Section 5. The subsections of Section 6 feature the less canonical forms of COMPREPS. The conclusions are drawn in Section 7.

2 Maltese – structural sketch

Maltese (Neo-Arabic, Semitic, Afro-Asiatic) is the native language of some 450,000 speakers. Their vast majority lives on Malta where the language not only enjoys the status of national language (co-official with English) but is also spoken natively by 98% of the population. In the course of its history, Maltese has undergone substantial lexical and grammatical Romancization via permanent contacts with Sicilian and (Tuscan) Italian including the borrowing of prepositions (henceforth: PREPS) (Saari 2009). More recently the wide-spread bilingualism and code-switching with English has begun to leave its mark especially on the language's lexicon.

In terms of its typological characteristics (Stolz 2011), Maltese is a predominantly fusional pro-drop language with nominative-accusative alignment. The basic word-order is pragmatically flexible SVO. The language is prepositional.¹ In the absence of morphological case, grammatical relations are expressed by juxtaposition (so-called construct state), word-order and/or PREPS. The language has a binary gender opposition $M \neq F$ and two productive grammatical numbers $SG \neq PL$. There is also a proclitic definite article which tends to fuse with some of the simple PREPS.

3 Background and technical matters

3.1 The basic problem

The primary motivation for investigating Maltese COMPREPS stems from the fact that there is as yet no linguistically satisfactory comprehensive account of Maltese

¹ Saari (2007) ponders the idea of *ilu* 'since' being the sole example of a postposition in Maltese. The author concludes that the function word under review is placed "somewhere on the blurred border between the categories of adpositions and adverbs." Fabri (1993: 198–201) classifies the same element among the so-called pseudo-verbs of Maltese. This classification is confirmed by Peterson (2009: 186–189). Note that the pseudo-verb stem is *il-* to which suffixal person markers are attached with *-u* encoding the 3SG.M. On account of the many open questions connected to the status of *ilu*, this function word is excluded from further discussion in this study.

PREPS, be they simple or complex.² Saari's (2003) very informative PhD-thesis notwithstanding, many basic issues still have to be settled – among them, for instance, the exact size and composition of the inventory of simple PREPS for which widely diverging proposals have been put forward (Stolz and Levkovich in press). As long as there is no general agreement as to what counts and what does not count as a simple PREP in Maltese, any attempt at doing justice to the properties of the COMPREPS of the very same language has to be taken with a grain of salt. This is why I refrain from covering the phenomenology of Maltese COMPREPS in its entirety in this study. Instead the focus is on those cases which are usually considered to be representatives of the class of COMPREPS in the extant literature on Maltese. Accordingly, of the following five subtypes of Maltese COMPREPS (a)–(e), (a) and to some extent also (d) will be paid the most attention whereas the remaining subtypes will be touched upon in ever decreasing intensity.

- a) productive binary PREP-PREP combinations as, e.g., *minn taħt* 'from under',³
- b) mandatory use of second PREP as, e.g., *permezz ta'* 'by means of',
- c) conditioned (optional or obligatory) use of second PREP as, e.g., *ġo (fi)* 'in',
- d) univerbations as, e.g., *matul* 'during',
- e) phrasal PREPS as, e.g., *mal-ġenb ta'* 'at the side of'.

Each of the subtypes deserves to be studied in-depth and extensively. Space restrictions do not however allow me to dwell on all of the interesting questions which arise in connection with (a)–(e) above. Their discussion thus has to be relegated to future dedicated studies.

3.2 Theoretical background

The above subtype (a) constitutes the class of PREP-PREP combinations. They are of general interest for the theory of COMPREPS because they pose problems as to the adequate way of analysing them. In Hagège's (2010: 58–61) very inspir-

² To save space, I mention the properties of simple PREPS only at those points where they are relevant for the discussion of the COMPREPS. There is thus no section dedicated to simple PREPS but the interesting bits about them are scattered about the main body of this paper.

³ For practicality, throughout this study, each PREP and COMPREP is usually given only one English translation equivalent based on the first meanings given in Aquilina's (1987, 1991) Maltese-English dictionary unless the meanings are quoted directly from other sources. In this case, different translations are possible. This does not mean that I deny the existence of polysemy.

(3) [Costa 1998: 13]

Sadattant, hi nizzlet landa vojta
 meanwhile 3SG.F <CAUS>go_down tin empty:F

tal-kunserva

of:DEF-preserve

minn fuq xkaffa mill-kċina.

from on shelf from:DEF-kitchen

‘In the meantime, she took an empty tin of preserve **from a shelf in the kitchen.**’

In (1)–(2), the PPs which are marked out in boldface are headed by the simple prepositions *minn* ‘from’ and *fuq* ‘on’, respectively. In contrast, (3) gives evidence of the combinability of these two prepositions as *minn fuq* ‘from on’ on the left margin of the PP. The structural analysis of the PPs headed by the simple prepositions is straightforward as shown in I–II. The situation is different with the PREP-PREP combination however, because according to Hagège’s above reasoning two different analyses are possible as reflected by the bracketed structures in III–IV.

- I. [*minn*_{PREP} [*skultur*_N [*ta’*_{PREP} [*fama*_N [*mondjali*_{ADJ}]_{NP}]_{PP}]_{NP}]_{PP} ‘by a sculptor of global fame’
- II. [*fuq*_{PREP} [*il*_{DEF}-*Kappillan*_N]_{NP}]_{PP} ‘on the parish priest’
- III. [*minn*_{PREP} [*fuq*_{PREP} [*xkaffa*_N [*mil*_{PREP} [*l*_{DEF}-*kċina*_N]_{NP}]_{PP}]_{NP}]_{PP}]_{PP}
- IV. [*minn fuq*_{PREP} [*xkaffa*_N [*mil*_{PREP} [*l*_{DEF}-*kċina*_N]_{NP}]_{PP}]_{NP}]_{PP} ‘from a shelf in the kitchen’ (the same for III and IV)

To determine which of the two analyses is adequate, it needs to be investigated whether or not there is a syntactic border dividing the PREP-PREP combination in two hierarchically distinct units. The problem vanishes if the cases are looked at from the point of view of Construction Grammar provided no hierarchically organized internal structure of constructions is assumed. Accordingly, a possible formula on the basis of which (3) could be construed is V:

- V. [*minn*_{PREP} [*fuq*_{PREP} NP]_{PP}]

The anonymous reviewers argue that different models might be applicable to different sequences. In this study, however, no decision is taken as to the appropriate choice of model.

3.3 Methodology and empirical foundations

My line of argumentation relies on the critical appraisal of the hypotheses about COMPREPS as suggested in the descriptive-linguistic literature on Maltese (to be tackled in Section 4). Since the previous accounts of the phenomenon under review belong to different linguistic schools of thought (traditional Semitic philology, structuralism, localism, minimalism, etc.), I adopt a relatively theory-neutral surface-oriented approach which is indebted to linguistic functionalism in the most general reading of the term.

The documentation of the empirical facts is severely limited in most of the linguistic work dedicated to Maltese PREPS and COMPREPS. Moreover, there are striking discrepancies not only between the hypotheses of different scholars but also within the work of one and the same author especially as to the number of COMPREPS and their classification according to internal criteria. To remedy these shortcomings, it is necessary to verify the extant hypotheses on the basis of a corpus which consists of a variety of original Maltese novels authored by different modern writers. I also make use of the electronic *Korpus Malti* (= KM) which has not been available to my predecessors.

The methodology is primarily qualitative. The phenomena are studied from a synchronic perspective. Evidence of synchronic variation and some very general ideas about possible diachronic developments are discussed unsystematically.

4 Previous accounts

The existence of COMPREPS in Maltese has been acknowledged already early on in the history of the descriptive linguistics of the language when Vassalli (1827: 33–37) wrote a paragraph about compound particles (“particelle composte”) which hosted among many other things also several COMPREPS. Since then however, in many grammars of Maltese, the issue of COMPREPS has not been deemed interesting enough to be mentioned at all.⁶

⁶ This is the case, for instance, with the otherwise recommendable grammars by Panzavecchia (1845), Sutcliffe (1936), and Brother Henry (1980). It especially strikes the eye that Procházka’s (1993: 250–256) selective survey of COMPREPS (“komposite Präpositionen”) in the neo-Arabic varieties does not include a single example from Maltese.

In the 1930's, Cremona (1970: 81)⁷ observed that a subset of the Maltese PREPS has a binary structure.⁸ The author referred to cases like *mingħajr* 'without', *għalfejn* 'why, whither', *sakemm* 'until', *safejn* 'up to where'. Except the first example, none of these words can be classified as PREP though because they function exclusively as conjunctions, adverbs, or relative-interrogative morphemes.

Aquilina (1973: 300)⁹ mentions seven examples of "the occurrence of two prepositions together." All of his examples involve initial *minn* 'from' as shown in (4).¹⁰

(4) Aquilina's examples of PREP-PREP combinations

- a) *minn bejnhom* 'from between them', b) *minn fuq* 'from over', c) *minn taht* 'from underneath', d) *mingħand* 'from with', e) *mingħala* 'in the opinion of', f) *minn qabel* 'beforehand', g) *mbagħad* 'afterwards' (< *minn bagħad*)

In these cases, there is fusion plus semantic bleaching. Examples f)–g) are problematic insofar as g) is an adverb or discourse particle that never fulfils prepositional functions whereas f) can be employed as COMPREP but not with the meaning given in my source which illustrates only the adverbial function of *minn qabel*.

The first use of the term COMPREPS ("komplexe Präpositionen") for the phenomenon under scrutiny is made by Fabri (1993: 187) who, like his predecessor Aquilina, illustrates the category exclusively with the pattern *minn* + X_{PREP} as shown in Table 1. Note that the lists of Aquilina's and Fabri's are not absolutely commensurate. The meaning of the COMPREPS (re-translated from German) is given in the cells below the second PREPS.

Fabri (1993: 187) is also the first to assume that the PREPS involved in a COMPREP fulfill a function jointly, namely that of describing a path. In addition, these COMPREPS can only combine with motion verbs ("komplexe Wegpräpositionen [...], die nur mit Bewegungsverben auftreten dürfen"). They thus constitute a functional unit. This unit-like character is further corroborated by the fact that only the second PREP can be inflected if there is a pronominal complement as in *minn quddiem-u* 'from in front of **him**' with the exponent of the 3SG.M -u being affixed exclusively to the second PREP (Fabri 1993: 187).

⁷ Quoted from the unaltered second edition.

⁸ Quote: "[b]iċċa minnhom huma magħulin minn żewġ kelmiet", my translation: 'some of them [= the prepositions] are made of two words.'

⁹ Originally published in 1959.

¹⁰ Aquilina's original IPA-based transcription of Maltese examples has been transferred to current Maltese orthography. The English translations are Aquilina's.

Table 1: Complex prepositions according to Fabri (1993).

Initial PREP								
<i>minn</i>								
<i>bejn</i>	<i>fuq</i>	<i>ġo</i>	<i>għand</i>	<i>ħdejn</i>	<i>ma'</i>	<i>quddiem</i>	<i>taħt</i>	<i>wara</i>
'from between'	'from on'	'through'	'from s.b.'s place'	'from beside'	'along'	'from in front'	'from under'	'from behind'
Second PREP								

Cachia (1994: 131) observes that some PREP-PREP combinations are motivated structurally in the sense that those PREPS, which cannot host inflections for pronominal complements call for the employment of an additional PREP which can be inflected pronominally. This is the case with *ġo* 'in' which requires the presence of the synonymous PREP *fi* 'in' if the complement is pronominal as in *ġo fi-h* 'in him' with *-h* representing the 3SG.M (cf. subtype (c) above). Therefore, not all PREP-PREP combinations are motivated semantically. Moreover, Cachia (1994: 131) also shows that there are candidates for the status of COMPREPS which involve an initial PREP other than *minn* 'from' as shown in (5).¹¹

- (5) Cachia's examples of PREP-PREP combinations
- initial *barra* 'outside': *barra minn* 'except',
 - initial *fuq* 'on': *fuq ġewwa ta'* 'on the inside of', *fuq barra ta'* 'on the outside of',
 - initial *ġewwa* 'inside': *ġewwa fi* 'inside',
 - initial *għal* 'for': *għal fuq* 'unto', *għal taħt* '(to) under', *għal ġewwa* 'into',
 - initial *minn* 'from': *minn fejn* 'from where', *minn ħdejn* 'from beside', *minn għand* 'from s.b.'s place', *minn ġo* 'through', *minn fuq* 'from on', *minn taħt* 'from under',
 - initial *sa* 'until': *sa fi* 'as far down as', *sa ġo fi* 'as far into as'.

In contrast to the prior statements about COMPREPS in Maltese, Cachia also lists combinations which involve more elements than just two PREPS (as in (5c) and (5f)).

For none of the authors mentioned in the foregoing paragraphs can it be assumed that it was their intention to exhaust the subject of COMPREPS. It is much more likely that they tacitly presuppose that their readers know that there

¹¹ Cachia (1994: 131) speaks of "prepożizzjonijiet li jhobbu jmorru flimkien", i.e. prepositions which frequently collocate (literally: 'love to go together' [my translation]).

are more cases so that they can limit themselves to presenting the most interesting or typical cases. This is different with the two subsequent texts because one is the internationally best known reference grammar of Maltese (Borg and Azzopardi-Alexander 1997) and the other the only book-length study dedicated to Maltese PREPS (Saari 2003) so far.

In accordance with the format of the LDS-grammars, Borg and Azzopardi-Alexander (1997: 297) have a section on COMPREPS. The authors identify three PREPS – *ta* ‘of’, *għal* ‘for’, *minn* ‘from’ – which “can be combined with (and precede)” other PREPS “expressing different spatial relations.” The second position within the binary combinations can be occupied by eight to ten different PREPS. The matrix of these combinations is given in Table 2. The symbol ✓ means that a given combination is attested. Empty cells are additionally marked by grey shading. The English translations are original.

Table 2: Complex prepositions according to Borg and Azzopardi-Alexander (1997).

second PREP	initial PREP		
	<i>minn</i> ‘from’	<i>għal</i> ‘for’	<i>ta</i> ‘of’
<i>barra</i> ‘out(side)’	✓	✓	✓
<i>bejn</i> ‘between’	✓	✓	✓
<i>fuq</i> ‘on, above’	✓	✓	✓
<i>ġo</i> ‘in(side)’	✓	✓	✓
<i>ħdejn</i> ‘near’	✓	✓	✓
<i>quddiem</i> ‘in front of’	✓	✓	✓
<i>taħt</i> ‘under, beneath’	✓	✓	✓
<i>wara</i> ‘behind’	✓	✓	✓
<i>għand</i> ‘at s.b.’s place’	✓	✓	
<i>qalb</i> ‘among’	✓		

In the detailed section on the expression of nonlocal and local semantic functions (Borg and Azzopardi-Alexander 1997: 147–172), there are further potential cases of COMPREPS which like *minn biswit* ‘from beside’ are excluded from Table 2 for no apparent reason. More importantly, both Borg and Azzopardi-Alexander (1997) and Fabri (1993) emphasize the spatial functions which are fulfilled by COMPREPS. They also concur as to the basically binary format of the COMPREPS. From both studies it results that PREP-PREP combinations with initial *minn* ‘from’ are particularly frequent in terms of types. On the other hand, Borg and Azzopardi-Alexander (1997) provide proof of the possibility that a PREP other than

minn can occupy the initial position. It strikes the eye however that their set of possible initial PREPS only partly overlaps with that of Cachia's given in (5).

Saari (2003: 84–100) discusses Maltese COMPREPS at length and identifies exactly eighty different members of this class with nine different candidates for the initial position although those with initial *minn* still constitute the largest subgroup (Saari 2003: 87–88). Saari's list of COMPREPS includes also a number of cases which involve more than two syntactic words. Table 3 contains only the 49 binary combinations which are identified as COMPREPS by Saari (2003: 95).¹²

Table 3: Complex prepositions according to Saari (2003).

	second position	initial position							sum	
		<i>minn</i>	<i>sa</i>	<i>għal</i>	<i>fi</i>	<i>qrib</i>	<i>barra</i>	<i>ġo</i>		<i>ma'</i>
		'from'	'until'	'for'	'in'	'near'	'out(side)'	'in(side)'		'with'
<i>ħdejn</i>	'near'	✓	✓	✓	✓					4
<i>barra</i>	'out(side)'	✓	✓	✓						3
<i>fuq</i>	'on, above'	✓	✓	✓						3
<i>ġo</i>	'in(side)'	✓	✓	✓						3
<i>taħt</i>	'under'	✓	✓	✓						3
<i>qalb</i>	'among'	✓			✓			✓		3
<i>għand</i>	'at s.b.'s place'	✓		✓						2
<i>ma'</i>	'with'	✓		✓						2
<i>qabel</i>	'before'	✓	✓							2
<i>qrib</i>	'near'	✓	✓							2
<i>quddiem</i>	'in front of'	✓	✓							2
<i>wara</i>	'behind'	✓	✓							2
<i>waqt</i>	'during'	✓			✓					2
<i>tul</i>	'along'	✓							✓	2
<i>bejn</i>	'between'	✓								1
<i>għajr</i>	'except'	✓								1
<i>għal</i>	'for'	✓								1
<i>fi</i>	'in'		✓				✓	✓		3
<i>minn</i>	'from'		✓			✓	✓			3
<i>kontra</i>	'against'			✓						1

¹² For expository reasons, Saari's original table has been modified to better illustrate which of the PREPS are more frequently involved in (putative) COMPREPS than others. Sum and total have been added for convenience.

