

Comparing typologies

MMM13 Online Proceedings

Edited by:

Jenny Audring
Katerina T. Frantzi
Nikos Koutsoukos
Giorgos Markopoulos
Kalomoira Nikolou

PUBLISHED AT UNIVERSITY OF PATRAS
2023



Comparing typologies

MMM13 Online Proceedings

Edited by:

Jenny Audring

Katerina T. Frantzi

Nikos Koutsoukos

Giorgos Markopoulos

Kalomoira Nikolou

Online Proceedings of the
13th Mediterranean Morphology Meeting (MMM13)
Rhodes (Greece), May 19-22, 2022

PUBLISHED AT UNIVERSITY OF PATRAS
2023

The MMM Permanent Committee

Jenny Audring (Leiden University)

Nikos Koutsoukos (University of Patras)

Francesca Masini (University of Bologna)

Online proceedings of the Mediterranean Morphology Meetings (MMM)

ISSN: 1826-7491

Table of contents

Foreword	v
Polysynthesis: lessons from Northwest Caucasian languages Peter Arkadiev.....	1
Motivating a Morpheme: Albanian case syncretism as a case study Đorđe Božović.....	27
Locative forms in Nakh-Daghestanian as an example of a transcategorial paradigm Marina Chumakina.....	37
Root reduplication and alignment overcome three challenges to the biradical, OCP-based analysis of Semitic QaTaT stems Noam Faust	47
The productivity of adverbs and adverbials in Modern Hebrew Malka Muchnik	57
Where did the Italian Verbal-Nexus N+N compounds come from? Jan Radimský	71
A typological comparison of infixes and circumfixes Tim Zingler	83

Foreword

The first Mediterranean Morphology Meeting (MMM) was held in Mytilene, Greece, in 1997. Since then, the conference has been organized biennially in various locations around the Mediterranean. The original founders and organizers were Geert Booij (Leiden University), Angela Ralli (University of Patras), and Sergio Scalise (University of Bologna). As of 2013, organization is in the hands of Jenny Audring (University of Leiden), Nikos Koutsoukos (University of Patras) and Francesca Masini (University of Bologna).

The aim of MMM is to provide a focused and informal setting for morphologists to present and discuss their work. The single-session setup guarantees maximal interaction between researchers, and gives young linguists the chance to present their work at a conference of moderate size, where fruitful contacts with senior linguists can be established. Twelve meetings – in 1997 (Mytilene, Greece), 1999 (Lija, Malta), 2001 (Barcelona, Spain), 2003 (Catania, Sicily), 2005 (Fréjus, France), 2007 (Ithaca, Greece), 2009 (Nicosia, Cyprus), 2011 (Cagliari, Sardinia), 2013 (Dubrovnik, Croatia), 2015 (Haifa, Israel), 2017 (Nicosia, Cyprus) 2019 (Ljubljana, Slovenia) and 2022 (Rhodes, Greece) – have proven the success of this formula.

In good tradition, this volume continues the MMM Online Proceedings series with a selection of papers presented at MMM13, which took place May 19-22, 2022 in Rhodes (Greece). MMM13 was made possible thanks to the excellent local organizing committee chaired by Katerina Frantzi and led by Giorgos Markopoulos (University of the Aegean). The topic of the conference was “Comparing morphologies: Typological generalizations at work”. Keynote speakers were Laura Michaelis (University of Colorado at Boulder) and Peter Arkadiev (Russian Academy of Sciences, Moscow). We are especially pleased that Peter Arkadiev was able to present his work despite the difficult circumstances, and to have his contribution included in the proceedings. The conference was preceded by a workshop on “Semitic morphology” with Noam Faust (Université Paris 8, CNRS SFL) as a keynote speaker.

The editors of this volume wish to thank the local organizing committee consisting of Katerina Frantzi (Chair of the committee), Marianthi Georgalidou, Hasan Kaili, Eleni Karantzola, George Kotzoglou, Giorgos Markopoulos and Kalomoira Nikolou, all student volunteers who made sure the conference ran smoothly, all attendants of MMM13, and especially the contributors to these Online Proceedings.

