



HAL
open science

Complex prepositions in Albanian: A first assessment

Guillaume Bonnet, Benjamin Fagard

► **To cite this version:**

Guillaume Bonnet, Benjamin Fagard. Complex prepositions in Albanian: A first assessment. Complex Adpositions in European Languages – A Micro-Typological Approach to Complex Nominal Relators, De Gruyter Mouton, pp.265-296, 2020, 10.1515/9783110686647-007 . halshs-03087879

HAL Id: halshs-03087879

<https://shs.hal.science/halshs-03087879>

Submitted on 24 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Guillaume Bonnet and Benjamin Fagard

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 possible origin of CAs in Albanian, showing a possible effect of language contact.

Keywords

Albanian, corpus linguistics, Balkan languages, contact linguistics, prepositions, case

Guillaume Bonnet
University of Bourgogne
15bis, rue du Pot-de-Fer, 75005 Paris, France
Guillaume.Bonnet@u-bourgogne.fr

Benjamin Fagard
CNRS, ENS & Sorbonne Nouvelle; PSL
Lattice laboratory, Ecole Normale Supérieure, 1 rue Maurice Arnoux, 92120 Montrouge,
France
benjamin.fagard@ens.psl.eu

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).

1 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

¹ 6 million by Ethnologue counts (Eberhard et al. 2019), which we used for the number of speakers of the different Albanian varieties.

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).

Table 1: Shared grammatical innovations in the Balkans (Lindstedt 2000: 232)

	Greek	Albanian	Balkan Slavic	Balkan Romance	Balkan Romani
<i>Argument marking</i>					
ENCLITIC ARTICLES	(+)	+	+	+	(+)
OBJECT REDUPLICATION	+	+	+	+	+
PREPOSITIONS INSTEAD OF CASES	(+)	(+)	+	(+)	(+)
DATIVE / POSSESSIVE MERGER	+	+	+	+	–
GOAL / LOCATION MERGER	+	+	+	+	(+)
<i>RELATIVUM GENERALE</i>	+	+	+	+	+
<i>Verb system</i>					
AUX (+ COMP) + FINITE VERB	+	(+)	+	(+)	+
<i>VOLO</i> FUTURE	+	(+)	+	+	+
PAST FUTURE AS CONDITIONAL	+	+	+	(+)	(+)
<i>HABEO</i> PERFECT	+	+	(+)	(+)	–
EVIDENTIALS	–	+	+	(+)	(+)
<i>Other</i>					
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.

2 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).

2.1 Case marking

2.1.1 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)

2.1.2 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* *e* *fundit*,

toward DET dawn.ACC when send.IND.AOR.3PL telegram.ACC DET.ACC.SG end.DET.GEN
ata ishin ende duke pirë.
 they.NOM be.IND.AOR.3PL still ADV.GER drink
 2

‘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

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*

² Gerund-like adverbial form.

- 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 phrase with one case only, most frequently the ablative, as indicated in Table 2.

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>jashë</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’	

³ This is a possessive use of the determiner.

2.2 Simple prepositions

2.2.1 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 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 *pmw* in both corpora (*tatëpjetë* ‘downward’, *tinëz* ‘furtively’ and *përposh* ‘below’, as well as the nouns *brinjë* ‘ribs’ and *vithë* ‘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’

2.2.2 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*

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.

‘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 (see details in Appendix).

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>në bazë</i> + GEN/ABL ‘sur la base de’; <i>në buzë</i> ‘alongside (lit. ‘on the edge of’); <i>krahas</i> + ABL ‘near, along with’; <i>me anën</i> + GEN ‘near, alongside’
< 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’

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).

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’
< 1	<i>fill e</i> + ABL ‘straight at’

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 compounds, 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 [Prep: ‘meaning_i’] → [*për* Prep: ‘meaning_{i/j}’] construction, 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 e njëjta.
yet earth.DET.F.NOM.SG around house.PL.OBL.DEF be.IND.IMP.3SG 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,
bark.IND.IMP.3PL also dogs.NOM DET.NOM.PL old.towers.GEN near stream.det.obl
lehnin qentë rreth e **qark** katundit...
bark.IND.PRS.3PL dog 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 (16). In this case, synchronically, there is thus a clear functional and semantic difference between the simple form *mjet* ‘means’, used only as a noun (17), and the compounds *ndërmjet* ‘between, among’, *përmjet* and *nëpërmjet* ‘by means of, through’ (18), used as prepositions.