Table 3 (continued)

<i>għeluq</i>	'end'				✓					1
<i>lejn</i>	'towards'					✓				1
<i>ta'</i>	'of'					✓				1
<i>dwar</i>	'about'								✓	1
total		17	11	8	4	3	2	2	2	49

According to Table 3, no PREP is compatible with the entire set of other PREPS to yield PREP-PREP combinations. Except *sa* 'until', all of the PREPS which occupy the initial position may also fill the second slot of the combination. It is worth noting too that with *kontra* 'against' Saari (2003) admits a PREP of Italian origin in his inventory whereas his predecessors have exclusively looked at combinations of elements which have a Semitic background. The synopsis of binary combinations provided by Saari (2003: 95) allows the reader to identify four PREPS which seem to be incompatible with the second position in PREP-PREP combinations. These PREPS are given in (6). Three of them are of Italian extraction. In contrast to *sa* 'until', they are depicted as also failing to qualify for the initial position of PREP-PREP combinations.

- (6) PREPS which are excluded from PREP-PREP combinations according to Saari (2003)
- favur* 'in favor of' (< Italian *a / in favore di* 'in favor of'), *rigward* 'as regards to' (< Italian *riguardo a* 'as regards to'), *sa* 'until', *skont* (~ <*skond*>) 'according to' (< Italian *secondo* 'according to')

Saari's (2003) discussion of COMPREPS has the merit of approaching the subject matter without any preconceived ideas so that it is unproblematic for the author to look beyond the domain of spatial relations on the functional side of the COMPREPS. He therefore identifies causal, modal, spatial, and other functions which are fulfilled by COMPREPS as is the case with the comparative PREP *bhal* 'like'. At the same time, Saari's (2003) account also raises a number of problems which suggest that the issue under review has not yet been settled completely.

A problem that comes to mind is the exclusion of the Italian PREPS in (6) from the initial position of PREP-PREP combinations although the author explicitly mentions that they are accompanied by the second PREP *ta'* 'of' under certain conditions (Saari 2003: 87). Since *ta'* is given as a possible second PREP in Table 3, the reason for omitting the Italian PREPS in the list of initial PREPS remains unclear. Similarly, the combination *minn ta'* is listed once (Saari 2003: 87) but

not entered in Table 3. The other way round, Table 3 allows for the combination *sa barra* which is missing however from the earlier inventory (Saari 2003: 88). A significant difference between Saari's (2003) account and that of Borg and Azzopardi-Alexander (1997) is that the latter postulates that *ta* 'of' can be in the initial position of PREP-PREP combinations whereas this possibility is not mentioned in Saari (2003). In contrast to Cachia (1994), Saari (2003) does not accept *għewwa* 'inside' as a PREP. On the other hand, Saari (2003) considers the noun *għeluq* 'conclusion' to be a PREP (in second position) although this item is mentioned nowhere else in the pertinent literature on Maltese PREPS – and so forth. On top of that, Saari (2003) lumps together synchronically transparent cases and those which require historical philological expertise to be analysed as diachronic instances of complex prepositions (cf. Subsections 5.3 and 6.1). This practice is avoided by most of the authors who have approached the subject of Maltese COMPREPS in the 20th–21st centuries.

Furthermore, it is relatively easy to fill many of the gaps that characterize the inventories of COMPREPS discussed in this section. In Appendix I, I provide sentential examples of several PREP-PREP combinations which are absent from the above studies but are attested in the Maltese texts used for this study. The initial PREPS *minn* 'from', *għal* 'for', and *sa* 'until' have been selected for this purpose because of their prominence in Table 3. It strikes the eye that with *fost* 'among', *madwar* 'around', *matul* 'during', and *maġenb* 'at the side of', there are four additional candidates for the second position in PREP-PREP combinations which are mentioned in none of the previous studies and which, at the same time, belong to Saari's (2003: 87–88) class of COMPREPS themselves. Possibly their COMPREP-status is somehow responsible for their absence from the previous inventories which tacitly look only at PREP-PREP combinations which involve exclusively simple PREPS. The data in Appendix I are indicative however of the possibility that PREP-PREP combinations may also involve COMPREPS as fillers of the two basic slots.

Admittedly, without the searchable KM many gaps would have remained empty because certain PREP-PREP combinations do not occur frequently. In contrast, several of the PREP-PREP combinations mentioned in Table 3 yielded no hits on the KM (such as *fi hdejn* and *f'qalb*). What is important to keep in mind is that the gaps of the previous inventories are not always there for structural or semantic reasons but often only because of the low frequency with which a given PREP-PREP combination is attested. The random check of the corpus data is revealing in the sense that PREP-PREP combinations can be shown to constitute a largely open and productive pattern with probably many more types than previously assumed.

There are thus many loose ends which need to be tied together before the chapter on Maltese COMPREPS can be closed for good, in a manner of speaking. As a first step in this direction, it makes sense to first narrow down the scope of

the investigation to those phenomena about which there is agreement among the experts. This can be achieved by way of defining the canonical type of Maltese COMPREP.

5 From productive PREP-PREP combination patterns to univerbations

5.1 The canon

Saari (2003: 87–88) distinguishes ten different classes of COMPREPS starting from nowadays opaque historical cases via different kinds of transparent PREP-PREP combinations and univerbations to phrasal constructions. To some extent, the order in which these classes come is reminiscent of a continuum based crucially on the criteria of (formal) opacity vs transparency in combination with that of complexity. In my opinion, there is an alternative way of imposing order on the seemingly heterogeneous phenomenology of COMPREPS. The point of departure for the comprehensive inquiry into the grammar of Maltese COMPREPS is the identification of the canonical COMPREP in this language. To this end I adopt the approach of canonical typology as exposed in Corbett (2005).

The canon can be defined on the basis of those assumptions about COMPREPS which are explicitly or tacitly shared by all scholars involved in the discussion. This is what the above approaches have in common. All of them at least

- accept binary combinations of PREPS as (one of the) instances of COMPREPS,
- involve combinations of synchronically productive simple PREPS,
- feature COMPREPS which fulfil spatial functions
- assume that the fillers of the two prepositional slots constitute (more or less sizable) closed classes with that of the initial position being smaller than that of the second position,
- depict the bulk of the COMPREPS as close-knit units which can neither be interrupted by intercalations of any kind nor change the linear order of their component parts.

To these properties, I add the general requirement that canonical COMPREPS are involved in PPs which have complements (Hagège 2010: 8). According to Hagège (2010: 58), the cross-linguistically most common types of complements are NPs, be they pronominal or nominal. If all these criteria are met, we are facing an instance of the canon.

5.2 PREP-PREP combinations as COMPREPS

The template for PREP-PREP combinations which meet the criteria of the above canon comprises three slots with the rightmost position being reserved for the complement. Given that the complement is an NP, we can focus on the fillers of the two prepositional slots. As a matter of fact, with reference to the canon, it comes to the fore that the replacement of the fillers of these slots yields relatively extended paradigms. The initial PREP and the second PREP interact with each other systematically on the level of semantics. This is what proves that there is a close bond between the two PREPS and that they are indispensable component parts of one and the same construction.

The template has a clear internal structure which I interpret along the lines of Lestrade (2010: 65–134), who distinguishes three major functional components of spatial situations the expression of which often reflects a division of labour among, for instance, the constituents of PPs. The complement serves the purpose of representing the GROUND whereas the accompanying elements in the PP take on the job of expressing configuration and directionality (Lestrade 2010: 118–126). Directionality is the cover term for the basic distinction of the spatial relations: static PLACE (e.g. AT) vs dynamic GOAL (e.g. TO), SOURCE (e.g. FROM), and PATH (e.g. THROUGH/ACROSS). Configuration, on the other hand, refers to the region (e.g. INSIDE, OUTSIDE, SURFACE) relative to the GROUND which is involved in a given spatial situation. In Maltese PREP-PREP combinations, the different functions are distributed systematically over the template as shown in Table 4.

Table 4: Template of PREP-PREP combinations.

initial	second	complement
PREP _{SPATIAL}	PREP _{SPATIAL}	NP
DIRECTIONALITY	CONFIGURATION	GROUND
PLACE/GOAL/SOURCE	REGION	

The division of labour and its paradigmaticity can be gathered from the confrontation of

- (a) the examples (7)–(9) which illustrate replacement of the fillers in the initial slot while the second PREP remains the same, and
- (b) the examples (10)–(12) which attest to different fillers in the second position in combination with an identical initial PREP.

In (7)–(9), the initial slots are variously filled by one of the three PREPS which are reported to be involved most frequently in PREP-PREP combinations according to Table 3. The second position is invariably occupied by *fuq* ‘on’. The GROUND-NP is always definite and refers to a concrete object.

- (7) *minn* ‘from’ [Briffa 1992: 49]

Kien żelaq minn fuq il-blata
 be.PERF slip **from on DEF-rock**
 ‘He had slipped **from the rocks** [...].’

- (8) *sa* ‘until’ [Cauchi 1999: 62]

jitiłghu sa fuq il-gverta.
 3IMPERF:climb:PL **until on DEF-deck**
 ‘[...] they climb **on deck** [...].’

- (9) *għal* ‘for’ [Cauchi 1999: 38]

imma ż-żewġ ġuvnotti qabżu
 but DEF-two youngster:PL jump:3PL.PERF
għal fuq il-bankina
for on DEF-pavement
 ‘[...] but the two youngsters jumped **onto the pavement** [...].’

In terms of the configuration, *fuq* ‘on’ indicates that the surface of the GROUND is affected by the motion event. The changing PREPS in the initial position however identify the directionality of the event as [+dynamic]. *Minn* ‘from’ expresses SOURCE whereas both *sa* ‘until’ (terminative) and *għal* ‘for’ (allative) specify different categories which can be subsumed under GOAL. The paradigm which results from the replacement of the initial PREPS is captured by Figure 1.



Figure 1: Paradigmaticity in the initial position.

In (10)–(12), it is the second position which is subject to changes whereas the initial PREP *minn* ‘from’ is kept constant. The GROUND-NP is inanimate in (10)–(11) but a human being in (12).

- (10) *qalb* ‘among’ [Briffa 1992: 11]
Mar-raba’ darba Superman deher ġie
 with:DEF-fourth time Superman appear come
minn qalb is-shab.
from among DEF-cloud.COLL
 ‘On the fourth try, Superman appeared **from among the clouds.**’
- (11) *ġo* ‘in’ [Cauchi 1999: 27]
s-Sur Ġilard kien qala’ l-batterija
 DEF-Mr Ġilard be.PERF remove DEF-battery
minn ġo-r-remote control
from in-DEF-remote control
 ‘[...] Mr Ġilard had removed the battery **from the remote control** [...].’
- (12) *ħdejn* ‘beside’ [Cauchi 1999: 6]
Is-Sur Ġilard telaq dritt minn ħdejn ix-xwejjah
 DEF-Mr Ġilard leave direct **from beside DEF-old_man**
 ‘Mr Ġilard went directly **away from the old man** [...].’

In all three of the examples, the directionality is realized as SOURCE since the PREP remains the same throughout.¹³ On the side of configuration however, the different fillers of the second position refer to different regions of the GROUND as shown in Figure 2.

$$\text{minn}_{\text{DIRECTIONALITY}} + \left\{ \begin{array}{l} \text{qalb}_{\text{AMONG}} \\ \text{ġo}_{\text{INSIDE}} \\ \text{ħdejn}_{\text{SIDE}} \end{array} \right\} + \text{NP}_{\text{GROUND}}$$

Figure 2: Paradigmaticity in the second position.

If the directionality is [+static] and thus realizes the spatial relation of PLACE, it seems that the leftmost slot of the template may remain empty¹⁴ or – put differ-

¹³ *Minn* is a particularly colorful PREP in terms of its range of meanings/functions. The basic ablative function competes with perelative, prolative, and comparative functions as well as that of marking the passive agent, etc. To keep the discussion within reasonable bounds, in this study only the ablative function will be taken account of.

¹⁴ There is the construction [*ta'* PREPSPATIAL NP_{GROUND}] which frequently invites a [+static] reading for directionality such as in [Costa 1998: 67] *Dahħlu fil-maqjel ta' wara wara l-bitha*

ently – a simple PREP expressing configuration alone may be sufficient. However, Maltese is characterized by PLACE-GOAL syncretism, meaning: simple PREPS are often ambiguous as to their spatial semantics. This problem is illustrated in Table 5 with the PREPS employed in second position in Figure 2.

Table 5: PLACE-GOAL syncretism of simple PREPS.

PREP	relation	example
<i>qalb</i>	PLACE	[Bartolo and Vella 2009: 75] <i>Fi triqithom iltaqgħu ma' ragħaj xiħ li kien miexi qalb merħla żgħira ta' nagħaġ.</i> 'On their way they met with an old shepherd who walked among a small herd of sheep. '
	GOAL	[Bartolo and Vella 2009: 82] [...] <i>daħlu qalb il-friegħi u l-weraq tal-ħarruba.</i> '[...] they went right among the branches and leaves of the carob tree. '
<i>ġo</i>	PLACE	[Bartolo and Vella 2009: 408] <i>Ĵda f'daqqa waħda ħass bħal tisbit kbir ġo rasu.</i> 'But suddenly he felt something like a big banging in his head. '
	GOAL	[Bartolo and Vella 2009: 368] <i>Sadanittant ħi daħlet idħa ġo borża tal-karti kannella u ħarġet xi ħaġa minnħa.</i> 'Meanwhile she put her hand into the brown paper bag and extracted something from it.'
<i>ħdejn</i>	PLACE	[Bartolo and Vella 2009: 210] <i>Ĥi u ħierġa lemħet xi ħaġa mal-art, ħdejn is-sodda ta' Antida.</i> 'While she was leaving she saw something on the ground beside Antida's bed. '
	GOAL	[Bartolo and Vella 2009: 40] <i>Dawn marru ħdejn il-mejda tal-Konti minnufiħ.</i> 'Those went immediately to the Count's bed. '

It is only logical that PREP-PREP combinations tend to be semantically more precise than simple PREPS because one of the slots can be used to disambiguate any potentially competing interpretations.

In conformity to Fabri (1993: 187), I consider the above PREP-PREP combinations to be functionally very closely interconnected binary units, meaning they are transparent morphologically and compositional semantically. Initial and second PREP together express the meaning which is crucial for understanding the spatial situation as such. This is what supports their interpretation as COMPREPS of Type IV or V as pictured in Subsection 3.2 above. What speaks against considering them instances of the alternative Type III is that the evidence in favour of a separation of the two PREPS on account of a syntactic border is not absolutely compelling. First of all, since the fillers of both slots

[...]. 'He made him enter the sty **behind the yard** [...].' However, the functions of the almost ubiquitous (genitival) PREP *ta'* 'of' are so manifold that its role in the system of COMPREPS can be determined only after a thorough check of the empirical facts, which is a task of a follow-up study.

(15) [KM news134523]

niżguraw li t-turisti jibbenefikaw
 1.IMPERF:ensure:PL SUBORD DEF-tourist:PL 3.IMPERF:benefit:PL
minn trasferiment effiċjenti, professjonali u kordjali
 from transfer efficient professional and cordial
minn u għal fuq il-vapur tagħhom
from and for on DEF-ferry of:3PL

‘We ensure that the tourists benefit from an efficient, professional and cordial transfer **from and to their ferry.**’

On account of the heterogeneity of their behaviour, the proper classification of PREP-PREP combinations in Maltese cannot be declared settled once and for all yet. Future research will have to determine the exact status of the phenomenon under review.

5.3 Univerbated erstwhile PREP-PREP combinations

One might argue further that none of the cases presented in this section is properly lexicalized, i.e. the meaning of the PREP-PREP combinations is largely compositional and transparent. This is a strong counter-argument against their being COMPREPS in the first place. Thus, it deserves to be taken into account.

Maltese boasts a number of PREPS which can be shown to result from univerbations of former syntactic neighbours. There are two different subtypes of univerbations, namely those which reflect an erstwhile PREP-PREP combination as opposed to those which go back to a PP whose head and complement NP have coalesced to give rise to a new lexeme. In this subsection, univerbated PREP-PREP combinations are discussed whereas the second subtype of univerbations is briefly addressed in Subsection 6.1. What the two subtypes have in common is that the formerly compositional semantics sometimes (but by no means always) becomes opaque.

The most prominent cases of univerbated PREP-PREP combinations involve either initial *ma* ‘with’ (Stolz, Stroh, and Urdze 2006: 294–304) or initial *minn* ‘from’ as shown in the following list (for formations with *tul*, cf. Subsection 6.1):

- *ma* ‘with’ + *dwar* ‘about’ → *madwar* ‘around’
- *ma* ‘with’ + *tul* ‘during’ → *matul* ‘during’
- *minn* ‘from’ + *barra* ‘outside’ → *minbarra* ‘except’
- *minn* ‘from’ + *għajr* ‘except’ → *mingħajr* ‘without’
- *minn* ‘from’ + *għal* ‘for’ → *mingħal* ‘in s.o.’s opinion’

- *minn* ‘from’ + *għand* ‘at s.o.’s place’ → *mingħand* ‘from s.o.’
- *minn* ‘from’ + *tul* ‘during’ → *mintul* ‘throughout’

The evidence of their univerbation is primarily orthographic, meaning: the two PREPS are written as one orthographic word. Secondly, the semantics of the univerbation is no longer compositional in a straightforward way.