Polysynthesis: lessons from Northwest Caucasian languages

Peter Arkadiev
Johannes-Gutenberg-Universität Mainz
alpgurev@gmail.com

1. Introduction

Polysynthesis has fascinated linguists ever since polysynthetic languages, characterised by exceptional morphological complexity of verbs, have come to their attention. However, despite considerable advances in the study of polysynthesis in the last few decades (e.g. Mithun 1988, Baker 1996, Evans & Sasse 2002, Mattissen 2004, 2006, Mahieu & Tersis 2009 and Fortescue et al. 2017), there is still no consensus as to how this notion should be defined and even whether it lends itself to a clear-cut definition at all, and, concomitantly, whether the class of “polysynthetic languages” can be delimited in a meaningful way (Zúñiga 2019).

Nevertheless, its problematic status notwithstanding, the notion of polysynthesis has proven useful for the advancement of typology and linguistic theory in that the study of polysynthetic languages has both allowed linguists to better understand a variety of apparently “exotic” phenomena, such as head-marking and polypersonalism, incorporation, “lexical affixation”, templatic organisation of morphology and others, and offered new insights into the fundamental questions concerning the relations between morphology and syntax, inflection and derivation, lexical storage and online production etc.

While not attempting to provide my own solutions to the problems of definition and delimitation of polysynthesis, in this article I shall first review the definitions of polysynthesis and its characteristic features proposed in the typological literature (section 2), then briefly introduce the major parameters of typological variation in polysynthetic morphology (section 3), and finally present an overview of the polysynthetic properties of the Northwest Caucasian languages focusing on how they fit into the typological classifications proposed (section 4).

2. Delimiting polysynthesis

The term “polysynthetic” was coined by the French-American linguist and philosopher Peter (Pierre-Étienne) Duponceau (1760–1844) to refer to language structures “in which the greatest number of ideas are comprised in the least number of words” (Duponceau 1819, quoted after Zúñiga 2019: 1). Initially, the term was applied to the native languages of North America, e.g. Mohawk (1)¹, then to Chukotkan languages (2), but later similar structures were found in many other languages of diverse geographical regions, e.g. Dalabon in Australia (3).

¹ Abbreviations used in glosses: 1 — 1st person; 2 — 2nd person; 3 — 3rd person; A — agent; ABS — absolutive; ACC — accusative; ADD — additive; ADV — adverbial; AND — andative; BEN — benefactive; CAUS — causative; CMPL — completive; CNT — continuative; COM — comitative; COND — conditional; COORD — coordinative; CSL — cislocative; DAT — dative; DCL — declarative; DEB — debitive; DEF — definite; DIST — distal demonstrative; DPR — depreciative; DU — dual; DUP — duplicative; DYN — dynamic; EMP — emphatic; ERG — ergative; EXC — excessive; F — feminine; FACT — factive; FUT — future; GEN — genitive; H — human; INCL — inclusive; IND — indicative; INT — intensive; INTR — intransitive; IO — indirect object; IPF — imperfect; ITER — iterative; LAT — lative; LOC — locative preverb; M — masculine; MSD — masdar; N — non-human; NEG — negation; NFIN — nonfinite; NPST — nonpast; OBJ — object; OBL — oblique; PASS — passive; PFV —

- (1) Mohawk (Iroquoian, Canada; Mithun 2017: 236)
ó:nenkati' ken t-en-s-ite-wa-htenno-'ók-h-a-'?
 now then Q DUP-FUT-REP-1INCL.A-PL-ball-hit-AND-PURP-PFV
 'How about we go there and play some golf?'
- (2) Chukchi (Chukotkan, Russia; Skorik 1961: 103)
tə-tor-taŋ-pəlwəntə-pojgə-pela-rkən
 1SG-new-good-metal-spear-leave-PRS.1SG
 'I am leaving a good new metal spear.'
- (3) Dalabon (Gunwinyguan, Australia; Evans 2017a: 769)
kah-marne-yerrûh-ye-rrudjm-inj
 3SG>1SG-BEN-ITER-COM-return-PST.PFV
 'He kept bringing them back for me.'

Notably, not only has “polysynthesis” been introduced as a holistic notion aimed to capture the overall “character” of a language, like many typological terms of the 19th century, but it has largely remained so up to now. Indeed, while modern linguists have largely given up the habit of speaking about e.g. “ergative” languages, obviously due to the recognition of the fact that most languages show a mixture of different alignment types in their morphosyntax (Bickel 2011), the phrase “polysynthetic language” still belongs to linguistic parlance. The reason for this must be the (often implicit) belief of linguists that “polysynthesis” is a typological trait with repercussions in many domains of linguistic structure (e.g. Mithun 1988, Baker 1996, Fortescue 2007).