- (17) Edhe muzika mund të bëhet një **mjet** torture.

also music can.ind.prs.3sg comp become.ind.prs.3sg one means torture.gen
 ‘Music can also become an instrument of torture’ (ANC, *Gazeta Shqiptare*, 2007.01.05)

- (18) 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’.

2.3 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’; Ghëg *deri*) and *ngjithë* (participle of *ngjit* ‘to attach’). The resulting sequences include the following:

- (19) *sa për mua*
 QUANT for me.ACC
 ‘as far as I am concerned’

- (20) *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.

3 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 “connexion 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.).

3.1 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)	<i>Kam</i>	<i>parë gjithashtu nën diell që në vend të arsyes</i>
	have.IND.PRS.1SG	see.PST.PTCP also under sun comp in place DET reason.GEN
	<i>kishte</i>	<i>paudhësi që në vend të drejtësisë</i>
		,
	have.IND.IMP.3SG	iniquity, COMP in place DET justice.GEN
	<i>kishte</i>	<i>paudhësi.</i>
	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 to the grammaticalization of CPs, i.e. of 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’.

(27)	<i>Mbi qiellin e Zajazit do të kalonte</i>
	over sky.ACC DET Zajaz.GEN PRT ⁵ PRT pass.IND.IMP.3SG
	<i>një skuadër aeroplanësh luftarak, që fluturonte</i>
	one squadron airplane military comp fly.IND.IMP.3SG
	<i>në drejtim të malit Drenovë.</i>

⁵ *Do* is a particle used to form the future tense; *të* is used here as a particle to form the subjunctive.

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) on the ANC corpus, lexeme “në” + lexeme “bazë”; (b) on the OPUS corpus, [word="(n|N)ë"] [word="bazë"]. Semantic bleaching is harder to identify; we only excluded sequences which seemed do 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).

(28) Por ai ishte i vetmi që pati guximin ta
 but PRO.3SG be.IND.IMP.3SG DET only PRO have.IND.AOR.3SG courage.ACC COMP
 refuzojë në mënyrë të qetë dhe dinjitoze ta
 refuse.SBJ.PRS.3SG in manner DET calm and dignified COMP
 braktisë besimin e vetë.
 abandon.SBJ.PRS.3SG faith DET PRO.3SG

‘But he was the only one who had the courage to refuse in a calm and dignified fashion to abandon his faith’ (ANC, *Spastrimet etnike. Politika gjenocidale serbe ndaj shqiptarëve* (Shqyrtime, komente, elaborate, dokumente) Hivzi Islami 2000–2001)

3.2 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
<i>në përgjigje të</i> ‘in response to’	2	11
<i>në periferi të</i> ‘on the periphery of (+/- met)’ (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’ (end)	5	1
<i>në emër të</i> ‘on behalf of, in the name of’	82	26
<i>në favor të</i> ‘in favor of’ (favor)	34	39
<i>në fund të</i> ‘at the end of’ (end)	182	291
<i>në (sh)këmbim të</i> ‘in exchange for’ (exchange)	8	17
<i>në krye të</i> ‘at the end (lit. head) of’ (end)	83	88
<i>në mbështetje të</i> ‘in favour of, in support of’ (favor)	12	35
<i>në mes të</i> ‘in the middle of’ (middle)	141	64
<i>në përfitim të</i> ‘in exchange for, in return for’ (exchange)	0	8
<i>në përputhje me (/të)</i> ‘in compliance with’	43	20
<i>në pikë të</i> ‘in the middle of’ (middle)	3	0
<i>në qendër të</i> ‘at the center of’ (middle)	49	50
<i>në zemër të</i> ‘at the heart of’ (middle)	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
qenë 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
in name DET redaction DET weekly Focus in.front.of colleague.OBL.DET.PL
dhe miqve të pranishmëve foli bashkëpunëtori i
and friend.OBL.DET.PL DET present.OBL.PL speak.IND.AOR.3SG colleague.NOM.SG DET
tij në këtë gazetë, Shase Dimevski, i cili evokoi
PRO in this.ACC.SG newspaper, Shase Dimevski, DET PRO evoke.IND.AOR.3SG
momente nga puna e përditshme me të.
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
 now several.ACC.PL year.ACC.PL reign.IND.PRS.3SG one climate.NOM
 ndërtimesh dhe mbindërtimesh pa leje mu
 construction.OBL.PL and over.construction.OBL.PL without permission.ACC until
në zemër të qytetit
 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