As to the orthographic side of the phenomenon, there is considerable variation across the literary corpus, some of which is explicable with reference to the vicissitudes of the discussion about the appropriate orthographic norm for Maltese. Especially in earlier texts, several of the above univerbations tend to be written separately as a sequence of two orthographic words. The examples (16)–(18) are meant to give an impression of the range of variation and at the same time prove how far removed the univerbations can be semantically from the simple PREPS on which they are based.

- (16) [Bartolo and Vella 2009: 133]

Imma din qatt ma
 but DEM.F.PROX never NEG
stennejtha mingħand-ha.
 expect:1SG.PERF:3SG.F **from-3SG.F**
 ‘But I never expected this **of her.**’

- (17) [Costa 1998: 66]

L-isqof, Ġann, jisthoqqlu
 DEF-bishop Ġann 3SG.M.IMPERF:REFL:merit:IO:3SG.M
l-akbar rispett minn
 DEF-big.COMP respect **from**
għand-na.
at_s.o.’s_place-1PL
 ‘The bishop, Ġann, deserves the highest respect **from us.**’

- (18) [Costa 1998: 99]

Infatti, il-flus m’għand Salvu s-Sifja hadhom
 in_fact DEF-money.PL **from Salvu s-Sifja** take.PERF:3PL
 ‘In fact, he took the money **from Salvu s-Sifja** [...].’

There are three different ways of representing the item orthographically. In (16), the two PREPS *minn* ‘from’ and *għand* ‘at s.o.’s place’ are treated as one word whereas in (17) they come as a succession of two distinct words. Interestingly,

sentence (18) stems from the same source as (17). However, in contrast to the separation of *minn* ‘from’ and *għand* ‘at s.o.’s place’ in the latter, example (18) attests to an advanced stage of fusion of the two PREPS because *minn* is reduced to its initial consonant /m/. This segmental erosion goes along with a moderate semantic dissociation of *mingħand* ‘from s.o.’ from *għand* ‘at s.o.’s place’. In the univerbated case, the reference to someone’s home usually associated with *għand* is not compulsory. In the examples (16)–(18), the place of residence of the referent of the complement is not at issue at all. In all of these instances, the participant is [+human] and is directly involved as SOURCE in the situation. Whether the event is taking place at the participant’s lodgings is of no avail. This is different with *għand* and *għal għand* in (19)–(20) the readings of which always involve the home of the person mentioned in the complement.

(19) [Bartolo and Vella 2009: 110]

kienu jafu li omm Roberto
 be:3PL.PERF 3.IMPERF:KNOW:PL SUBORD mother Roberto
kwazi dejjem għand oħtha Rose
 almost always at_s.o.’s_place sister:3SG.F Rose
tkun.

3SG.F:be.FUT

‘[...] they knew that Roberto’s mother would be almost always **at her sister Rose’s (place).**’

(20) [Casha 1996: 12]

telqet tghaġġel
 leave:3SG.F.PERF 3SG.F.IMPERF:hasten
għal għand oħtha Pina
 for at_s.o.’a_place sister:3SG.F Pina

‘[...] she left hurriedly **for her sister Pina’s (place)** [...].’

Mingħand ‘from s.o.’ is thus – if not already completely lexicalized – on the verge of lexicalization. Mutatis mutandis, this diagnosis also holds for the other cases of univerbation of former PREP-PREP combinations. I therefore conclude that the attested processes of lexicalization of erstwhile PREP-PREP combinations lend support to the idea that we are dealing with COMPREPS, be they genuine or still in the making.

6 Beyond the canon

In this section, I sketch very briefly further kinds of complexity in the domain of PREPS in Maltese without aspiring at exhausting the subject matter.

6.1 Phrasal COMPREPS and the other kind of univerbation

Saari (2003: 87–88) mentions thirteen (supposedly) phrasal COMPREPS which basically reflect the construction pattern [PREP (DEF-)N (*ta'*)]. In Table 6, these phrasal COMPREPS are listed in accordance to the construction formula.

Table 6: Phrasal COMPREPS according to Saari (2003).

PREP	DEF	N	<i>ta'</i>	meaning
<i>ma-</i> 'with'	<i>n-</i>	<i>naħa</i> 'side'	(<i>ta'</i>) 'of'	'at the side of'
<i>mi-</i> [<i>< minn</i>] 'from'	<i>n-</i>	<i>naħa</i>	(<i>ta'</i>)	'from the side of'
(<i>fi-</i>) 'in'	(<i>n-</i>)	<i>naħa</i>	(<i>ta'</i>)	'at the side of'
<i>fi-</i>	<i>l-</i>	<i>ħin</i> 'time'	<i>ta'</i>	'at the time of'
<i>ma-</i>	<i>l-</i>	<i>ħin</i>	<i>ta'</i>	'at the time of'
<i>fi-</i>	<i>l-</i>	<i>waqt</i> 'moment'	(<i>ta'</i>)	'during'
(<i>fi-</i>)	(<i>n-</i>)	<i>nofs</i> 'middle'	(<i>ta'</i>)	'in the middle of'
<i>fi-</i>	<i>r-</i>	<i>rigward</i> 'regard'	<i>ta'</i>	'as to'
<i>għa-</i> [<i>< għal</i>] 'for'	<i>l-</i>	<i>favur</i> 'favour'	<i>ta'</i>	'in favour of'
<i>għa-</i> [<i>< għal</i>]	<i>l-</i>	<i>ħabta</i> 'collision'	<i>ta'</i>	'at the time of'
<i>ma-</i>	<i>l-</i>	<i>bidu</i> 'beginning'	<i>ta'</i>	'at the beginning of'
<i>ma-</i>	<i>t-</i>	<i>tmiem</i> 'end'	<i>ta'</i>	'at the end of'
<i>ma-</i>	<i>l-</i>	<i>ġenb</i> 'side'	<i>ta'</i>	'at the side of'

As the brackets indicate, several of the phrasal COMPREPS involve optional constituents. Only four different PREPS occupy the initial slot of the constructions in Table 6. With ten types, the number of nouns which are used in these patterns is also relatively limited. The range of meanings expressed by the phrasal COMPREPS shows that temporal concepts and the spatial relation of location at the side of a Ground are especially prominent.

However, for several of these candidates it remains to be demonstrated by empirical fact that they have advanced sufficiently in the process of grammaticalization. As it seems, the decategorialization of their constituents is still

in its incipient phase. Borg and Azzopardi-Alexander (1997: 163–164) show, for instance, that the common noun *naħa* ‘side’ can be further modified e.g. by the adjectival attribute *l-oħra* ‘the other’ just like the definite article can be accompanied by the proximal demonstrative *din* ‘this’, etc. The possibility to manipulate the pattern speaks against classifying it as a genuine COMPREP on a par with the cases discussed in Section 5.

This is not to exclude phrasal structures from being involved in the formation of COMPREPS in Maltese. In point of fact, there are again several cases of univerbation which prove that erstwhile PPs have been lexicalized as unanalyzable units with prepositional functions. A case in point is the comparative PREP *bħal* ‘like’ < Arabic *bi* ‘with’ + *ħāl* ‘condition, way’ (Aquilina 1987: 112), for which it can be assumed that its historical compositionality is opaque to the vast majority of contemporary speakers of Maltese, not the least because the cognate Maltese noun *ħal* is obsolete now and is said to occur exclusively in a certain proverb (Aquilina 1987: 482). Much the same is true of the original noun in:

- *biswit* ‘facing’ < Arabic *bi* ‘with’ + *swit* ‘opposite side’. Aquilina (1987: 126) states that **swit* “is never used alone.”
- *fost* ‘among’ < Arabic *fī* ‘in’ + *wast* ‘middle’ (Aquilina 1987: 358) for which Aquilina (1991: 1521) states that the cognate Maltese noun *wast* is “hardly ever used alone but it occurs in *fost* ‘amongst’.”

It is doubtful that Maltese speakers recognize that *bla* ‘without’ is (historically) dimorphemic as it goes back to the univerbation of Arabic *bi* ‘with’ + *lā* ‘not’ (Aquilina 1987: 128–129) although both of the components are still (highly) frequent in contemporary Maltese, viz. the simple multipurpose PREP *bi* ‘with’ and *la* with a variety of functions in the domain of negation (Aquilina 1987: 718).

A higher degree of transparency may be assumed for

- *flok* ‘instead of’ < Maltese *fī* ‘in’ + *lok* ‘place’ (also *minflok* ‘instead of’) (Aquilina 1987: 350),
- *maġenb* ‘beside’ < Maltese *ma* ‘with’ + *ġenb* ‘side’ (Aquilina 1987: 386),
- *minkejja* ‘in spite of’ < Maltese *minn* ‘from’ + *nkejja* ‘vexation’ (Aquilina 1991: 900–901).

A completely different story is that of *minħabba* ‘on account of’ which superficially seems to involve *minn* ‘from’. However, there is also the alloform *mħabba* ‘on account of’ which is identical with the noun *mħabba* ‘love’ (Aquilina 1987: 466–467). This suggests that the form *minħabba* results from a reanalysis of *mħabba*, motivated by the many PREP-PREP combinations with initial *minn* ‘from’ in Maltese. The erstwhile common noun had already been grammaticalized as simple PREP

to be reanalyzed according to the familiar pattern of PREP-PREP combinations afterwards.

Slightly different are the cases which involve *tul* ‘during’ as second component (cf. Subsection 5.3). There is the noun *tul* ‘length’ in Maltese. It is tempting to assume that, in analogy to *maġenb* ‘beside’, the COMPREPS *matul* ‘during’ and *mintul* ‘throughout’ are univerbations of erstwhile PPs with a complement noun. In this article however, I defend the hypothesis that both *matul* and *mintul* are based on PREP-PREP combinations since *tul* itself has been grammaticalized from noun to PREP. This PREP-function of *tul* is illustrated in (21).

- (21) [Bartolo and Vella 2009: 71]
Ħadd minnhom ma kien qal kelma
 Nobody from:3PL NEG be.PERF say.PERF word
tul il-vjaġġ kollu.
during DEF-journey all:3SG.M
 ‘None of them had uttered a word **during the entire journey.**’

In (21), *tul* can be replaced with *matul* and vice versa. It cannot be ruled out however that *tul* as PREP is a backformation via aphaeresis from *matul*.

6.2 Obligatory vs optional second PREP

In this subsection, I mention summarily further phenomena which connect to the issue under scrutiny. All of those have to do with the presence of a second PREP.

The second PREP *ta’* ‘of’ is mandatory with the Italian loan-PREP *permezz* (< Italian *per mezzo di* ‘by means of’). Example (22) illustrates its use.

- (22) [Camilleri 2013: 63]
il-pożizzjoni eżatta taġhom misjuba
 DEF-position exact:F of:3PL PASSPCPL:find:F
permezz ta’ metal detector
by means of metal detector
 ‘[...] their exact position (was) found out **by means of a metal detector** [...].’

For other PREPS of Italian origin, the presence of *ta’* is largely optional (Saari 2009). In Table 7, the variation is illustrated with the loan-PREP *kontra* ‘against’.

This free variation is blocked for a number of Semitic PREPS in Maltese which require the presence of a second PREP if the complement is pronominal (see 24).

Table 7: Presence and absence of *ta'* with *kontra* 'against'.

N/PRO	<i>ta'</i>	Example
N	yes	[KM Parl10595] <i>Dan jiġri mhux għax ta' barra huma kontra ta' Malta [...].</i> 'This happens not because the foreign states are against Malta [...].'
	no	[KM Academic399] [...] <i>sbatax -il Malti ġew akkużati bi tradiment kontra Malta waqt il-gwerra.</i> '[...] seventeen Maltese were accused of treason against Malta during the war.'
PRO	yes	[KM Academic236] <i>Beethoven beda jaħseb li kien hemm konfossa kontra tiegħu.</i> 'Beethoven began to think that there was a conspiracy against him .'
	no	[KM Sport1488] <i>Qatt ma kelli xejn kontrih personali [...].</i> 'I never had anything personal against him [...].'

In some of these cases, the second PREP is semantically related to the initial PREP (Stolz and Levkovych in press).

(23) [Bartolo and Vella 2009: 319]

Huwa kien ilu kważi tlett ijiem maqful
 he be.PERF since:3SG.M almost three day.PL PASSPCPL:lock
go ċella mudlama u umduża.
in cell PASSPCPL:darken:F and humid:F
 'For almost three days he was locked up **in a dark and humid cell**.'

(24) [Briffa 1992: 18]

Is-sodda taparsi fortizza.
 DEF-bed as_if fortress
Aħna t-tnejn biss fadal go fi-ha.
 we DEF-two only remain **in in-3SG.F**
 'The bed is as if a fortress. We two alone remain **in it**.'

The complement noun in (23) combines directly with the PREP *go* 'in' whereas the pronominal complement in (24) calls for the use of the second PREP *fi* 'in' because the latter is able to inflect for person. Similar cases which require the presence of a second PREP with pronominal complements are

- *barra* 'outside' + *minn* 'from'
- *minhabba* 'on account of' + *fi* 'in'
- *minkejja* 'in spite of' + *fi* 'in'
- *sa* 'until' + *fejn* 'near'

In a way, the above PREPS are part-time COMPREPS because the presence of the second PREP is dependent upon the type of complement. Only if the complement is pronominal does a PREP-PREP combination arise. In contrast to the PREP-PREP combinations discussed in Subsection 5.2, there is no combinatory semantics, i.e. the second PREP does not contribute anything new to the meaning of the construction. The second PREP is there for purely formal reasons in the sense that it serves as host of the pronominal complement which cannot directly be attached to the initial PREP. Since the choice of the second PREP is not free but dictated by the initial PREP, cases like that in (24) have been analysed in terms of a government relation: the initial PREP governs a PP if the complement is pronominal (Stolz and Levkovych in press) – an interpretation which is in line with type III in Subsection 3.2.

7 Conclusions

The presentation of the Maltese evidence of COMPREPS in the foregoing sections strongly suggests that the situation is characterized by heterogeneity. The definition of the canon in Subsection 5.2 notwithstanding, it is hardly possible to claim that there is only one candidate for the status of COMPREP in Maltese. On the contrary, several distinct types of COMPREPS emerge, the members of which share different subsets of properties with the canonical COMPREP. The diachronic connections between PREP-PREP combinations or erstwhile PPs and certain univerbations have been highlighted above. The diachronic developments are suggestive of gradual transitions from one class to the other. These transitions, in turn, presuppose that the domain of COMPREPS is organized like a continuum but not exclusively along the lines of the polar distinction transparent vs opaque. Univerbations result from processes which affect two different categories, namely PREP-PREP combinations and PPs. It is these two sources which occupy the opposing extremes of the continuum with univerbations being situated in the continuum's centre. The goal of the processes is the creation of new complex prepositional lexemes (Saari 2002: 273), which ultimately may be reanalysed as monomorphemic (i.e. unanalysable) units and thus enrich the inventory of simple PREPS.

The results are largely preliminary. Future investigations will have to reveal the systematics of Maltese COMPREPS in full. For the time being, the lesson to be learned is that the study of Maltese COMPREPS raises many questions which call for being studied in-depth to the benefit of the general theory of COMPREPS.

Acknowledgments: Sandra Abela kindly shared her native-speaker knowledge of Maltese with me. For their indefatigable technical support, I am grateful to Nataliya Levkovych and Beke Seefried. Sandro Caruana made important reading matter available to me. Slavomír Čepeló kindly provided me with bibliographical information. I also express my gratitude to the editors and the anonymous reviewers. The content and form of this paper is however exclusively my own responsibility.