How can “polysynthesis” be defined beyond the rather impressionistic, even if insightful, characterisation by Duponceau? The most straightforward way to lend precision to Duponceau’s description was proposed by Greenberg (1960: 194), who introduced a quantitative “index of synthesis”, the morpheme-to-words ratio in a 100-words long text, and defined as polysynthetic the languages with an index of synthesis of 3.0 or more. Simple as this definition might seem, it is unsatisfactory for two interrelated reasons (cf. Sadock 2017). First, the cut-off point of 3.0 morphemes per word is purely arbitrary; second, the high morpheme-to-words ratio in and of itself does not reveal anything about the morphological structures making it possible. It is therefore not surprising that since Greenberg, linguists working on polysynthetic languages have been focusing on qualitative rather than quantitative characteristics. For example, Mithun in her seminal article, starting with a simple statement that “[p]olysynthetic languages are by definition those that exhibit a high number of morphemes per word” (Mithun 1988: 442) ends with the insight that “[t]he propensity of polysynthetic languages to develop morphological complexity primarily within their verbs can have a subtle effect both on the semantic nature of the categories they grammaticize, and on the resulting structure of their morphological systems” (Mithun 1988: 451–452). It is clear that “[p]olysynthetic languages are more than just languages with very long words” (Fortescue 2016: 2) and must differ qualitatively, not just quantitatively, from non-polysynthetic ones.

perfective; PL — plural; POSS — possessive; POT — potential; PR — possessor; PRED — predicative; PRF — perfect; PROX — proximal; PRS — present; PST — past; PTCL — particle; PURP — purposive; PVB — preverb; Q — interrogative; RE — reffective; REC — reciprocal; REL — relativizer; REP — repetitive; RFL — reflexive; RPST — remote past; RR — reflexive/reciprocal; RSTR — restrictive; S — single argument of canonical intransitive verb; SBD — subordinator; SBJ — subject; SEQ — sequential; SG — singular; SML — similitive; TEMP — temporal; TRL — translocative; VERT — motion upwards.

That said, there is still no general and agreed-upon definition of polysynthesis, and the very notion and the class of languages it is supposed to characterise lack clear-cut boundaries. This becomes apparent if we consider some of the qualitative definitions proposed in the literature (setting aside Baker's (1996) definition or rather redefinition of polysynthesis within the generative framework). Thus, Evans & Sasse (2002: 3) speak about "a prototypical polysynthetic language" as one "in which it is possible, in a single word, to use processes of morphological composition to encode information about both the predicate and all its arguments <...> to a level of specificity allowing this word to serve alone as a free-standing utterance without reliance on context". Likewise, Fortescue (2017: 122) requires of a "core polysynthetic" language to "display holophrasis (i.e. be able to represent a whole clause – including all bound core pronominals – by a single word)" as well to "allow more than one lexically 'heavy' morpheme within the holophrastic verb, whether it be lexical or affixal". Fortescue underscores that polysynthesis cannot be reduced to head-marking (Nichols 1986, 2017) alone and "contains an essential derivational component" (Fortescue 2016: 6), i.e. semantically loaded morphological processes altering and modifying the verb's meaning. Indeed, there are languages such as Basque or Kinyarwanda, which exhibit even more head-marking than such "core" polysynthetic languages as Yupik, but have never been considered polysynthetic precisely for their lack of this derivational component.

The latter notion is central in De Reuse (2009), who introduced the concept or "productive non-inflectional concatenation" (PNC), i.e. optional and formally transparent affixes expressing semantic content often bordering on lexical, and especially in Mattissen (2004, 2006, 2017), who defines polysynthetic languages as those which "have complex, polymorphemic verbal units which necessarily integrate productive<ly> non-root bound morphemes with 'lexical' and grammatical meanings [...] and optionally allow concatenation of lexical roots within a verbal wordform" (Mattissen 2017: 72). By contrast, head-marking and holophrasis are not wrought into Mattissen's definition and are treated as parameters of variation.