- (32) Stavreski, ka qenë të hënën në qendër të akuzave
 Stavreski AUX.IND.PRS.3SG be.PTCP DET Monday in center DET accusation.OBL.PL.DET
të opozitës, të cilët iniciuan votëbesim
 DET opposition.GEN DET PRO initiate.IND.IMP.3SG vote.of.confidence.acc
 për punën e tij.
 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, datë
 in support DET paragraph.OBL 16 DET law.OBL number 7491 date.ACC
 29.4.1991 «Për dispozitat kryesore kushtetuese», propozim të
 4.29.1991 about disposition.ACC main.ACC constitutional.ACC proposal.NOM DET
 Këshillit të Ministrave, Kuvendi popullor i Republikës
 council.OBL DET 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 nuk e
 commune PRT AUX.IND.PRS.3SG find.PTCP solution.ACC.DET COMP NEG DET

⁶ *Mu* is glossed ‘bis’ (i.e. ‘until’) by Meyer (1891), who indicates that it is a preposition.

lë *më* *të* *varur* *nga* *ujësjellësi* *i*
 leave.IND.PRS.3SG more DET dependent from water.line.OBL in
cili *reduktimet* *nuk* *na* *i* *hoqi* *as*
 which reduction NEG CL.1PL.OBL CL.ACC remove.IND.AOR.3SG not.even
në *pikë* *të* *dimrit*.
 in heart DET winter.OBL

‘The solution found by the municipality is such that we are no longer dependent on the water supply, which did not fail us even in the dead of winter’ (ANC, *Zëri*, 2013.03.16)

Some sequences, while they display some variation, still seem to warrant an analysis as CPs, for instance *në shkëmbim të / në këmbim të* ‘in exchange for’, which are found in comparable frequencies (this is true for both corpora).

As far as semantic extension is concerned, there are both sequences with evidence of metaphorical extension, for instance *në periferi të* ‘on the periphery of’ (35) and sequences which seem very limited in terms of context of appearance, for instance *në fillim të* ‘at the beginning of’, which is generally found with a temporal complement (36) – year, century, month; also e.g. *në fillim të viteve* ‘80’ ‘in the early 80s’.

(35) Vedat Kokona e paraqet gruan si figurë që
 Vedat Kokona CL.ACC present.IND.PRS.3SG woman.ACC.DET as figure.ACC COMP
jeton *në* *periferi* *të* *jetës*
 live.IND.PRS.3SG in periphery DET life.OBL

‘Vedat Kokona presents the woman as a figure living in the periphery of life’ (ANC, *Femina litterarum: Romanticism, sentimentalism, symbolism, realism, expressionism. Resmije Kryeziu 2000-2009*)

(36) Filluan që në janar dhe s’pushuan gjer
 start.IND.AOR.3PL since in January and NEG.stop.IND.AOR.3PL until
në *fillim* *të* *qershorit*.
 in beginning DET June.OBL

‘They started in January and did not stop until the beginning of June’ (ANC, *Trëndafili në gotë – fragment*, Dritëro Agolli, 1970–1979)

3.3 Areal effects?

An interesting result appears from the contrast between the data found in the ANC corpus and the OPUS corpus. Some CPs are much more frequent in the former, others in the latter. For instance, the sequence *në shkelje të* ‘in violation of, in breach of’ is relatively frequent in OPUS and almost absent from the ANC; conversely, *në bazë të* ‘on the basis of’ and *në kuadër të* ‘in the framework of’ are much more frequent in the ANC. There seems to be a link between the semantics 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).