Abbreviations

1/2/3	1st/2nd/3rd person	M	masculine
ADJ	adjective	N	noun
CAUS	causative	NEG	negation
COLL	collective	NP	noun phrase
COMP	comparative	OBJ	object marker
COMPREP	compound preposition	PASSPCPL	passive participle
DEF	definite article	PERF	perfective
DEM	demonstrative	PL	plural
DIR	directional	PP	prepositional phrase
DIS	distal	PREP	preposition
F	feminine	PRON	pronoun
FUT	future	PROX	proximal
IMPERF	imperfective	REFL	reflexive
IO	indirect object	SG	singular
KM	Korpus Malti	SUBORD	subordinator

Primary sources

- Bartolo, Simon & Vella, Lorraine. 2009. *Il-ġnien tad-dmugħ. Il-fiddien: it-tielet ktieb*. Blata l-Bajda: Merlin.
- Briffa, Charles. 1992. *Teżor fl-imħabba*. Valletta: Bugelli.
- Camilleri, Mark. 2013. *Volens*. Blata l-Bajda: Merlin.
- Casha, Charles. 1996. *Fra Modest. Stejjer umoristiċi oriġinali miktuba għat-tfal*. Il-Belt: Union Press.
- Cauchi, Carmel G. 1999. *Dak l-imbieriek Mastru Gerefex*. Blata l-Bajda: Merlin.
- Costa, Tusè. 1998. *Il-ġimgha mqaddsa tal-kappillan*. San Ġwann: PEG.
- Korpus Malti = <http://mlrs.research.um.edu.mt/>

References

- Aquilina, Joseph. 1973. *The Structure of Maltese*. Malta: The University of Malta Press.
- Aquilina, Joseph. 1987. *Maltese-English Dictionary. Volume One: A–L*. Malta: Midsea.
- Aquilina, Joseph. 1991. *Maltese-English Dictionary. Volume Two: M–Z and Addenda*. Malta: Midsea.
- Borg, Albert & Marie Azzopardi-Alexander. 1997. *Maltese*. London: Routledge.
- Brother, Henry. 1980. *Grammatika Maltija* [Maltese Grammar]. Gżira (Stella Maris College): De La Salle Brothers Publications.
- Cachia, Lawrenz. 1994. *Grammatika ġdida tal-Malti* [New Maltese Grammar]. Żabbar: Veritas Press.
- Corbett, Greville G. 2005. The canonical approach in typology. In Zygmunt Frajzyngier Adam Hodges & David S. Rood (eds.), *Linguistic Diversity and Linguistic Theories* (Studies in Language Companion Series 72), 25–50. Amsterdam & Philadelphia: Benjamins.
- Cremona, Antonio. 1970. *Tagħlim Fuq il-Kitba Maltija* [Instruction about the Maltese orthography]. Malta.
- Fabri, Ray. 1993. *Kongruenz und die Grammatik des Maltesischen* (Linguistische Arbeiten 292). Tübingen: Niemeyer.
- Hagège, Claude. 2010. *Adpositions* (Oxford Linguistics). Oxford: Oxford University Press.
- Lestrade, Sander. 2010. *The Space of Case*. Radboud Universiteit te Nijmegen Dissertation.
- Panzavecchia, Fortunato. 1845. *Grammatica della lingua maltese*. Malta: Weiss.
- Peterson, John. 2009. “Pseudo-verbs”: an analysis of non-verbal (co-)predication in Maltese. In Bernard Comrie et al. (eds.), *Introducing Maltese linguistics*. Amsterdam & Philadelphia: Benjamins, 181–204.
- Procházka, Stephan. 1993. *Die Präpositionen in den neuarabischen Dialekten* (Dissertationen der Universität Wien 238). Wien: Verband der wissenschaftlichen Gesellschaften Österreichs.
- Saari, Rami. 2002. *Quddiem* and some remarks on grammatical aspects of Maltese prepositions. In Susanne Feigenbaum & Dennis Kurzon (eds.), *Prepositions in their Syntactic, Semantic and Pragmatic Context* (Typological Studies in Language 50), 269–282. Amsterdam, Philadelphia: Benjamins.
- Saari, Rami. 2003. *Milot hayaxas hamalteziyot* [The Maltese prepositions]. Jerusalem: Carmel.
- Saari, Rami. 2007. Maltese *ilu* – The only postposition in a language of prepositions? In Tali Bar & Eran Cohen (eds.), *Studies in Semitic and General Linguistics in Honor of Gideon Goldenberg* (Alter Orient und Altes Testament 334), 75–88. Münster: Ugarit-Verlag.
- Saari, Rami. 2009. Some remarks on Maltese prepositions of Italian origin. In Gideon Goldenberg & Ariel Shisha-Halevi (eds.), *Egyptian, Semitic and General Grammar: Studies in the Memory of H. J. Polotsky* (Sources and Studies in Jewish Languages), 270–276. Jerusalem: The Israel Academy of Sciences and Humanities.
- Stolz, Thomas. 2011. Maltese. In Bernd Kortmann & Johan Van der Auwera (eds.), *The languages and linguistics of Europe. A comprehensive guide*. Berlin & Boston: De Gruyter Mouton, 241–256.
- Stolz, Thomas & Ahrens, Carolin. 2017. On prepositional ellipsis and the factors which block its application in Maltese. In Benjamin Saade & Mauro Tosco (eds.), *Advances in Maltese Linguistics* (Studia Typologica 21), 121–147. Berlin & Boston: De Gruyter Mouton.

- Stolz, Thomas & Lestrade, Sander & Stolz, Christel. 2014. *The Crosslinguistics of Zero-Marking of Spatial Relations* (Studia Typologica 15). Berlin & Boston: De Gruyter Mouton.
- Stolz, Thomas & Levkovich, Nataliya & Urdze, Aina. 2017. When zero is just enough... In support of a *Special Toponymic Grammar* in Maltese. *Folia Linguistica* 51 (2). 453–482.
- Stolz, Thomas & Levkovich, Nataliya. In press. From variation towards the grammar of Maltese prepositions – first steps. In Slavomír Čepelák & Jaroslav Drobný (eds.), *Maltese linguistics on the Danube*. Berlin & Boston: De Gruyter Mouton.
- Stolz, Thomas, Cornelia Stroh & Aina Urdze. 2006. *On Comitative and Related Categories. A Typological Study with Special Focus on the Languages of Europe* (= Empirical Approaches to Language Typology 33). Berlin & New York: Mouton de Gruyter.
- Sutcliffe, Edmund F. 1936. *A Grammar of the Maltese Language with Chrestomathy and Vocabulary*. London: Humphrey Milford.
- Vassalli, Michele Antonio. 1827. *Grammatica della lingua maltese*. Malta: Stampata per l'autore.

Appendix I: Additional PREP-PREP combinations

Table A: Initial *minn*.

initial	second	Example
<i>minn</i>	<i>biswit</i>	[KM literature80] [...] <i>35 ajruplan Taljan kienu resqin lejna, minn biswit il-Port il-Kbir</i> [...]. ‘[...] 35 Italian airplanes were approaching us, from beside the Grand Harbor [...].’
	<i>fost</i>	[Cauchi 1999: 84] <i>U minn fost il-kljenti qam applaws għal Mastru Gerefex</i> [...]. ‘And among the customers there arose applause for Mastru Gerefex [...].’
	<i>madwar</i>	[Camilleri 2013: 135] [...] <i>xi hadd ilmenta minn madwar il-mejda f’nofs il-Murder Conference Room temporanja</i> . ‘[...] someone at the table in the middle of the temporary Murder Conference Room complained.’
	<i>maġenb</i>	[Camilleri 2013: 257] <i>Minn maġenbha kien għadu għaddej inħir storbjuż</i> . ‘ On her side some heavy snoring was still going on.’

Table B: Initial *sa*.

initial	second	Example
<i>sa</i>	<i>biswit</i>	[KM literature77] [...] <i>il-għarajjes telgħu flimkien sa biswit l-altar il-kbir</i> [...]. ‘[...] bride and bridegroom together went up to beside the main altar [...].’
	<i>għand</i>	[Cauchi 1999: 4] <i>Għalhekk l-għada, qabel mar il-Każin, għadda s’għand Menzju ta’ l-Inbid, xtara borża tabakk mit-tajjeb</i> [...]. ‘Therefore the next day, before he went to the band-club, he passed by Menzu the wine merchant , to buy a bag of the best tobacco [...].’
	<i>lejn</i>	[KM news83623] [...] <i>wrietni pupa Maltija li tmur lura sa lejn l-aħħar tas-seklu tmintax</i> [...]. ‘[...] she showed me a Maltese doll which goes back to the end of the 18th century [...].’
	<i>maġenb</i>	[KM literature79] [...] <i>iwassluhom sa maġenb il-kaless</i> [...]. ‘[...] they take them to [the side of] the coach [...].’
	<i>matul</i>	[KM parl1494] [...] <i>ix-xogħol jinsab fi stadju avanzat u fil-fatt għandu jitlesta sa matul il-lejl li ġej</i> [...]. ‘[...] the work is in an advanced state and indeed it has to be ready until the coming night [...].’
	<i>waqt</i>	[KM news119730] [...] <i>ħafna nisa ma jkunux konxji minnhom sa waqt u wara t-tqala</i> . ‘[...] many women aren’t aware of them until during and after the pregnancy .’

Table C: Initial *għal*.

initial	second	Example
<i>għal</i>	<i>bejn</i>	[Briffa 1992: 70] <i>Malli Lina waslet tiġri bl-iskejz quddiemha, Marisa mbuttat l-iskejt-bord għal bejn saqajha.</i> ‘When Lina came skating in front of her, Marisa pushed the skate-board between her legs. ’
	<i>biswit</i>	[KM literature77] [...] <i>mexa mbagħad mat-triq għal biswit id-dar ta’ Lippu [...].</i> ‘[...] then he went along the road to beside Lippu’s house [...]. ’
	<i>maġenb</i>	[KM literature77] [...] <i>u medd riġlu sa maġenb il-mejda fejn kien qiegħed Wenzu [...].</i> ‘[...] and he stretched his legs as far as beside the table where Wenzu was sitting [...].’
	<i>matul</i>	[Casha 1996: 64] <i>Jekk trid ġib xi basket li fih tista’ tpoġġi xi bżonnijiet għal matul il-vjaġġ.</i> ‘If you want fetch a basket in which you can put some necessary things for [during] the voyage. ’
	<i>qabel</i>	[KM news107340] <i>Il-każ imur lura għal qabel l-Elezzjoni Ġenerali [...].</i> ‘The case goes back to before the General Elections [...]. ’
	<i>qrib</i>	[KM news75280] [...] <i>tikkoordina s-salvataġġ f’wisa ta’ baħar li twassal minn ħdejn Lampedusa għal qrib il-ġżira ta’ Kreta.</i> ‘[...] she coordinates the rescuing operations on the open sea which reaches from near Lampedusa to close to the island of Crete. ’
	<i>quddiem</i>	[KM literature77] [...] <i>u kuljum għalhekk il-qaddejja kienu jgħorru n-nies għal quddiem il-Ħakem u l-imħallfin.</i> ‘[...] and every day therefore the servants brought the people before the Ruler and the judges. ’
	<i>tul</i>	[KM european1435] [...] <i>għandhom jieħdu abbord sitt baħrin minn São Tomé u Príncipe għal tul l-istaġun tas-sajd.</i> ‘[...] they have to take on board six sailors from São Tomé and Príncipe for [the time of] the fishing season. ’
	<i>waqt</i>	[Casha 1996: 45] <i>Imbagħad xi ngħidu għal waqt it-tqarbin.</i> ‘What do we say then on occasion of the baptismal ceremony. ’ [lit. for during the baptism]
	<i>wara</i>	[KM literature5] <i>U mingħajr ma qagħad jaħsibha darbejn qabad is-senter u ġera għal wara ċint.</i> ‘And without thinking twice he grabbed the gun and ran behind a fence. ’



Discussion

Benjamin Fagard, Elena Smirnova, Dejan Stosic, and
José Pinto de Lima

12 Complex adpositions in Europe and beyond: A synthesis

Abstract: In the chapters of this volume, the authors describe the paradigms of complex nominal relators (CRs for short) in various languages of Europe, both Indo-European (Romance, Germanic, Slavic, Celtic, Greek and Albanian) and non-Indo-European (Estonian and Permic (Finno-Ugric), Basque (isolate), Turkish and Maltese (Afro-Asiatic)). In this final chapter, we propose to go beyond individual descriptions, and discuss the results presented in all chapters from a wider typological perspective.

Keywords: adpositions, case, complex prepositions, nominal relators, typology, constructionalization

1 The rise of complex nominal relators – The diachrony of CRs in European languages

In the introductory chapter, we formulated two research questions related to the diachrony of complex adpositions. In what follows, we will take up these questions and elaborate on them, taking into account the situations in the individual languages and language families dealt with in the chapters of this volume.

One of the aims of this volume was to focus more closely on the EMERGENCE AND EVOLUTION of different types of CAs in European languages by gaining more empirical evidence from different languages and on different types of expressions. Our intention was thus to bring together contributions from a varied array of European languages, so that a good picture of the (presumably) complex phe-

Benjamin Fagard, CNRS, ENS & Paris Sorbonne Nouvelle; PSL, Montrouge, France. benjamin.fagard@ens.psl.eu

Elena Smirnova, Institut de langue et littérature allemandes, Université de Neuchâtel, Neuchâtel. elena.smirnova@unine.ch

Dejan Stosic, University of Toulouse Jean Jaurès & Laboratory CLLE (CNRS & UT2), Toulouse, France. dejan.stosic@univ-tlse2.fr

José Pinto de Lima, Universidade de Lisboa, Faculdade de Letras, Lisboa, Portugal. pintodelima@outlook.pt

<https://doi.org/10.1515/9783110686647-013>

nomena leading to the emergence and the evolution of CAs ensues. As mentioned in the introduction, the background has already been laid out in the literature (Lindqvist 1994, Schwenter and Traugott 1995, Di Meola 2000, Cifuentes Honrubia 2003, Hoffmann 2005, Kurzon and Adler 2008, Hagège 2010, Libert 2013, among others). This volume has widened the object under analysis from the traditional one of simple prepositions, widely studied, to the more embracing one of complex adpositions, including other complex nominal relators, especially gaining the advantage of contemplating studies on complex postpositions and combinations of adpositions and case marking. With regard to this goal, the volume adds substantially to the existing body of research on complex prepositions.

First, as concerns the so far most studied syntactic pattern of **complex prepositions** [P N P], with its less frequent variant with the genitive marking on the complement noun phrase [P N CASE], the chapters on *West Germanic* and on *Romance* languages look in more detail at the developmental paths of a series of individual prepositions, as well as at the development of the schematic pattern itself. They show that, on one hand, there is a great amount of variance between the individual adpositions with respect to frequency, to the degree of formal fixedness and to the possibility of modification and omission of individual elements within them. On the other hand, the schematic pattern at a more general level of abstraction seems to be more stable, in that it gains in productivity and attracts an ever larger paradigm of individual adpositions. The chapters on *Celtic* and *Greek*, on the other hand, represent a first account of the inventory of Celtic and Greek CAs following this pattern. Most importantly, the observations from these two languages confirm the general developmental regularities which have been taken into account and described for other languages, such as Germanic and Romance. For example, the diachronic processes of unit formation and reanalysis are particularly relevant for the development of complex prepositions of this type, which are sometimes followed by and may thus be considered a prerequisite for the univerbation (in a broad sense) of the erstwhile free syntactic combinations with compositional semantics. On the way from free syntactic combination to complex adposition, syntagmatic fixedness increases, which leads to the formation and entrenchment of a syntagmatic string. During this diachronic process, the individual elements of the syntagm successively become less independent, gradually coalescing into a holistic and idiosyncratic unit (= unit formation). This loss of autonomy of individual elements is often (and quite naturally) accompanied by decategorialization and followed by univerbation. A good example of unit formation followed by univerbation and complete loss of internal structure is the Celtic preposition *abalamour da* ‘because of’, which originated in (Medieval or Classical) French *par l’amour de* ‘out of love for’ and is often pronounced [blam] in spoken Breton, its etymology thus completely obscured. What we observe here

are more or less usual reduction processes commonly found in the processes of grammaticalization and lexicalization.

As concerns other formation patterns of complex **prepositions**, the individual chapters have revealed a great variety of syntactic and morphological patterns, which, however, display a high degree of similarity at a more general level. For example, the adverbial pattern [ADV P] is discussed prominently in chapters on *Greek* (e.g. *andi ja* ‘instead of’, *mazi me* ‘together with’, *mesa s(e)* ‘in; inside’, *ekso apo* ‘out of; outside’ etc.), but is also mentioned in the chapters on West Germanic and Romance languages, where its productivity seems to be weaker and has faded away in the last centuries, most likely due to the spread of the currently dominant [P N P] pattern mentioned above¹; in Serbian, it is the third most frequent pattern of formation (with 24 CPs formed). *Maltese*, on the other hand, displays a structure which serves as a main productive pattern to form complex adpositions in this language, the combination of two prepositions [P P], e.g. *barra minn* ‘except’, *għal fuq* ‘unto’, *minn fuq* ‘from on’, *minn taht* ‘from under’. Remarkably, the patterns [ADV P] and [P P], documented in different languages, behave very much alike in that they form a continuum from free and compositional combinations to non-analyzable holistic (often unverbated) units, suggesting that they follow the same general diachronic tendencies just described for the pattern [P N P]. To give an example, *Maltese* has a number of unverbated [P P] combinations either with the initial *ma* ‘with’ or initial *minn* ‘from’, which obviously originate in free and compositional syntactic combinations but now function as non-analyzable linguistic units (see the chapter on *Maltese* for details): *madwar* ‘around’, *matul* ‘during’, *minbarra* ‘except’, *mingħajr* ‘without’, *mintul* ‘throughout’. In other words, we find once again the stable development from erstwhile free syntactic combinations to more fixed elements – with the recurrent question of where to draw the line: for instance, in the chapter on Serbian, such elements have been considered as non-canonical simple adpositions.

As concerns complex **postpositions**, the volume shows that several languages have this type of expressions at their disposal. Whether a language has pre- or postpositions in the domain of CAs generally follows from the basic structural characteristics of the language in question. If a language predominantly makes use of prepositions in the domain of simple relators, it will most likely have prepositions in the domain of complex relators. Of course, this rule is not without exceptions, and languages may have several different types of adpositions, like

1 The situation in Romance is not homogeneous, however, since Ibero-Romance and, to a lesser extent, Italo-Romance still have a productive paradigm of [Adv P] sequences.

e.g. German, where pre-, post- and circumpositions are attested in both subcategories of nominal relators. Nevertheless, the “normal” situation is that there is one preferred and prototypical type of adpositions in a language, and both, simple and complex relators, show an identical or similar distribution in this respect.