What can be distilled from these characterisations of polysynthesis is a cluster of morphological properties such as head-marking and polypersonalism, productive and optional semantically loaded derivational morphology, incorporation and composition in general, as well as more concrete features such as use of applicatives in the verb to fulfill the function of nominal case (see Mithun 1988, Fortescue et al. 2017). Different languages traditionally conceived of as polysynthetic show different constellations of these properties, and some of the polysynthetic properties can be found, even if to a limited extent, in languages that linguists have never included into this class, e.g. in Lithuanian (4) (see e.g. Arkadiev 2021).

- (4) Lithuanian (Indo-European, CCL)
ne-be-su-si-tik-inėj-o-me
 NEG-CNT-together-RR-meet-ITER-PST-1PL
 'we did not (repeatedly) meet each other any more'

The existing definitions of polysynthesis crucially rely upon on the notion of "word", whose validity and consistency as a cross-linguistic concept have been repeatedly put to doubt (e.g. Haspelmath 2011, Tallman 2020). "Wordhood" has been shown to raise considerable methodological and analytical problems at least in some polysynthetic languages as well (e.g. Bickel & Zúñiga 2017, Tallman 2021), in particular with respect to mutual correlations (or rather lack thereof) between phonological and morphosyntactic criteria defining the relevant syntagmatic domains. Unless the boundaries of the "word" or at least of the "verb" are robustly defined, speaking about its morphological complexity is hardly a meaningful endeavour.

Do these internal heterogeneity, fuzzy boundaries and problematic foundations lend the very notion of polysynthesis “at best an impressionistic label, and at worst an ill-defined buzzword, without much practical usefulness” (Zúñiga 2019: 15)? Perhaps they do, but one can still conceive of polysynthesis as not entirely hopeless by trying, on the one hand, to make the notion more restrictive (as suggested by Zúñiga 2019: 14–15), and, on the other, by looking more deeply into the individual characteristics associated with polysynthesis and exploring their cross-linguistic variation as well as their possible mutual correlations and their relations with other relevant properties of grammar and lexicon.

3. Typologising polysynthesis

Languages traditionally considered polysynthetic differ widely along many parameters, both quantitative and qualitative. These include average or maximal number of morphemes per word or slots in a morphological template as well as number of paradigmatically opposed affixes for each slot, number and type of semantic features grammaticalised, maximal number of participants expressed by pronominal affixes on verbs, types of arrangement of morphemes within the word, degree of morphological and morphophonological opacity manifested in such phenomena as fusion, cumulation, multiple exponence or suppletion, presence, productivity and types of incorporation, etc.

That said, it is remarkable that there are few comparative works based on representative samples of polysynthetic languages aiming at mapping their diversity and exploring the similarities and differences between them (e.g. Fortescue 1994, Drossard 1997). Indeed, most significant publications on polysynthesis deal with just one language or several selected languages. Perhaps the most famous exception to this trend is Baker (1996) approaching polysynthesis from a generative perspective. However, most typologists would disagree with Baker’s rather restrictive definition of polysynthesis based on obligatory expression of arguments within the verb by means of either bound pronominals or productive noun incorporation, as well as counter his claims about the correlations between these definitional features and other grammatical properties such as absence of grammatical case marking, reflexive pronouns or non-finite verbal forms (e.g. many contributions to Evans & Sasse 2002 and Mahieu & Tersis 2009).

A genuine typological approach to polysynthesis on a basis of a 75-language sample has been advanced in a series of articles by Johanna Mattissen (2004, 2006, 2017). Her typology is based on the following three major parameters:

- (i) Whether the verb stem can contain more than one lexical root: *compounding*² vs. *affixal polysynthesis*.
- (ii) Internal organisation of the polysynthetic morphology: *scope-ordered* vs. *templatic*.
- (iii) Number of arguments indexed in the verb by pronominal affixes: *polypersonal* (two or more), *monopersonal* (just one), *apersonal* (none).

The first parameter is probably the most important one and can be elaborated further. First, different types of compounding can be singled out: besides the best-known noun incorporation shown in example (2) above, Mattissen singles out adverb incorporation (5) and verb-root serialisation (6).

² Mattissen uses the term “compositional”, which creates unwarranted associations with semantic compositionality and should preferably be avoided.