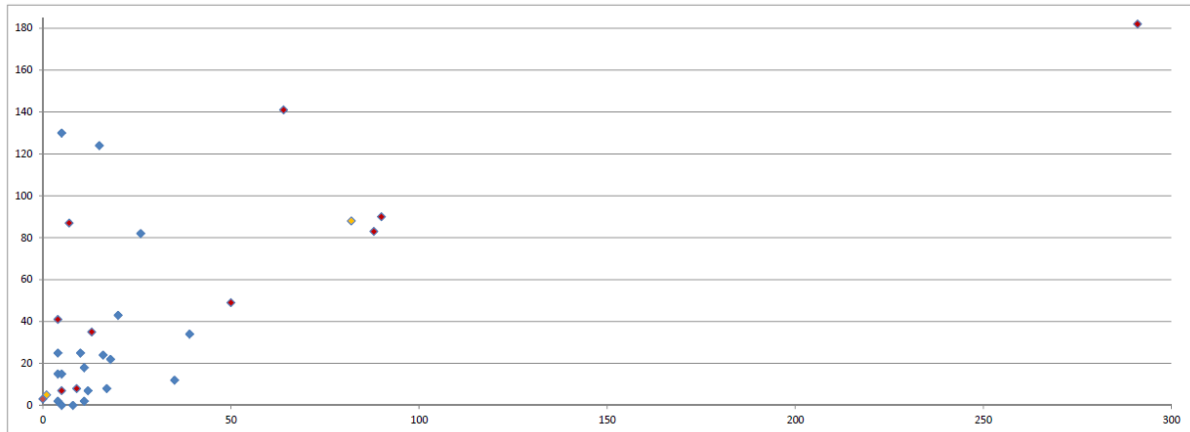


Figure 1: Relative frequency (pmw) of CPs in the ANC (on the horizontal axis) and OPUS (vertical axis) corpora, depending on the semantic nature of the nucleus: spatial (red), temporal (yellow), other (blue).

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 a nice illustration of the tendency for CPs to appear in European languages as a result of language contact (cf. Van der Horst 2013, Hüning 2014).

4 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]).

4.1 Case marking: moving away from the Indo-European case-marking system

4.1.1 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 locative⁷. 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.

4.1.2 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)	<i>endeh</i>	<i>stepii</i>	<i>teh</i>	<i>atit</i>	<i>tim</i>
	in	house.ACC	DET	father.DET.OBL	my.OBL
	‘in my father’s house’				

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”.

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*.

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’

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 anpacken.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.

4.1.3 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] sequence⁹. 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** + 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)*

⁹ As e.g. with Spanish *según* in *según yo* ‘according to me.NOM’ (where one could expect *mi* ‘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)).

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)’

4.2 Simple prepositions and compounds

4.2.1 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. *epí*) (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 compound 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).

4.2.2 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 the 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)

¹⁰ This was also the idea of the great Albanian specialist Norbert Jokl (1877-1942), as reported by Sandfeld (1930: 158).

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 (*largohem tinëz* ‘I took French leave (lit. *leave.PST.ISG 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 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’

¹¹ See e.g. *jashtë*, etymologically related to Latin *ex* but with a final element *-të* which is difficult to account for precisely.

- (48) *banoj afër*
live.PRS.1SG close
'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

4.2.3 *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 section¹². 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).

4.3 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

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 bainō káta* ≡ *katà óreos bainō* ≡ *kata-bainō óreos*).

(i) *óreos bainō káta*
mountain.G.SG walk.1SG down

'I come down from the mountain.'

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 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.

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	Gr.	Greek
ACC	accusative	IE	Indo-European
ADV	adverb	IMP	imperative
AOR	aorist	IND	indicative
AUX	auxiliary	Lat.	Latin
CL	clitic	M	masculine
COMP	complementizer	NEG	negative
DAT	dative	NOM	nominative
DEF	definite	OBL	oblique
DEM	demonstrative	PL	plural
DET	determiner	PRET	preterite
F	feminine	PRO	pronoun
GEN	genitive	PRS	present
GER	gerund	PRT	particle

PST	past	REF	reflexive
PTCP	participle	SBJ	subjunctive
QUANT	quantifier	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 Lafe & 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 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.
- 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. Belgrad.
- 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. *Slavische 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.