Among the languages described in this volume, there are complex postpositions in *Estonian*, *Turkish*, *Basque* and *Permic*. They are of different but closely related structural types. It is in the domain of complex postpositions that the distinction between **syntactic** and **morphological complexity** becomes interesting. In the case of complex prepositions, which we have focused on above, the diachronic origin is usually a syntactic structure, namely a free syntactic combination with compositional semantics, which in the course of time becomes more and more formally fixed and semantically non-compositional. For most complex **post**positions, however, the original structure appears to be a morphologically complex structure. In *Permic*, for example, the basic morphological structure of the complex postposition consists of a relational noun with case marking [N_{rel}-CASE] (1).

- (1) keñer **dor-âš** (Beserman Udmurt)
 fence near-EL
 ‘not far from the fence’ (Usacheva and Arkhangel'skiy, this volume,
 example (16))

This structure functions as a postposition and assigns case to the preceding complement noun phrase. In the course of the development, the relational noun in the nucleus of the complex postposition gradually loses its ability to inflect – an instance of decategorialization –, and often ends up as a fixed expression which cannot be further modified.

In *Basque*, the default or canonical construction of a complex postposition is schematically the same as in *Permic*: [N_{rel}-CASE], e.g. *menpe-an* = domination-LOC ‘under the authority’, where the case of the preceding nominal complement is governed by the morphologically complex adposition, whose nucleus is the relational noun (2) (see chapter on Basque for more details).

- (2) *zuzendari-a-ren menpe-an* (Basque)
 director-DET-GEN domination-LOC
 ‘under the authority of the director’ (Jendraschek, this volume, example (22a))

Diachronically, this construction has undergone reduction processes in several ways, most notably by losing case marking on the relational noun, similarly to the situation in *Permic* just described.

In *Turkish*, a similar situation holds: the prototypical productive formation pattern of complex postpositions is [N-(POSS)-CASE], thus analogous to the complex postpositions in Basque and Permic. These morphologically complex postpositions undergo formal and semantic changes over time. On the formal side, the noun present as nucleus loses the ability to take different cases and often ends up as a fossilized form with only one specific and generally nontransparent case marker. On the semantic side, many complex postpositions with spatial semantics develop metaphorical and/or metonymical meanings which are often clearly differentiated from the original spatial meanings by the assignment properties of the postposition itself: whereas the spatial meaning is associated with genitive marking on the preceding complement noun phrase (3), metaphorical uses are associated with the ablative case assignment (4).

(3) ağac-in **yan-in-da** (Turkish)
 tree-GEN side-POSS.3-LOC
 ‘**next** to the tree’

(4) demokrasi-den **yan-a** (Turkish)
 democracy-ABL side-DAT
 ‘**in favor** of democracy’ (Jendraschek, this volume, examples (16ab))

In *Estonian*, a typical complex postposition follows a more complex structure and consists of an inflected relational noun, typically in the genitive case, and a simple postposition, i.e. [N_{rel}CASE P], thus displaying syntactic and morphological complexity at the same time (5).

(5) selja taga (Estonian)
 back.GEN behind
 ‘behind the back of’ (Jürine, this volume, from example (18))

As the development of complex postpositions is very recent in Estonian, there is not enough diachronic evidence to date to reconstruct the developmental paths of these elements in detail. However, available synchronic evidence suggests that Estonian postpositions follow developmental paths which are very similar to those of other complex adpositions described above. More specifically, the complex postpositions form a continuum from fully compositional structures to semantically non-compositional, i.e. holistic units, and thus display usual signs of unit formation and lexicalization (see chapter on Estonian for details).

In sum, in all these cases of the attested developments of complex **postpositions**, very similar general diachronic mechanisms and processes seem to be at

work: unit formation, decategorialization, loss of compositionality and analyzability, univerbation and semantic bleaching, or loss of referential meaning. Generalizing over these observations, the diachronic development of complex adpositions and other CRs, prepositions and postpositions alike (and irrespective of the presence of case-marking), appears to be a relatively homogeneous phenomenon, following known and typical mechanisms and paths. In this connection, the present volume has shown that (i) complex adpositions exist in many languages – at least in the languages described here: this could be either a typical feature of European languages, i.e. a Euroversal rather than a universal phenomenon; or it could be that nominal relations naturally come to be marked by complex structures, beyond the widely attested simple markers of nominal relations (case, simple adpositions and relational nouns); (ii) complex adpositions can come in different shapes, with a few typical patterns for European languages, such as [P N P], [P N CASE] or [ADV P], and (iii) irrespective of the main type of adposition found in a given language, i.e. *pre-*, *post-* or *ambiposition*, and of the presence or absence of case marking, they follow similar developmental paths in the course of their history.

The second central question raised in the introductory chapter concerned the conceptual status of the diachronic changes in which complex adpositions are involved and their theoretical modelling in terms of GRAMMATICALIZATION, LEXICALIZATION and/or CONSTRUCTIONALIZATION. The authors of this volume seem to share the idea that complex adpositions arise via the process of **grammaticalization**. The development of complex adpositions in the languages described here follows typical characteristics of grammaticalization processes, as for example decategorialization, semantic bleaching, syntagmatic coalescence, etc. (see above). Some authors in this volume explicitly mention **lexicalization** as a possible conceptual embedding of the observed changes, for example Anni Jürine in her chapter on Estonian (see also chapter on West Germanic for a similar view):

It is important to highlight that the main mechanism for semantic change in the development of Estonian (complex) postpositions is lexicalization. Lexicalization, in this case, is defined as a type of language change whereby a ‘new contentful’ form is created with formal and/or semantic properties that cannot be directly derived from its components (Brinton and Traugott 2005: 96). It must be noted that lexicalization is here taken to be a phenomenon that is not the opposite of the grammaticalization process, but rather the two are taken as processes that complement each other, and which can occur in a single instance of language change. To put it in Lehmann’s (2002) terms, lexicalization is here understood as ‘adoption into (mental) lexicon’, which means that the linguistic item is analyzed holistically, not analytically. (Jürine, this volume)

A significant insight from the discussions presented in this volume is that the authors, as a rule, do not see grammaticalization and lexicalization as conflicting, opposite or complementary processes in which complex adpositions might

be involved (see e.g. Moreno Cabrera 1998, Giacalone Ramat and Hopper (eds.) 1998, Brinton and Traugott 2005, Lightfoot 2005, Prévost and Fagard 2007). Rather, they consider these two processes as operating at the same time and in parallel during the development of complex adpositions. Whereas the concept of lexicalization, as creation of new holistically accessed linguistic items, i.e. linguistic signs or lexemes, applies to the initial emergence of a new complex adposition in a language, the concept of grammaticalization is applied to the small-scale changes accompanying and following it, leading to the establishment of a linguistic sign with the dedicated grammatical function, i.e. an adposition.

One possible way to reconcile the discussion of the role of lexicalization and/or grammaticalization in the development of complex adpositions is offered by the relatively new concept of **constructionalization** or constructional change (see esp. chapters on Germanic and Romance). The emergence and increasing productivity of systematic schematic patterns such as [P N P], [P N CASE], [P P] for complex prepositions or [N_{rel}-CASE]² for complex postpositions, which become available for the creation of new CAs, may just as well be conceptualized in terms of constructionalization, i.e. the emergence of a new schematic construction in a language (see also the term “grammaticalization by analogy” in Hoffmann 2004). A pattern often appears to be closely linked to a specific period, in which it emerges, gathers momentum (attracting or producing new constructs); after that, it often ceases to be productive, as is the case with [P N CASE] constructions in Medieval French, or [N CASE] constructions in Modern Serbian.

The issue of **univerbation**, which featured prominently in most individual chapters of this volume, is closely connected to the important question of the “ultimate” goal category: do complex adpositions constitute an intermediate step on the grammaticalization cline towards simple adpositions and eventually towards case markers? In a very abstract manner, the answer to this question is positive. Since the empirical evidence from different languages indicates that, in many cases, complex adpositions either shed parts or become univerbated, and thus become unanalyzable linguistic units, we can treat the resulting elements as simple adpositions. In this respect, the available evidence corroborates the assumption of the cline **complex** > **simple** adposition. Note, however, that there also seems to be a good deal of renewal within the category of CAs, many of which disappear before undergoing univerbation. This can be illustrated with French *par l'amour de* ‘for the love of, thanks to’, which was frequent enough in Medieval

² A recurrent pattern in Serbian (and other Slavic languages), which yields items generally analyzed as secondary prepositions: (Serbian) *krajem* + GEN ‘by the end of’ (*kraj* ‘end’), *vrh* + GEN ‘above’ (*vrh* ‘top’), etc.

or Classical French to be borrowed by Breton, but then disappeared altogether from French. In the corpora many examples can be found of individual CAs (i.e. tokens, not types or constructions) that become somewhat frequent, and then disappear, perhaps ousted by a new wave (see e.g. the chapter on Romance for more examples).

The development of the German preposition *wegen* ‘due to’ is a good example of the ‘shedding strategy’ (see chapter on Germanic for details). Rottet (this volume) notes the existence of a “grammaticalization cycle” from morphologically complex to simplex preposition in *Celtic*. As in the case of *a-enep* ‘against’ (6), a complex adposition based on the simple preposition *da* ‘to’ and a lost noun presumably meaning ‘face’ (cf. Welsh *wyneb* ‘face’), some of the complex prepositions “come to function like simplex prepositions, taking personal inflections directly” (Rottet, this volume).

- (6) *en da enep* ‘in your against’ (Celtic)
 → *a-enep dit* ‘against to-you’
 → *enepdout* ‘against.2SG’

For *Basque*, Jendraschek (this volume, (1)) reconstructs a “continuum of grammaticalization” with three consecutive stages leading from complex to simple adposition and eventually to case marker (7).

- (7) $N_{\text{case}} [N_{\text{rel}}\text{-CASE}]_P$ (*Joni buruz* ‘about Jon’) (Basque)
 > $N_{\text{case}} [N_{\text{rel}}\text{-}\emptyset]_P$ (*legearen arabera* ‘according to the law’)
 > $N [\text{CASE}]$ (*etxe-an* ‘in the house’). (Jendraschek, this volume, Figure (2))

First, adpositions emerge and are usually identifiable within a relevant construction. In the intermediate stage, the adposition is no longer complex and is juxtaposed to its nominal complement. In the final stage, grammaticalization has reached its endpoint and the adposition has become bound and turned into a case suffix. For *Romance* languages, it can also be said that some of the older CAs grammaticalize into simple non-analyzable adpositions, even though univerbation does not always show up in the orthography (8).

- (8) *à travers* ‘through’ (French)
par (la) faute de / *à faute de* > *faute de* ‘for want of’
du côté de > *du côté* > *côté* ‘on the side of, concerning’

As can be seen in these examples, the reduction of complex adpositions can take different forms, mainly (a) morphosyntactic reduction via univerbation and (b)

shedding, i.e. loss of constitutive elements. Though the lexical nucleus is typically maintained even when the complex adposition becomes a simple one, it can then be formally distinct from its source, as in the case of Italian *attraverso* ‘through’ (Hoelbeek 2017) (cf. Italian *traverso* ‘transverse’) or not, as in the case of German *richtung* ‘towards’ (Stathi, *forthc.*) (cf. German *Richtung* ‘direction’). One question that arises in this connection is whether the tendency for complex adpositions to eventually become simple is important in the renewal of adpositional paradigms (cf. Svorou 1994).

However, looking into the systems of simple and complex adpositions in individual languages more closely, one may as well hypothesize that – at least in some languages – CAs form a more or less independent paradigm or a (sub-) category of their own. For instance, in German, the complex prepositions predominantly follow the productive pattern [P N P/ CASE_{GEN}], whose inventory may be more or less freely expanded to accommodate new members of the class: they serve to build prepositional phrases with adverbial functions, and they do not always have semantic equivalents in the class of simple adpositions. Importantly, from this inventory of complex prepositions, only rare exceptions proceed on the grammaticalization cline towards simple prepositions, whereas the majority of them either remain complex or become obsolete over time. Taking into account these synchronic and diachronic paradigmatic properties of complex adpositions, it could be argued that they pertain to a separate category with prototypical structure. Whether or not this idea represents a plausible theoretical embedding of complex adpositions cannot be answered in the context of this discussion, and will remain an question for further research.

2 Complex nominal relators and other typological features

From a typological perspective, it seems we might expect (at least) two types of relationships between nominal relators and other typological features of a language. The first one is the well-known tendency for SVO languages to have prepositions, and for SOV languages to have postpositions (see e.g. Dryer 2013a-b):

‘Lyons (1967:302) observes that the difference between prepositions and postpositions is trivial and that ‘many linguists would say that it is mere pedantry to maintain the terminological distinction’. Research on the order of meaningful elements carried out four years earlier (Greenberg 1963d) had established that this terminology is not all that trivial.’

(Heine 1989:78)

The second is the obvious link between the global tendency of a language towards syntheticity or analyticity and its use of case markers vs adpositions to mark nominal relations. Thus, while varieties described in this book all seem to have morphologically and/or syntactically complex nominal relators, this complexity takes many forms. The word *continuum* appears in more than one chapter of this volume to describe the paradigms of CAs, and it seems safe to say that paradigms of nominal relators typically range from simple items to complex ones, in languages displaying such a category. We can posit a cline from morphologically and syntactically simple nominal relators (simple adpositions without case marking and simple case markers) to complex ones, some languages having both morphologically *and* syntactically complex nominal relation markers (see e.g. the variation between [P₁ (DET) N P₂] and [P₁ (DET) N + CASE] in German and Romanian, in chapters 2 and 3). The data from Permic (chapter 9) suggests that, in some cases, one could establish a gradient of complexity from syntax to morphology. Though the most obvious type of morphologically complex relators are combinations of case markers, such as those found in Permic, Estonian (chapter 8) or Basque (chapter 10), compound prepositions are sometimes transparent enough to license an analysis in terms of morphological complexity, e.g. in Serbian (chapter 3), Albanian (chapter 7) and Maltese (chapter 12). The opposition between morphological and syntactic complexity can be illustrated by the prevalence of [P₁ DET N P₂] or [CASE + N-CASE] structures, as illustrated in Table 1. The analysis of a construction as a Compound, rather than as a Simple or Complex adposition, is problematic, and may depend on the perspective adopted by the authors (more synchronic, or more diachronic): we maintained this distinction in Table 1, but highlighted the category in grey to show that it may have a different status.

A third relationship could be expected between typological features and CRs – CAs in particular: since languages with case markers often also have simple adpositions, one might think that there is some kind of (inverse) correlation between the presence of case marking in a language and its having complex adpositions: schematically, the first degree of complexity in a language with case marking and simple adpositions is the combination of the two, while in a language without case marking it is compound adpositions, as illustrated in Table 2.

This is a very schematic view, since the complexity and weight of case marking vary greatly from one language to the next, as illustrated by the data on Romanian, German, Serbian, Greek, Albanian, Estonian, Permic, Basque and Turkish (chapters 2–4 and 6–11). If we do accept this idea, it could be thought that languages with case marking have less developed paradigms of CRs. Some languages do seem to fit this hypothesis, for instance Old French and Albanian (with case marking and limited paradigms of CRs) vs Modern French, English and

Table 1: Degrees of complexity of nominal relators (in bold, underlined) within PPs (*NP* complements in italics).

Use type	Marker type	Languages		
		without case marking	with case marking	
			no semantic differentiation	semantic differentiation
Functional (not a head, grammatical uses)	Simple	<u>PREP_i NP</u> Fr., Pt., En., Dt.	<u>NP-CASE</u> Gr., Rm., Bs., Pm., Tk., Al., Es., Gm., Sb.	–
		<u>PREP NP</u> Fr., Pt., En., Dt., Mlt.	<u>NP-CASE^(STRUCTURAL**)</u> Gr., Rm.*	<u>NP-CASE</u> Bs., Pm., Tk., Al., Es., Gm., Sb.
Lexical (functions as a head)	Simple	<u>PREP NP-CASE^(ARCHAIC)</u> Dt.	<u>PREP NP-CASE^(STR)</u> Gr., Rm.*, Gm., Sb.	<u>PREP NP-CASE</u> Gm., Sb.
		Compound (e.g. PP, AdvP)	<u>PREP₁PREP₂ NP</u> Fr., Pt., En., Mlt.	<u>PREP₁PREP₂ NP-CASE</u> Sb.
Complex	Compound (e.g. PP, AdvP)	<u>ADVPREP NP</u> Sp., Pt., It.	<u>ADVPREP NP-CASE^(STR)</u> Gr.	<u>NP N-CASE</u> Bs. (inanimate nouns)
		<u>PREP₁NPREP₂ NP</u> Fr., Pt., En.	<u>PREP₁N NP-CASE^(STR)</u> Gr.	<u>N-CASE NP-CASE</u> Bs.
		–	–	<u>NP-CASE N-CASE</u> Bs.
		<u>PREP₁detNP₂ NP</u> Fr., Pt., En.	–	<u>PREPdetN</u> <u>NP-CASE</u>
		–	–	–

Notes:

* in Modern Romanian, some Ps take the accusative, others the genitive, but there is no single P with the two cases.