- (5) Biniŋ Gun-wok (Gunwinyguan, Australia; Evans 2017b: 315)
a-ban-yawoyʔ-wargaʔ-marŋe-gaŋ-giŋe-ŋ
 1SG.SBJ-3PL.OBJ-again-wrongly-BEN-meat-cook-PST
 ‘I cooked the wrong meat for them again.’
- (6) Yimas (Lower Sepik-Ramu, New Guinea; Foley 1991: 331)
num-n na-way-mpi-waraca-mpi-ya-ntut
 village-OBL 3SG.S-turn-SEQ-return-SEQ-come-RPST
 ‘He came back to the village.’

Since various types of incorporation are not limited to polysynthetic languages (e.g. one finds noun incorporation in Frisian and Soninke, see e.g. Olthof 2020, Vinyar 2021, while productive verbal compounds are found e.g. in Japanese, Kageyama 2016), Mattissen (2004: 203; 2017: 94) suggests that criterial for polysynthesis are “non-root bound morphemes with rather concrete (“lexical”) meanings” (Mattissen 2004: 190), also known as “lexical affixes” (Mithun 1997) or “lexically heavy morphemes” (Fortescue 2017: 122). A detailed cross-linguistic analysis of such affixes is given in Mattissen (2006: 297–333), where the following ontological domains are singled out (the list below is non-exhaustive):

- (i) direction and position
- (ii) body parts
- (iii) classifiers
- (iv) artefacts and living creatures
- (v) motion and manner
- (vi) degree
- (vii) chronology
- (viii) phasal
- (ix) quantification and focus

Some of these as well as additional domains are illustrated below; thus (7) from Purépecha shows lexical affixes with locational meanings, (8) from Bella Coola shows body-part affixes expressing patients or instruments, while (9) from Central Alaskan Yupik shows affixes with verb-like meanings.

- (7) Purépecha (isolate, Mexico; Chamoreau 2017: 683)
- a. *waxa-nu-x-ti* ‘He sat *on the patio*.’
 - b. *waxa-ru-x-ti* ‘He sat *in the street*.’
 - c. *waxa-k’ara-x-ti* ‘He sat *inside the house*.’
- (8) Bella Coola (Salishan, Canada; Mithun 1997: 361)
- a. *cp-ak-m-c* ‘I am wiping my *hands*.’
 - b. *ip’-ak-m-tic* ‘I am grabbing them *with my hand*.’
 - c. *pusm-ak-c* ‘My *hand* is swelling.’
- (9) Central Alaskan Yupik (Eskimo-Aleut, USA; Woodbury 2017: 551)
- a. *qaya-ngqer-tua* ‘I have a *kayak*.’
 - b. *taryaqvag-tur-tua* ‘I’m *eating* king salmon.’
 - c. *citegta-ingu-unga* ‘I’m *tired* of tomcods.’

Of course, none of the languages in Mattissen’s sample has lexical affixes from all of the ontological domains, while at least some of these domains are attested by productive affixation in the languages traditionally not considered polysynthetic. For example, the so-called preverbs of Slavic, Baltic, Germanic, Hungarian or Kartvelian languages express locational as well as some more abstract meanings, e.g. degree or repetition (10), and Lithuanian has a prefix with the focus-related meaning ‘only’ (11) (Arkadiev 2010).

(10) Bulgarian (Indo-European; Istratkova 2004: 313)

iz-raz-pre-pro.da-m

CMPL-EXC-REP-sell-PRS.1SG

‘I completely sell (it) again in excess’

(11) Lithuanian (CCL)

Te-mat-au

j-os

tams-us

plauk-us.

RSTR-see-PRS.1SG

3-GEN.SG.F

dark-ACC.PL.M

hair-ACC.PL

‘I *only* see her dark hair.’

The main problem with “lexical affixes” is again that of delimitation of this class in a meaningful way (cf. Zúñiga 2017, Haspelmath 2018). While cases shown in (7)–(9) above appear uncontroversial, because those affixes indeed refer to concrete locations, objects or actions, the Bulgarian and Lithuanian affixes in (4) and (10)–(11) have quite abstract meanings, even if translatable by lexical words into English. Excluding such affixes from the domain of “lexically heavy” morphemes would substantially reduce the ontology given above. Mattissen’s (2017: 94) conjecture that the Slavic or German preverbs differ from non-root bound morphemes in polysynthetic languages such as Ket or Purepécha in that the former are “lexicalized on their roots” is factually incorrect as well as unhelpful, since, on the one hand, the Bulgarian and Lithuanian affixes shown in (10) and (11) are fully transparent and productive, while, on the other hand, root-affix combinations in polysynthetic languages are just as prone to lexicalisation (e.g. Dorais 2017, Mithun 1998). Thus “non-root bound morphemes with rather concrete meanings” is a fuzzy concept, possibly with a prototype-structure.

Turning to the second parameter of Mattissen’s typology, that of morphological organisation, we find the familiar distinction between “layered” (or scope-based) and “templatic” types of ordering (Stump 2006, Bickel & Nichols 2007: 214–220; Mithun 2016: 149–152). Clear manifestations of both types of morphological structure are not hard to find, e.g. successive attachment of suffixes in Central Alaskan Yupik in (12) vs. interlacing of derivational and inflectional elements in Ket (13), a hallmark of templatic morphology.

(12) Central Alaskan Yupik (Eskimoan, USA; Mithun 2016: 15151)

a. *quya-yuumi-it-u-a*

thankful-yearn-NEG-INTR.IND-1SG

‘I don’t want to be thankful.’

b. *quya-yuumi-ite-llru-u-nga*

thankful-yearn-NEG-PST-INTR.IND-1SG

‘I didn’t want to be thankful.’

c. *quya-yuumi-ite-llru-yugnarq-u-a*

thankful-yearn-NEG-PST-probably-INTR.IND-1SG

‘I guess I didn’t want to be thankful.’

- (13) Ket (Yeniseian, Siberia; Vaida 2017: 907)
da=in-ba-h-a-ted
 3F.SBJ=needle-1SG.OBJ-area-PRS-hit.endwise
 ‘She pokes me with a needle (once).’

However, both purely scope-ordered and purely templatic organisation are idealised types to which morphological structures of real languages adhere to different degrees, being shaped by diachronic processes of grammaticalisation, reanalysis and analogy (Mithun 2000, 2016), and Mattissen (2004: 208) acknowledges that “[t]wo types of mixed organization are observed”. In the first one, some affixes show scope-driven ordering in an otherwise rigid template as in Southern Sierra Miwok (14); in the second one, “different parts of the verb form, e.g. the pre-root and post-root part, may differ in their organizational principles” (Mattissen 2004: 208), as in Nivkh, where, according to Mattissen, suffixes follow a strict template while the order of pre-root elements is scope-driven.

- (14) Southern Sierra Miwok (Miwok-Costanoan, California; Broadbent 1964: 39–40)
- a. *ʔetal-nuk:u-lumhu-:*
 return-CAUS-ready-PRS.IPF.3SG
 ‘He is ready to make him go home.’
 - b. *ʔetla-lamhy-nuk:u-:*
 return-ready-CAUS-PRS.IPF.3SG
 ‘He is making him ready to go home.’

The last parameter of Mattissen’s typology pertains to head-marking. As said above, in contrast to other approaches, Mattissen considers polypersonalism to be a variable rather than a defining feature of polysynthesis. This is not unreasonable, since, on the one hand, polypersonalism is quite widespread (Siewierska 2005) and is in no way limited to polysynthesis, see Basque in (15), and, on the other, there exist languages exhibiting polysynthetic features while lacking argument indexing altogether, e.g. Awtuw (16).

- (15) Basque (isolate, Spain, France; Saltarelli 1988: 238)
Ni-k aita-ri diru-a eska-tu d-i-o-t.
 1SG-ERG father-DAT money-DEF[ABS] ask-PRF 3.ABS.PRS-DAT-3SG.IO-1SG.ERG
 ‘I have asked father for (some) money.’

- (16) Awtuw (Sepik, Papua–New Guinea; Feldman 1986: 52)
ka-d-ma-taw-owra-t-akla-kow-kay-e
 NEG-FACT-go-yet-again-DU-dig-BEN-PRF-PST
 ‘(two) hadn’t gone and dug again for (someone) yet’

The three parameters of Mattissen’s typology are largely independent of one another and cross-classify languages in a number of subtypes, as shown in Table 1 based on Mattissen (2017: 82)³.