** the default is that cases add a semantic value to the construction; in some instances, however, there seems to be no opposition between different cases, and thus no semantic value: we then consider them to have only a structural value, and they are noted CASE^(STR).

– the category is not attested / there is no corresponding category.

Table 2: Degrees of complexity and presence of case marking.

Degrees of complexity	Language with case marking	Language without case marking
single marking	case	simple adposition
double marking	case + simple adposition	compound adposition
complex marking marker	case + compound adposition	complex adposition
highly complex marking	case + complex adposition	–

Portuguese, for instance. The case of German, which has case marking and large paradigms of CAs – paradigms which are, in fact, quite comparable to those of Dutch and English –, seems to indicate that the hypothesis is at best partly valid. Besides, Serbian (chapter 3), which has an even more elaborate system of case marking, also has a large paradigm of CAs.

3 The emergence of CAs in Europe

From a diachronic perspective, the emergence of CRs is quite easy to understand, at least in very general terms. For instance, looking at adnominal relations in Indo-European languages, we can see a global paradigm shift from proto-Indo-European (with no case marking, see e.g. Meillet 1948:544) to a closed-class paradigm of case marking in Indo-European (8 cases, initially lexical elements which gradually grammaticalized to various extents, see e.g. Lehmann 1974:49, Haudry 1982:20, Meier-Brügger 2000:12). What is striking in the later evolution is that new relational elements keep stacking up onto these original ‘simple’ adnominal relation markers: first adverb-like elements (see e.g. Beekes 1995:218), which then become reanalyzed as simple prepositions (Giacalone Ramat and Ramat 1993:102) and end up, in quite a few Indo-European languages, replacing case marking altogether as adnominal relation markers. In this perspective, CAs constitute just another layer of complexity. This general trend could be analyzed as follows: at any point in time, a language typically has a closed-class paradigm for adnominal relation marking. But closed-class paradigms eventually prove to be functionally insufficient,³ the need or desire to specify or disambiguate further the possible

³ Cf. for instance Hopper & Traugott (1993:88): “As grammaticalized forms become increasingly syntacticized or morphologized they unquestionably cease over time to carry significant semantic or pragmatic meaning”.

⁴ In the case of Indo-European, these elements were adverbs, accompanying the initial set of case markers. Once they were used frequently, these adverbs were reanalyzed not as *accompanying* case markers, but as *specifying* their meaning, and further as *governing* the noun phrase initially headed by case alone – a situation which is found in Classical languages, e.g. Latin and Ancient Greek. This is roughly how (simple) prepositions were introduced in the language system. In the next step, this renewed paradigm of nominal relation markers has also proven to be insufficient, which resulted in the broadening of the class either by combining simple prepositions in order to form compound ones, or by recruiting new markers in the nominal domain. A whole range of relational and spatial nouns were thus integrated, alone in an inflected form or in combination with a preposition, in the paradigm of nominal relation markers in order to express in a finer-grained way, at least in the first instance, various concrete relationships, and later many other kinds of semantic relations

semantic relations triggering the use of new elements.⁴ Whatever the stage of evolution, the trend seems to be the same: a deficiency of an initial paradigm triggers an enlargement making it possible for language to respond to increasing communicative needs.

However, the case of CAs in European languages remains particularly interesting, because they appear to have emerged as an areal phenomenon (e.g. Popović 1966, Van der Horst 2013, Hüning 2014). Thus, their emergence could be the result of the ‘natural’ evolution described above – just another layer of complexity to remedy the limitations of closed-class paradigms – or it could be linked to socio-cultural changes which took place at some point between the Renaissance and Modern Times. It is quite widely accepted by now that the cultural proximity across the whole of Europe may have led to a slow but steady leveling of linguistic differences (Haspelmath 2001: 1506–1507), the outcome being what has been termed ‘Standard Average European’ (Whorf 1956 [1939], quoted in Van der Auwera 2011: 291). And it is quite clear that CAs have appeared in several European languages during the last few centuries: as noted for instance in the chapter on Romance, there are [PNP] constructions in the Middle Ages, but CAs to have emerged as a class (i.e. with consistent patterns and important type and token counts) only later on. Thus, for French, to take just one example, there is a striking contrast between Old and Middle French, on the one hand, and Classical and Modern French, on the other. In other chapters, slightly different periods are mentioned: 18th and 19th centuries in Serbian, 19th century in the Germanic languages.

Another striking feature of the emergence of CAs in European languages is that two forces seem to have played an important role: *cultural drift* (Van der Horst 2013:182) and *language contact*. The use of CAs is strongly linked to formal registers and/or the Press, an aspect which has been noted for various languages and by several authors (see e.g. the chapters on Germanic and Serbian; cf. Benes 1974, Blumenthal 2018, Stefanowitsch forthc.). Popović (1966:195–196) namely writes that it is linked to “the development of modern sciences, administration and journalism”, a statement which sums up neatly the cultural trends possibly involved in the rise of CAs (see also Lehmann’s “bureaucratic and journalistic jargon”: Lehmann 1991:501). The transformation of Europe in the 18th and 19th centuries – first and foremost in England, France and Germany, but also in neighbouring countries – was quite radical, not only in industrial terms but also in terms of society and politics (see e.g. Piel 1989:138 sq.). For one thing, there is the slow regression of Latin, which had been the language of culture, diplomacy and science for the whole period of the Middle-Ages, and well into the Renaissance. Latin may be said to have remained the language of prestige until the 17th and even 18th centuries (Febvre and Martin 1976); however, there was a slow but regular progression of vernacular languages in that period: more and more people came to have access

to writing without mastering Latin, scientific works began appearing in French, English, German, Dutch, etc. The important industrialization process (with what could be called an “industrial Enlightenment”, see Mokyr 2009) could also have had an impact on language: as noted by Popplow (2016:282), this process brought on a set of interrelated developments resulting “in the fact that technical knowledge was no longer nearly exclusively embodied in an artisan [...] Towards the end of the early modern period, such formalized technical knowledge was collected, discussed, and taught in institutions such as scientific academies, economic societies, and engineering schools”. Finally, in the 19th century (and even more so in the early 20th century), the press achieved a wide distribution, from some tens of thousands of copies around the middle of the 19th century to hundreds of thousands at the turn of the century and millions shortly thereafter (Weill 1934).

The importance of language contact has also been noted in various chapters of the volume. The existence of similarities in cognate languages is quite obvious, and can be seen for instance in the chapters on Romance and Germanic; but the borrowing of CAs from (genealogically) more distant languages is noted in various chapters, e.g. on Breton, Serbian, Maltese and Greek; additionally, it is mentioned as a possibility in the chapter on Albanian. In the chapter on Estonian, though the emergence of CAs is not ascribed to language contact, the author does mention the importance of contact with German, Swedish, and Russian.

Two counter-arguments could be brought to this hypothesis that CAs might have emerged in European languages mainly as a result of a global cultural drift associated with close language contact across Europe. The first one is that, among the languages investigated in this volume, the importance of contact is not as clear in languages which do not have CAs but CRs: Turkish, Estonian, Basque, Maltese and Permic. For these languages, it would be interesting to dig further in order to check the possibility of an independent development of CRs. The second counter-argument is even more obvious: there are CAs outside Europe.⁴ We address this in the following section.

4 Beyond Europe

Concerning *simple* nominal relators, previous studies (e.g. Svorou 1994, Hagège 2010) have shown that there are important differences across languages. These differences may for instance concern the type of items which encode nominal

⁴ Of course, this is a counter-argument only if these CAs outside Europe developed independently from any similar ‘cultural drift’.

relations, mainly case suffixes and prepositions, but also postpositions, prefixes (much less common than case suffixes, according to Dryer 2013c) and relational nouns (typically with some items which grammaticalize into adpositions, see e.g. Heine 1989, DeLancey 1997). There are also important differences in paradigm size, with e.g. only *one* simple adposition in some languages, such as Burmese (Tibeto-Birman, Sino-Tibetan), Likpe (Kwa, Niger-Congo), Samoan (Eastern Malayo-Polynesian, Austronesian), Tagalog (Western Malayo-Polynesian, Austronesian), or Tzeltal (Mayan), and long lists in other languages (e.g. Romance languages, see Fagard 2006). One might expect similar variations for complex nominal relators. Indeed, the previous sections present a few core patterns: constructions with *adpositions* as relators, with *case markers* as relators, and with a combination of the two. The situation is actually even more complex, since the distinction between case markers and adpositions is not entirely discrete: rather, while some case markers are clearly affixes or even clitics, and some adpositions clearly independent words, quite a few items fall in-between these categories. For instance, functional adpositions may grammaticalize to the point that they resemble case markers in at least some respects, and some items can be analyzed both as case affixes and postpositions, for instance in Hungarian (Uralic; see Creissels 2006) or Korean (isolate; see Choi-Jonin 2008).

Complex nominal relators, as we define them, are apparently found in all languages having simple nominal relators. It is quite easy to find languages with adpositions which also have CAs, as in Malagasy. To give just a few examples, Malagasy (Austronesian) displays a construction which seems perfectly parallel to those described in this volume: alongside simple adpositions, which are typically quite polysemous, such as *amy* ‘at’ (9–10), there are several sequences of simple adposition and noun (sometimes in disuse, see Ferrand 1903:232), which can introduce an NP, as in (11).

- (9) Dokotera izy ka manao fitsdihana ny olona (Malagasy)
 Doctor PRO and do visit DET person
 eny *an-tsambo*
 there on-boat
 ‘He’s a doctor and visits people on the boat’ (Fugier 1999:136)
- (10) *An-katezerana no itenenan -dRaso*a (Malagasy)
 with-anger EMPH speak.PASS Raso
 ‘Raso speaks angrily’ (Fugier 1999:82)
- (11) *Indro ity zazavavy nataon’ izy roa lahy* (Malagasy)
 behold DEM.PROX young.girl put.by PRO two man

am-povoa ny nenti-ny nidina mankany Isotry any
 at-middle POSS taken- POSS descend go.there.PRS Isotry there
 ‘And lo the young girl was put by the two men between them and taken
 down toward Isotry’ (Fugier 1999:223)

Another example with simple and complex adpositions is Teochew, an isolating, SVO Southern Min dialect of Chinese. Alongside simple prepositions such as [pu_ŋ³³] ‘at, by’, [do⁵] ‘at, by, in’ or [dui²¹³] ‘by, via, along’, Teochew has constructions with a simple preposition and a relational noun, the latter specifying the position of the target with respect to the landmark (see My Dung, in prep.), as in (12–13).

(12) *do⁶ cwj⁵ dej²* (Teochew)
 at table top
 ‘on the table’

(13) *do⁶ cwj⁵ bin¹tao²* (Teochew)
 at table side
 ‘next to the table’

Languages with case marking such as Japanese also display sequences which correspond to what we call CRs, in which a nominal or adverbial nucleus (*naka* in (14)) is linked to the complement (*ie* ‘house’) by a genitive case marking (*-no*) and to the rest of the sentence by a locative case marking (*-de*).⁶

(14) *Mary wa ie-no naka-de hashit-ta.* (Japanese)
 Mary TOP house.GEN inside.LOC run.PRS
 ‘Mary was running in the house.’

Other types of CRs are attested, such as the combination of a nucleus and a case marker, illustrated in (15) by Jaminjung (Yirram, Mirndi), or the combination of case markers illustrated in (16) by Tabasaran (Lezgetic, Nakh-Daghestanian), including rare types of case marking as in Maasai (Nilotic, Eastern Sudanic, Nilo-Saharan), in which the case is marked by tone (17).

(15) *dibard ga-w-ijga walthub langiny-bina.* (Jaminjung)
 jump 3SG-FUT-go inside wood-ALL
 ‘It will jump into the trees.’ (Schultze-Berndt 2000: 425)

⁵ Some examples are adapted from Fortis & Fagard (2010).

- (16) *nir-q-an-di* (Tabasaran)
 river-behind-ABL-TSL
 ‘From the direction of (the bank of) the river.’ (Comrie and Polinsky 1998:99)
- (17) *me sídáí toó-(i)l-áyìð-(ò)k* (Maasai)
 NEG good for-M.PL-boy-PL
 ‘it is not good for the boys’ (Tucker and Mpaayei 1955:216, quoted in Hagège 2010:15)

As noted by Fortis and Fagard (2010), in languages with only one simple adposition, this adposition has a generic meaning, and is used alone, i.e. with no other adnominal marker, only for canonical spatial relations (18): When further specification is needed, it may be done with the help of spatial nouns (19) or positionals, thus yielding CRs.

- (18) *na sa Silya Ang kanya-ng damit.* (Tagalog)
 REAL PREP chair NOM 3SG.OBL-LNK dress
 ‘Her dress is / was on the chair.’
- (19) *na sa harap ng silya siya.* (Tagalog)
 REAL PREP front GEN chair 3SG.NOM
 ‘S/he is in front of the chair.’

In (19), the complexity is purely syntactic; examples (20–21) (Wilkins 2006:33) illustrate for Arrernte (Central Pama-Nyungan) the existence of morphological complexity (20) or both morphological and syntactic (21) (also (10), above).

- (20) *panikane-ø tipwele akertne-le (ane-me).* (Arrernte)
 cup.NOM table up/top.LOC (sit.NPP)
 ‘The cup is on top of the table.’
- (21) *alkngenthe-ø tipwele-nge akertne-le.* (Arrernte)
 light.NOM table.ABL up/top.LOC
 ‘The lamp is above the table.’

In Korean (isolate), there is a debate concerning the status of adnominal markers (see Choi-Jonin 2008), which are post-posed and could be analyzed as case markers or as postpositions (as we mentioned for Hungarian) – some linguists consider only a subparadigm to be actual case markers, while the others should be analyzed as postpositions. While they appear to be morphologically simple

in synchrony (22–23), some seem to arise from the simplification of complex markers. For instance, diachronic studies show that the ablative-inessive *eyse* results from the combination of the locative *ey* and an ancient verb *se* ‘to exist’ (24); a similar analysis could be proposed for the exceptive (25), based on *pak* ‘outside’ in the locative case *ey*.

- (22) *abeci-ga cha-lul sa-ss-ta* (Korean)
 father-NOM car-ACC buy-PFT-TS(DECL)
 ‘The father bought a car’
- (23) *abeci-eygey cenhwa-hay* (Korean)
 father-DAT phone-do.IMP
 ‘Call the father’
- (24) *edi-eyse o-ass-eyo?* (Korean)
 where-ABL come-PFT-TS.INT
 ‘Where do you come from?’
- (25) *ne-pakkey eps-ta* (Korean)
 you-except not.exist-ST(DECL)
 ‘There’s only you’⁶

Hagège (2010) restricts what he calls “Complex Adps” to “a word-type in which one of the constituent parts is an Adp, and the other is a case affix, in a form required by the Adp” (Hagège 2010: 38–39).⁷

As further noted by Fortis and Fagard (2010), “In a number of languages, a part which functions as a spatial noun must be ‘locativized’ (receive a locative marker)” (cf. Heine 1989: 103–4 for Bantu languages), thus also yielding CRs, as in Swahili (Bantoid, Benue-Congo, Niger-Congo) *mbele* ‘front’ < locative class prefix *mu-* + **-bele* ‘breast’).

⁶ Examples (22–25) are from Choi-Jonin (2008 and p.c.).

⁷ Excluding from the core category, i.e. what he calls *Complex Adps proper*, marginal instances such as phenomena of *case heaping* (“a phenomenon involving adnominal complements marked by a succession of two or more contiguous case markers or case marker-like elements”, Hagège 2010:39), *case agreement* (“situations in which two case-markers seem to be applied to the same element”, Hagège 2010:40) and *case association* (in which two words are marked with the same case, one “signaling location in general, while the other refers to the specific orientation of the located entity” (Hagège 2010: 42–43); this “can be considered as a type of double marking” (Hagège 2010:43)).

It would be interesting to see whether rare systems of adpositions, such as inpositions – as found in Anindilyakwa (Gunwinyguan; Dryer 2013a) –, also include CRs (26).

- (26) ...*narri-ng-akbilyang-uma* [eyukwujiya=**manja** eka] (Anindilyakwa)
 ...NC₁.PL-NC₂-stick.to.end-TA [small=LOC stick]
 ‘... they stuck them (the feathers) to a little stick.’
 (Groote Eylandt Linguistics-langwa 1993: 202)

While prepositions and case-marking systems are widely attested in the languages of the world, a question that comes to mind is what would correspond to a CR in a language which has neither.

According to DeLancey (2005), Klamath (Klamath-Modoc, Penutian) probably has no adposition, as spatial relations are encoded in the verb, in Locative Directive Stems, as in (27).