³ It has to be observed that some of Mattissen’s classifications are inaccurate, e.g. the listing of Yimas among the languages with at most bipersonal indexing (cf. Foley 1991: 208–215) or of Tiwi as lacking incorporation (cf. Osborne 1974: 46–50).

Table 1: Classification of polysynthetic languages according to Mattissen

scope-ordered	templatic	noun incorporation	verb-root serialisation	no. of indexed participants	language
+	+	+	+	2	Lakhota
-	+	+	+	2	Wichita
+	-	+	+	2	Pano
-	+	+	-	2	Takelma
+	+	+	-	2	Blackfoot
-	+	-	+	2	Tonkawa
-	+	-	+	0	Awtuw
+	+	-	+	3	Yimas
+	+	-	+	0	Maidu
+	-	-	+	2	Capanawa
-	+	(+)	-	3	Creek
+	+	-	(+)	0	Klamath
+	+	(+)	-	2	Spokane
-	+	-	-	2	Navaho
+	+	-	-	1	Tariana
+	-	-	-	2	Greenlandic

The typological parameters discussed above (as well as other conceivable traits) are all synchronic in their nature. However, a highly important question concerns the diachronic origins of polysynthesis. Since most of the known polysynthetic languages either lack historical records altogether or have not changed much during their written history, only speculations based on internal and comparative reconstruction are possible. Thus Fortescue (2007) proposes to distinguish between “old” and “new” polysynthesis on the basis of synchronically observable traits summarised in Table 2.

Table 2: “Older” vs. “newer” polysynthesis according to Fortescue (2007: 21)

Older polysynthesis (e.g. Nuuchahnulth)	Newer polysynthesis (e.g. Chukchi)
(i) few if any lexical sources of derivational affixes to be found;	(i) lexical sources of derivational affixes transparent;
(ii) no independent stress on incorporated morphemes;	(ii) residual stress on incorporated stems;
(iii) entangled ordering of derivational and inflectional morphemes;	(iii) derivational morphemes closer to stem than inflectional morphemes;
(iv) evidence of successive layering of affixes, with fossilisation.	(iv) productivity of incorporation or verb-root serialisation.

In a more recent article, Fortescue (2016) postulates several diachronic pathways by which polysynthesis may arise, all presupposing as a prerequisite “embedding into a larger geographical region where head-marking is already dominant” (Fortescue 2016: 6). These idealised pathways differ in the major types of productive stem-derivational morphology and are therefore reminiscent of Mattissen’s “compositional” and “affixal” types (Fortescue 2016: 6):

- (i) productive verbalising affixes but little or no compounding → affixing type (e.g. Eskimoan, Wakashan);
- (ii) compounding of various kinds but no productive verbalising affixes → compounding type (e.g. Iroquoian, Chukotkan, Gunwinyguan);
- (iii) clause chaining or verb serialisation in fixed order → clause-combining type (e.g. Yimas, Athabaskan)

While the general consensus seems to be that the extreme morphological complexity of the polysynthetic kind takes extended periods of time to develop (Fortescue 2016, Dahl 2017) under such specific sociolinguistic conditions as isolation and no asymmetric language contact (Trudgill 2017), one may ask whether polysynthetic structures can also develop “abruptly” via univerbation of analytic constructions or clitic clusters. Indeed, it has been suggested that some of the analytically-looking European languages, when analyzed in their colloquial spoken form and without orthography-based preconceptions, may turn out to look rather polysynthetic, cf. Lambrecht (1981), Arkadiev (2005), Kibrik (2011: 253–259) on spoken French (17), Charitonidis (2008) on Modern Greek and Moreno Cabrera (2014) on spoken Spanish.

(17) Written vs. spoken French⁴

parce qu'il me les a toujours fait envoyer

paʁsk-i-mə-lez-a-tuʒuʁ-fɛ-ɑ̃vwajɛ

because-3SG.SBJ-1SG.OBJ-3PL.OBJ-PST.PFV-always-CAUS-send

‘because he has always had them sent to me’

Such examples, even if debatable, clearly bear on the issues of “wordhood” and of delimitation of polysynthesis.