- (27) *c-eq'ya* *honk doo* *wqepl'ags-dat.* (Klamath)
 sit-in.doorway –DEM over.there summer.house-LOC
 ‘She sat at the door of the house.’ (Barker 1963:4.1;107)

However, Klamath does have a case marking system. Besides, Locative Directive Stems constitute a very large paradigm – containing 133 items, according to DeLancey (2000, following Barker 1964): while they “do not represent any familiar or well-understood category” (DeLancey 2000:71), DeLancey shows that they are verbal “at least in origin” (ibid.). Dryer (2013a) notes that languages with no adpositions “are most common in Australia and North America”. Those that have case markers, such as Yidiny (Northern Pama-Nyungan; Dixon 1977), could have CCs, as described e.g. in Permic (Chapter 9). But in languages which have neither adpositions nor case markers, such as Kutenai (isolate; western North America), it would be interesting to see the range of linguistic items used for nominal relations. More generally, it would be interesting to test the link between the *type of items* found in a language for marking nominal relations (*case markers, adpositions, relational nouns*), their *position* (in front of the complement, behind it, around it or infixes, i.e. within the complement itself), their *number* (i.e. the paradigm size) and the respective proportions of simple vs complex markers.

To conclude, let us turn back to our initial methodological decisions: in this volume, we chose to focus on a limited area of the world – Europe –, to proceed, as much as possible, within a common theoretical framework, and to ask authors to combine (wherever such data were accessible) synchronic and diachronic approaches. We believe that this approach has yielded fruitful results: it has

made it possible to provide a good coverage of the phenomenon under investigation, i.e. Complex Adpositions and more widely Complex Nominal Relators, and to bring to light interesting phenomena, raising new research questions or at least leading us to reframe the questions we had started out with. It seems quite clear that, being half-way between lexicon and grammar, CAs uniformly play a fundamental structuring role in languages, whatever their position in adpositional phrases, their patterns of formation, or their paths of evolution. However, it is equally obvious that there is still substantial work to be done: first, in order to explain by testing, for instance, the ‘cultural drift’ hypothesis, which has only been briefly touched upon here, second, in order to probe the universality or, at the very least, the communality, outside Europe, of the main construction schemas of complex adnominal relations, as revealed in this volume, and third in order to grasp the whole range of functional and semantic relations that CAs convey across languages. We hope that the results of this volume are enough to open this new line of research.

Abbreviations

2, 3	second, third person	CR	complex nominal relator	DET	determiner
ABL	ablative	DAT	dative	EL	elative
ACC	accusative	DECL	declarative	FUT	future
ALL	allative	DEM	demonstrative	GEN	genitive
CA	complex adposition	N _{rel}	relational noun	IMP	imperative
INT	interrogative	OBL	oblique	PRS	present
LNK	linker	PASS	passive	REAL	realis
LOC	locative	PFT	perfect	SG	singular
M	masculine	PL	plural	TA	tense/aspect
NC	noun class	POSS	possessive	TOP	topicalizer
NEG	negation	PREP	preposition	TS	terminal suffix
NOM	nominative			TSL	translative

References

- Auwera, Johan van der (ed.). 1998. *Adverbial constructions in the languages of Europe*. Berlin & New York: Mouton de Gruyter.
- Barker, Muhammad Abd-al-Rahman. 1963. *Klamath Texts*. Berkeley & Los Angeles: University of California Press.
- Barker, Muhammad Abd-al-Rahman. 1964. *Klamath Grammar*. Berkeley & Los Angeles: University of California Press.

- Beekes, Robert. 1995. *Comparative Indo-European linguistics: an introduction*. Amsterdam & Philadelphia: John Benjamins.
- Benes, Eduard. 1974. Präpositionswertige Präpositionalfügungen. In Ulrich Engel & Paul Grebe (eds.), *Sprachsystem und Sprachgebrauch: Festschrift für Hugo Moser zum 65. Geburtstag* (Sprache der Gegenwart 33), vol. 1, 33–52. Düsseldorf: Schwann.
- Blumenthal, Peter. 2018. Locutions prépositionnelles « phénoménologiques ». In Céline Vaguer-Fekete (ed), *Quand les formes prennent sens. Hommage à Danielle Leeman*, 195–204. Limoges: Lambert-Lucas.
- Brinton, Laurel & Elizabeth C. Traugott. 2005. *Lexicalization and language change*. Cambridge: Cambridge University Press.
- Choi-Jonin, Injoo. 2008. Particles and postpositions in Korean. In Dennis Kurzon & Silvia Adler (eds.), *Adpositions: Pragmatic, semantic and syntactic perspectives*, 133–170. Amsterdam & Philadelphia: John Benjamins.
- Cifuentes Honrubia, José Luis. 2003. *Locuciones prepositivas – sobre la gramaticalización preposicional en español*. Alicante: Universidad de Alicante.
- Comrie, Bernard & Maria Polinsky. 1998. The great Daghestanian case hoax. In Anna Siewierska & Jae Jung Song (eds.), *Case, Typology and Grammar*, 95–113. Amsterdam & Philadelphia: John Benjamins.
- Creissels, Denis. 2006. Suffixes casuels et postpositions en hongrois. *Bulletin de la Société de Linguistique de Paris* 101(1), 225–272.
- DeLancey, Scott. 1997. Grammaticalization and the gradience of categories: Relator nouns and postpositions in Tibetan and Burmese. In Joan Bybee, John Haiman & Sandra A. Thompson (eds.), *Essays on Language Function and Language Type*, 51–69. Amsterdam & Philadelphia: John Benjamins.
- DeLancey, Scott. 2000. Argument structure of Klamath bipartite stems. In Andrew Simpson (ed.), *Proceedings of the 26th Annual Meeting of the Berkeley Linguistics Society: Special Session: Syntax and Semantics of the Indigenous Languages of the Americas*, 15–25. Berkeley: Berkeley Linguistic Society.
- DeLancey, Scott. 2005. Adposition as a non-universal category. In Zygmunt Frajzyngier, Adam Hodges & David Rood (eds.), *Linguistic diversity and language theories*, 185–202. Amsterdam & Philadelphia: John Benjamins.
- Di Meola, Claudio. 2000. *Die Grammatikalisierung deutscher Präpositionen*. Tübingen: Stauffenburg.
- Dixon, Robert M. W. 1977. *A Grammar of Yidin*. (Cambridge Studies in Linguistics, 19.) Cambridge: Cambridge University Press.
- Dryer, Matthew S. 2013a. Order of Adposition and Noun Phrase. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Dryer, Matthew S. 2013b. Order of Subject, Object and Verb. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Dryer, Matthew S. 2013c. Position of Case Affixes. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Fagard, Benjamin. 2006. Evolution sémantique des prépositions dans les langues romanes: illustrations ou contre-exemples de la primauté du spatial ? PhD Thesis, University of Paris 7.

- Febvre, Lucien & Henri-Jean Martin. 1997 [1976]. *The Coming of the Book. The Impact of Printing 1450–1800*. London & New York: Verso.
- Ferrand, Gabriel. 1903. *Essai de grammaire malgache*. Paris: E. Leroux.
- Fortis, Jean-Michel & Benjamin Fagard. 2010. *Space in language*, Summer School on Linguistic Typology. (online: <http://htl-test.linguist.univ-paris-diderot.fr/Fortis>).
- Fugier, Huguette. 1999. *Syntaxe malgache*. Louvain-la-neuve: Peeters.
- Giacalone Ramat, Anna & Paolo Ramat. 1993. *Le lingue indoeuropee*. Bologna: Il Mulino.
- Giacalone Ramat, Anna & Paul Hopper (eds.). 1998. *The Limits of grammaticalization*. Amsterdam & Philadelphia: John Benjamins.
- Greenberg, Joseph H. 1963. *Universals of language*. Cambridge, Mass.: MIT Press.
- Hagège, Claude. 2010. *Adpositions*. Oxford: Oxford University Press.
- Haspelmath, Martin. 2001. The European linguistic area: Standard Average European. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language Typology and Language Universals, 1492–1510*. Berlin: De Gruyter Mouton.
- Haudry, Jean. 1982. *Préhistoire de la flexion nominale en indo-européen*. Lyon: Institut d'études indo-européennes de l'université Jean Moulin.
- Heine, Bernd. 1989. Adpositions in African languages. *Linguistique Africaine* 2. 77–127.
- Hoelbeek, Thomas. 2017. *The Evolution of Complex Spatial Expressions within the Romance Family*. Leiden: Brill.
- Hoffmann, Sebastian. 2004. Are Low-Frequency Complex Prepositions Grammaticalized? On the Limits of Corpus Data – and the Importance of Intuition. In Hans Lindquist & Christian Mair (eds.), *Corpus Approaches to Grammaticalization in English*, 171–210. Amsterdam & Philadelphia: John Benjamins.
- Hoffmann, Sebastian. 2005. *Grammaticalization and English Complex Prepositions. A corpus-based analysis*. London & New York: Routledge.
- Horst, Joop van der. 2013. *Taal op drift: lange-termijnontwikkelingen in taal en samenleving*. Amsterdam: Meulenhoff.
- Hüning, Matthias. 2014. Over complexe preposities en convergentie [About complex prepositions and convergence]. In Freek Van de Velde, Hans Smessaert, Frank Van Eynde & Sara Verbrugge (eds.), *Patroon en argument. Een dubbelfeestbundel bij het emeritaat van William Van Belle en Joop van der Horst* [Pattern and argument. A double festschrift on the occasion of the retirements of William Van Belle and Joop van der Horst], 433–445. Leuven: Universitaire Pers Leuven.
- Kurzon, Dennis & Silvia Adler (eds.). 2008. *Adpositions: Pragmatic, semantic and syntactic perspectives*. Amsterdam & Philadelphia: John Benjamins.
- Lehmann, Christian. 1991. Grammaticalization and related changes in contemporary German. In Elizabeth C. Traugott & Bernd Heine (eds.), *Approaches to grammaticalization*, vol. 2, 493–535. (Typological Studies in Language 19). Amsterdam & Philadelphia: John Benjamins.
- Lehmann, Christian. 2015 (3rd edition). *Thoughts on Grammaticalization*. Berlin: Language Science Press.
- Lehmann, Winfred. 1974. *Proto-Indo-European Syntax*. Austin: University of Texas Press.
- Libert, Alan. 2013. *Adpositions and Other Parts of Speech*. Frankfurt am Main: Peter Lang.
- Lightfoot, Douglas J. 2005. Can the lexicalization/ grammaticalization distinction be reconciled? *Studies in language* 29/3, 583–615.
- Lindqvist, Christer. 1994. *Zur Entstehung der Präpositionen im Deutschen und Schwedischen*. Tübingen: Max Niemeyer Verlag.

- Linguistics-langwa, Groote Eylandt. 1993. *Eningerribirra-langwa jurra* [Book about all sorts of things]. Angurugu: Groote Eylandt Linguistics.
- Lyons, John. 1967 / 1977. *Semantics*, 2 vol., Cambridge: Cambridge University Press.
- Meier-Brügger, Michael. 2000. *Indogermanische Sprachwissenschaft*. 7th edn. Berlin & New York: Walter de Gruyter.
- Meillet, Antoine. 1948. *Linguistique historique et linguistique générale*. Paris: Champion.
- Mokyr, Joel. 2009. *The Enlightened Economy. An Economic History of Britain, 1700–1800*. New Haven: Yale University Press.
- Moreno Cabrera, Juan C. 1998. On the relationship between grammaticalization and lexicalization. In Anna Giacalone Ramat & Paul Hopper (eds.), *The Limits of grammaticalization*, 211–227. Amsterdam & Philadelphia: John Benjamins.
- Piel, Jean. 1989. *Esquisse d'une histoire comparée des développements dans le monde jusque vers 1850*. Nanterre: Erasme.
- Popović, Ljubomir. 1966. Predložki izrazi u savremenom srpskohrvatskom jeziku [Phrasal prepositions in contemporary Serbian]. *Naš jezik* [Our language], XV, vol. 3–4. 195–220.
- Popplow, Marcus. 2016. Technology and technical knowledge in the debate about the “great divergence”. *Histoire et Sciences Humaines 4: L'Europe technicienne. XVe–XVIIIe siècle. Artefact. Techniques*. Rennes: Presses Universitaires de Rennes, 275–286.
- Prévost, Sophie & Benjamin Fagard. 2007. Grammaticalisation et lexicalisation: la formation d'expressions complexes. *Langue française* 156. 3–8.
- Schultze-Berndt, Eva. 2000. *Simple and complex verbs in Jaminjung. A study of event categorisation in an Australian language*. Nijmegen: University of Nijmegen. (= MPI Series in Psycholinguistics, 14).
- Schwenter, Scott & Elizabeth C. Traugott. 1995. The semantic and pragmatic development of substitutive complex prepositions in English. In Andreas Jucker (ed.), *Historical pragmatics*, 243–273. Amsterdam: John Benjamins.
- Stathi, Katerina. forthc. Die Entstehung der komplexen Präposition *in Richtung* im Deutschen: Zur Emergenz und Verfestigung von Konstruktionen. In Alexander Lasch & Alexander Ziem (eds.), *Konstruktionsgrammatik VI. Historische Konstruktionsgrammatik: Konvergenzen und Divergenzen im Sprach- und Konstruktionswandel*. Tübingen: Stauffenburg.
- Stefanowitsch, Anatol. forthc. Complex prepositions of contrast and opposition in English. In Marcin Grygiel and Barbara Lewandowska-Tomaszczyk (eds.), *Contrast and analogy in language: Perspectives from cognitive linguistics*. Amsterdam & Philadelphia: John Benjamins.
- Svorou, Soteria. 1994. *The Grammar of Space*. Amsterdam & Philadelphia: John Benjamins.
- Tucker, Archibald N. & John Tompo Ole Mpaayei. 1955. *A Maasai Grammar with Vocabulary*. Leyden: Publications of the African Institute / London: Longmans, Green & Co.
- Weill, Georges. 1934. *Le journal. Origines, évolution et rôle de la presse périodique*. Paris: La renaissance du livre.
- Whorf, Benjamin. 1956. *Language, Thought and Reality. Selected Writings of Benjamin Lee Whorf*. Edited by John B. Carroll. Cambridge, Mass.: MIT Press.
- Wilkins, David. 2006. Towards an Arrernte grammar of space. In Stephen Levinson & David Wilkins (eds.), *Grammars of Space: Explorations in Cognitive Diversity*, 24–62. Cambridge: Cambridge University Press.

Index

- abessive 302, 304
- ablative 22–3, 47, 195, 220–8, 240, 245–6, 267–72, 287–8, 302, 341, 359, 361, 363, 370, 373, 375, 381–2, 384, 390, 393–5, 404–5, 410–11, 420, 424–5, 427, 454, 477, 490
- ablative-locative (transfer) 195–6, 225, 227–9
- absolute 370–1, 377
- accusative 82–5, 105, 111, 115, 145–6, 152, 177, 235–6, 253–4, 259, 261, 267–70, 272, 287–8, 345, 359, 374–5, 403–5, 425, 440, 483
- adessive 302, 329–30, 363
- adverb 1, 4, 15, 19, 21–3, 41–2, 44, 52–5, 67, 69–71, 78, 83, 102, 104–10, 122–5, 148–9, 149, 151–2, 159, 161, 166–7, 180, 199, 205, 207, 233, 236–9, 243, 246–9, 260, 269, 277, 291–3, 301, 311, 330, 348, 367, 373, 375, 379–81, 383–5, 387–9, 395–6, 398, 405, 407, 411, 425, 430, 433, 440, 445, 475, 481, 484, 488
- complex adverb 301, 306, 311, 318, 330
- affixation
- affix 69, 90, 109–10, 375, 408, 426, 445, 487, 490
- infix 18, 195, 491
- prefix 1, 48, 104, 109, 122, 124–5, 151, 195, 206, 222–5, 228, 290, 487, 490
- suffix 20, 22–3, 100, 104, 109, 201–2, 217, 267–8, 277, 290, 301–2, 305, 339, 340–2, 344–6, 348–51, 353–4, 360–1, 367, 369–72, 375, 377–8, 383–4, 386–7, 395–6, 398, 403–6, 408–9, 411–14, 418, 420, 422–5, 430–1, 440, 480, 487
- allative 146, 220, 245–6, 302, 304, 369–70, 372, 375–6, 381–2, 384, 390, 393, 395, 404, 424–5, 453
- allative-dative 22, 404
- ambiposition 3, 77, 79, 81–3, 302, 478
- analogy 59, 114, 131, 133, 153, 200, 222, 330, 456, 462, 479
- analyticity 35, 59, 87, 109, 125–6, 130, 141, 154, 262, 266–7, 482
- approximative 306
- areal 9, 15, 59, 285, 485
- Balkan features 35
- Balkan Sprachbund 235, 265
- Balkanic language 266
- (language) contact 18–20, 59, 163–4, 207, 223, 228, 247–8, 265–6, 265–6, 286, 294, 301, 335–6, 353, 361–3, 403, 439, 440, 485–6
- contact-induced change 18–19, 336, 361
- associational strength 314, 317–19
- attrition 144, 368, 386, 394, 397–8, 421
- Balkans. *See* areal
- benefactive 367, 371–2, 385
- body-part words 20, 283, 314–19, 322, 324, 328–9, 388
- borrowing 114, 122, 163, 202, 209–10, 257, 265, 290, 371, 382, 389, 406–7, 439–40, 480, 486
- case. *See also:* abessive, ablative, ablative-locative, absolute, accusative, adessive, allative-dative, approximative, benefactive, comitative, dative, destinative, egressive, elative, ergative, essive, genitive/nominative, genitive-dative, illative, inessive, lative, motivative, non-nominative, partitive, perlative, prolative, separative, stative, terminative, translative, vocative
- case assignment 5, 83, 115, 477
- case marking 11–13, 16, 20–2, 24, 35, 37–40, 48, 56, 84–5, 87, 107, 115, 140, 143–4, 150, 155–7, 171, 177, 186, 208, 259, 265, 267–8, 270, 272, 286, 287, 294, 347, 350, 357, 367–72, 374–7, 379–80, 384–5, 387, 390–6, 403, 405–7, 413–14, 418, 420–1, 423, 474, 476–9, 482–4, 487–91
- case-adposition equivalence 236
- case-alternation 82, 288, 375, 391, 403, 412
- structural case 236, 256

<https://doi.org/10.1515/9783110686647-014>

- cliticization 22, 68, 99, 101, 122, 163, 198,
213–15, 218–20, 228, 266–7, 278, 290,
369, 385, 406, 408, 412, 431, 440, 487
– weak clitic pronoun 238
- coalescence 15, 111, 128, 139, 154, 158, 187,
375, 386, 398, 423, 457, 474, 478
- comitative 22, 146, 302, 304–5, 358, 370,
372, 385, 408, 410–11, 425
- conjunction 1, 5, 15, 35, 41–2, 71–2, 75,
96, 130, 149, 166, 200, 205, 278, 289,
380–1, 383–4, 406, 408, 414, 426, 445
– complex conjunction 42, 312
– secondary conjunction 403, 407–8, 414–5
- construction 9, 11–2, 15, 17, 19, 22, 33, 38,
44–5, 47–51, 55–9, 65, 100–1, 105–7,
112, 118, 122, 126, 129–33, 142, 144–6,
153, 157–61, 166, 169–74, 177, 186–7,
197–8, 212–3, 216, 227, 234, 236, 255,
257, 261, 275, 278, 288–9, 291, 317–25,
328, 335, 341, 348–9, 351, 353, 362,
367–8, 371–3, 375–7, 385–7, 390,
392–6, 398, 405, 411, 419, 422, 425–6,
443, 451–2, 454, 464, 476, 479–80,
482–3, 485, 487–8
– Construction Grammar 49, 443
– construction pattern 140, 209, 211, 460
– construction schema 131, 139, 166, 171,
178–9, 185, 492
- constructionalization 8, 15, 18, 49, 130, 132,
187, 478–9
- contact. *See* areal
- conversion 69, 110, 151–2, 167, 171, 407, 411
corpus
– corpus-based (study, analysis) 17, 33, 36,
139, 140, 179, 362
– diachronic corpus 130
– parallel corpus 20, 165, 265, 268, 279
- criteria (for adpositional status) 240–50
- dative 22, 35, 82–5, 100, 105–8, 111, 115, 117,
124, 145–6, 198, 235, 253–4, 259, 261,
267–71, 287–8, 345, 370, 374, 389, 393,
404–5, 407, 410–11, 413, 425, 429
- decategorialization 12, 20, 101, 111, 128–9,
280–1, 292, 299, 315, 317, 320, 322–9,
460, 474, 476, 478
- degrammaticalization 367–8, 385
- desemanticization 20, 280, 299, 314–15,
317, 393, 403, 419
- destinative 372
- diachrony 9, 12–13, 18, 21, 23–4, 34, 36, 47,
65–8, 73, 75–6, 78–9, 85, 89–90, 93,
96, 101–3, 111, 113–14, 126, 128, 130,
151, 153–4, 187, 233–4, 243, 251, 260,
265, 286, 289, 293, 305, 312–14, 351,
368, 370, 372, 375, 381, 384–5, 390,
398, 404, 407–8, 416, 418, 439, 444,
450, 464, 473–8, 481–2, 484, 490–1
- diagnostics for adpositional status. *See*
criteria
- distributed semantics 269–70
- egressive 347, 359
- elative 246, 302, 304
- ergative 370
- essive 302, 371
- functional equivalence 12, 40, 139, 161, 164,
186, 236, 281, 375
- genitive 13, 21–2, 38–40, 47, 53–4, 63, 76,
82–3, 85–8, 103–4, 106, 109, 111–12,
115, 117–18, 124, 145–7, 152, 156, 168,
175, 185, 187, 197–8, 208–9, 211–13,
216, 235–6, 238, 244, 249–51, 259,
267–9, 278, 287–8, 294, 302, 304, 306,
310, 312, 330, 335, 340–1, 345, 351–7,
359, 361–2, 368, 370–2, 374–5, 378,
389–93, 395–6, 403, 405, 408, 412–13,
420–3, 425–6, 433, 474, 477, 488
– genitive/nominative split 363
– genitive-dative 35, 270
- grammaticalization. *See also*
degrammaticalization
– grammaticalization by analogy 59, 131, 479
– grammaticalization cline 57, 300, 350, 373,
479, 481
– grammaticalization cycle 480
- illative 302, 360
- inessive 302, 490
- inflected prepositions 21, 35, 67, 99–101,
105–6, 110–12, 144, 195, 197–8, 201–7,
209, 211, 215, 217–20, 228, 234, 250,

- 260–2, 301–3, 307–8, 339, 344–5,
347–8, 350–5, 358–60, 362, 367, 369,
373, 385, 387, 392, 445–6, 463, 476–7,
480, 484
- initial consonant mutation 18, 195, 198, 201
- instrumental 22, 36, 145–6, 152, 177, 275,
287–9, 345, 358, 370, 374, 376, 381–2,
384–7, 393–4, 407–8
- lative 303, 307, 308
- lexical prepositions 49, 130, 271, 273
- lexicalization 8, 12, 15–17, 33, 50, 74, 88,
126, 128–30, 132, 153, 178, 187, 207,
288, 294, 299, 301, 303, 311–14, 319,
328–31, 375, 383, 385, 403–4, 412, 419,
422, 426, 457, 459, 461, 475, 477–9
- linking article 268, 378
- locative 22, 81, 84, 145–6, 175, 177, 195–6,
221–2, 224–5, 227–9, 245–6, 261, 269,
287–8, 302–3, 307–9, 316, 318–19, 322,
323–4, 328–9, 369–70, 372–3, 375–7,
382–3, 389–90, 394–7, 403–5, 410–11,
423–5, 488, 490–1
- marked 23, 38, 56, 84, 109, 115, 117, 124,
197–8, 212–13, 259, 261, 267–8, 275–6,
301, 341–2, 349, 351, 359, 370–1, 373,
375–7, 384–6, 390–6, 403, 405, 407,
410, 413, 424, 442–3, 447, 478, 488,
490
- markedness 392, 395, 425
- meaning 9, 11, 19, 21, 41, 54–6, 74–6, 80–4,
95–6, 116–9, 121, 125, 131, 133, 144–7,
150–2, 155, 159, 162–3, 168, 178, 187,
216, 221, 224–7, 233, 239, 243, 246–7,
250–1, 256–7, 260–1, 271, 275, 277–8,
289, 291, 307, 311–2, 315, 323–4, 330,
342, 344–7, 349–50, 357–8, 360, 367,
373–4, 377–82, 384, 388–90, 392–4,
398, 403, 409, 411, 413–5, 417, 419–22,
426–36, 441, 445, 454–5, 457, 460, 464,
477, 484, 489
- combinatory meaning 245
- referential meaning 280, 282, 478
- unit meaning 314, 317, 320–2, 325–8
- metaphorical extension 285
- metaphorical shift 403–4
- morphological affinity 367, 392
- motivative 367, 372, 380
- multi-word units 313–14, 319
- neologisms 23, 130, 403, 407, 408, 416–18
- nominal relator 10–12, 22, 24, 47, 90, 299,
302, 329–30, 476, 481, 483, 486–7
- complex nominal relator 1–24, 47, 90,
299, 302, 329–30, 473–4, 476, 481–3,
486–7, 492
- nominative 21–2, 84–5, 145, 212, 235,
267–8, 271–2, 289, 302, 304, 335,
344–5, 351–4, 356–8, 361–3, 403, 405,
412–13, 415, 425, 440
- nominative-genitive 412–13
- non-nominative 5, 84
- orthography 89, 102, 122, 124, 197, 236,
240, 254, 256, 442, 445, 480
- partitive 57, 235, 302, 304, 371, 376,
386, 390
- pattern (of formation)*. See construction
(pattern)
- perlative 240, 245–6, 454
- phrasal adposition 143, 250, 441, 460
- prolative 347, 371–2, 454
- polysemy 7, 13, 39, 75, 144, 262, 271, 273,
275, 289, 441, 487
- possession 17–18, 22–3, 33, 58, 87, 115, 160,
195, 198–9, 203, 213–15, 218–19, 223,
228, 235, 267–8, 271, 301, 335, 339,
342, 345, 348–50, 353–4, 395–6, 403,
405–9, 413–14, 418, 425–6, 430
- preposition stranding 78, 206
- preposition-case (construction) 144, 146,
171–4
- reanalysis 24, 48, 73, 98, 111, 115–16, 124,
130, 200, 216, 290, 328, 368, 377,
382–4, 411, 461, 462, 464, 474, 484
- relational noun 11, 21–4, 109, 200, 211–14,
216–17, 221, 283, 291, 345, 371, 374–6,
381, 398, 403, 409, 419, 423, 476–8,
487–8, 491

schema (of formation/of construction). See

construction (pattern)

separative 303, 307–8

stative 80, 82, 84

terminative 302, 304, 347, 370, 384–5,

413, 453

translative 302, 371

univerbation 13, 23, 69, 72, 74, 102, 106–7,

117, 122–5, 153, 240, 439, 441, 451,

457–62, 464, 474, 478–80

vocative 145, 235, 267, 268

Language Index

- Afro-Asiatic 439–40, 473
Albanian 19–20, 146, 265, 266–73, 279, 283,
286–94, 473, 482, 486
Altaic 22, 404
Anindilyakwa 491
Arabic 23, 34, 404, 407, 411, 416–9, 426,
432–3, 435, 444, 461
– Neo-Arabic 440
Arrernte 489
Austronesian 487
- Basque 21, 367–75, 377–2, 385–93, 395–8,
473, 476–7, 480, 482, 486
– Southern Basque 389
Bizkaian 368
Bosnian 141
Breton 19, 195–209, 211, 213, 216–7, 219–25,
228–9, 474, 480, 486
– Central Breton 218
Brythonic 18, 195–6, 198, 213, 223
Bulgarian 140–1, 266
- Celtic 18, 34, 195–201, 205–6, 211, 223, 228,
473–4, 480
Chuvash 362
Cornish 196–7
Creole 34
Croatian 140–1
Czech 140–2, 164
- Dutch 2–3, 17, 43, 59, 65–74, 76–82, 84–91,
93–97, 99–100, 104–8, 111–2, 114,
117–20, 122, 125–6, 131–4, 241, 254,
258, 484, 486
- English 1–2, 6, 8, 10, 14, 17–9, 36,
38, 43, 65–79, 81–3, 85–91, 93–9,
101, 106, 108–10, 113–4, 118, 120,
122–6, 130–3, 150, 165–6, 179,
217, 223, 233–4, 237, 240–3, 248,
250, 252, 254–6, 258–9, 269, 286,
331, 369, 371–2, 375, 381, 385, 411,
422–3, 425, 440–2, 445, 447, 482,
484, 486
– American English 67, 121, 422
– British English 68
Erzya 361
Estonian 20–1, 299–307, 309–15, 317,
320–1, 324, 328–31, 473, 476–8,
482, 486
- Finnic 300, 305
Finno-Ugric 301, 305, 361, 473
French 2, 4–9, 17, 19, 33–9, 41–4, 47, 49–51,
53–5, 57–9, 114, 132, 150, 157–8, 165,
169, 171, 176, 179, 195–6, 200, 208–10,
217, 225, 241–2, 250, 252–6, 258, 286,
291, 293, 368–9, 372, 376, 396, 417,
426, 474, 479, 480, 485–6
– Classical French 49, 58
– Medieval French 50, 58, 479
– Middle French 49, 485
– Modern French 38, 42, 44–7, 49–51, 58–9,
180, 225, 482, 485
– Old French 38, 46–9, 59, 225, 293, 482
– Pre-Classical French 49
- Gaelic 196, 198, 206, 212
Gascon 368
German 3, 7, 10, 13, 14, 17, 38, 43, 65–72,
74, 76–9, 81–3, 85–91, 93–95, 97–104,
106–7, 109–12, 114–9, 122, 124–6,
132–4, 146, 243, 250, 252, 255–6, 258,
269, 286, 294, 301, 374, 376, 396, 445,
476, 480–2, 484, 486
Germanic 17, 34, 49, 102, 105, 109, 134, 293,
362, 473–4, 479–80, 485–6
– West Germanic 17–8, 65–7, 72, 76, 85, 87,
89, 102, 109, 114–5, 157, 475, 478
Gheg 266, 271, 277, 287
Greek 19, 233–8, 240–6, 250–2, 254–8,
260–2, 266–7, 290, 294, 473–5,
482, 486
– Ancient Greek 19, 235, 240, 243–4, 251,
257, 261, 484
– Classical Greek 234, 236, 260, 269
- Hungarian 300, 361, 487, 489

<https://doi.org/10.1515/9783110686647-015>

- Ibero-Romance 53–4, 59, 475
 Indo-European 31, 34, 140, 141, 144, 151,
 200, 234–5, 260, 265–6, 269, 286–7,
 288, 290, 294, 297, 301, 473, 484
 Irish 121, 196, 200
 Italian 2–3, 7, 35–6, 42–3, 52–5, 57–9, 195,
 242–3, 252, 255–6, 258, 289, 368, 371,
 439–40, 449, 462, 468, 481
- Jaminjung 488
 Japanese 22, 404, 488
- Kashubian 141
 Katharévoussa 234, 257
 Komi 21, 335–7, 339, 341–4, 346–7, 349–51,
 354, 356–62
 Komi–Zyrian 339, 343, 346–7, 350,
 356–8, 360
 Korean 22, 404, 487, 489–90
 Kutenai 491
 Kwa 487
- Latin 34, 47, 53, 57, 59, 200, 209, 223,
 225–6, 228, 234, 242–3, 261, 266, 288,
 290–1, 294, 381–2, 416, 484–6
 – Classical Latin 47, 265
 Likpe 487
- Maasai 488–9
 Macedonian 141, 266
 Malagasy 487
 Maltese 23, 439–42, 444–8, 450–2,
 455, 457–8, 460–4, 468, 473, 475,
 482, 486
 Manx 196, 206
 Mari 361
 – Hill Mari 362
 Mayan 487
 Mongolian 22, 404
 Montenegrin 141
 Mordvinian 361
- Niger-Congo 487, 490
 Nilotic 488
- Occitan 2, 35, 52–4, 368
 Ottoman 416
- Pama-Nyungan 489, 491
 Permic 21, 335–7, 339, 342–6, 348, 350–1,
 359, 361–3, 473, 476–7, 482, 486, 491
 Permyak 336, 339–40, 342, 360–1
 – Southern Permyak 361
 Persian 23, 416–7, 426, 428
 Polish 7, 140–1, 241–2, 250, 252, 255–6, 258
 Portuguese 2, 6, 9, 17, 33–9, 41–2, 52–5,
 57–8, 165, 368, 484
- Romance 2, 17, 33–37, 43, 47, 52–9, 225–6,
 228, 234, 242, 261, 267, 292, 293, 362,
 368–9, 371, 382, 473–5, 479–80, 485–7
 – Central Romance 35
 – Gallo-Romance 53
 – Italo-Romance 54, 475
 Romanian 2, 17, 33, 35–8, 43, 48, 52–4,
 56–7, 59, 208, 242, 250, 266, 482–3
 Russian 1, 140, 142, 164, 301, 335, 337, 343,
 345, 361, 363, 486
- Samoan 487
 Scottish 196, 198, 206, 212
 Semitic 440, 444, 449, 462
 Serbian 4, 7, 13, 18, 139–52, 154, 156–8,
 161–72, 178–87, 192, 266, 268, 475,
 479, 482, 484–6
 Silesian 141
 Sino-Tibetan 487
 Slavic 14, 18, 34, 139–42, 151–2, 155, 186–7,
 266–7, 293–4, 473, 479
 – South-Slavic 139
 – West Slavic 140
 Slavomolisano 141
 Slovak 141
 Slovenian 141–2
 Sorbian
 – Lower Sorbian 141
 – Upper Sorbian 141
 Southern Min 488
 Spanish 1, 5, 17, 34–5, 37, 43, 52–4, 57, 165,
 242, 289, 368, 372, 382
 Swedish 241–2, 252, 255–6, 258, 301, 486
- Tabasaran 488–9
 Tagalog 487, 489
 Tatar 353, 361

- Tibeto-Birman 487
Tosk 266
Turkic 22, 335–7, 353, 361–3, 404, 416–7
Turkish 22–3, 266–7, 403–9, 412, 414,
416–9, 422–3, 425–7, 473, 476–7,
482, 486
Tzeltal 487
- Udmurt 21, 335–7, 339–62, 476
– Beserman Udmurt 336, 339–49, 351–5,
358–60, 476
– Central Udmurt 336
– Northern Udmurt 336
– Southern Udmurt 336
Ukrainian 140
Uralic 21, 300, 336, 344–5, 487
- Welsh 19, 196, 200–2, 216, 221, 223–4, 480
- Yazva 21, 336, 339, 343–4, 357–8
Yidiny 491
- Zuberoan 368
Zyrian 336, 339, 341–2