4. Polysynthesis in the Northwest Caucasian languages

The Northwest Caucasian (or Abkhaz-Adyghean; further NWC) is one of the three indigenous language families of the Caucasus alongside the Northeast Caucasian (Nakh-Daghestanian) and the South Caucasian (Kartvelian). It comprises three branches: Circassian consisting of West Circassian (a.k.a. Adyghe) and Kabardian, Abkhaz-Abaza and the now extinct Ubykh. The languages are spoken by about 1 million people in the Russian republics of Adyghe, Karachay-Cherkessia and Kabardino-Balkaria and some districts of the Stavropol and Krasnodar regions, in the self-proclaimed Republic of Abkhazia (officially part of Georgia but *de facto* under Russian protectorate), as well as in the diaspora in Turkey and other countries of the Middle East. All NWC languages are head-marking and polysynthetic, and below I shall focus on the various manifestations of these traits trying to situate NWC in the typology of polysynthesis outlined above and to discuss the implications of their data for the understanding of the concept in general. Further information about these languages, including the sociolinguistic situation, state of research and most important traits of their phonology and morphosyntax, can be found in Hewitt (2005) and Arkadiev & Lander (2020). The foregoing discussion is based on the material collected by myself and my colleagues during fieldwork in Adyghe on various dialects of West Circassian and Kabardian (2004–2016) and in Karachay-Cherkessia on Abaza (2017–2021), as well as on published grammars, special studies and text collections. Examples marked as “texts” and “elicited” come respectively from the oral narratives and elicited examples collected during our fieldtrips.

All NWC languages possess highly complex verbal morphology with many suffixal and prefixal positions:

(18) Abaza (texts)

awəj árqan s-z-á-la-nəqʷa-wa-ʒə-j-š'a-t

DIST DEF+rope 1SG.ABS-POT-3SG.N.IO-LOC-walk-IPF-PVB-3SG.IO-seem-DCL

‘He thought that I would be able to walk on that rope.’

⁴ <https://www.babelio.com/livres/Dupuis-Apaches/1286313>, accessed October 10, 2022.

NWC nominals also show polysynthetic features, albeit of a different nature, i.e. productive and in many cases obligatory compounding of lexical stems corresponding to nouns, adjectives and numerals into so-called “nominal complexes” showing properties of coherent morphosyntactic words (Lander 2017):

(19) Standard West Circassian (published text)

jə-[ʒ'ene-šχ'ente-daxe]-re *jə-[c'weqe-λedeqe-λage]-re*
 POSS-dress-green-beautiful-COORD POSS-shoe-heel-high-COORD
 ‘her beautiful green dress and her shoes on high heels’

Verbal and nominal polysynthesis in NWC arguably instantiate different morphosyntactic mechanisms (Ershova 2020). Below I shall focus on verbal complexity, although the borderline between verbs and nominals in NWC is rather fluid (cf. Testelets & Lander 2017: 951–952).

The polysynthetic properties of NWC languages include the following:

- (i) exuberant polypersonalism coupled with limited (Circassian, Ubykh) or no (Abkhaz-Abaza) case marking of core grammatical relations;
- (ii) many productive affixes with different degrees of “lexicality”, most notably from the domain of spatial semantics, as well as vestiges of incorporation;
- (iii) an intricate mixture of templatic and layered organisation;
- (iv) a complex system of morphological expression of syntactic information.

4.1. Polypersonalism and “open head-marking”

Head-marking and polypersonalism can be illustrated by the following example showing as many as four person-number-gender prefixes indexing participants neither of which is expressed by an overt noun phrase:

(20) Abaza (texts)

š'ta j-šə-z-j-á-s-h^w-p'
 PTCL 3SG.N.ABS-2PL.IO-BEN-3SG.M.IO-DAT-1SG.ERG-say-NPST.DCL
 ‘OK, I’ll tell this to him (God) about you.’

Verbal forms indexing four participants like the one in (20) are infrequent but are attested in texts, being constructed by speakers when necessary. Tripersonal forms are fairly common, and in Circassian, verbal forms indexing five participants are reported in grammars (Kumakhov 2006: 200–202; Kumakhov & Vamling 2009: 38) and even attested in written texts (21).

(21) West Circassian (AdCorp)⁵

t-jə-wəram asfal't
 1PL.PR-POSS-street asphalt
Ø-qə-t-fə-Ø-tər-a-r-jə-ke-λha-κ
 3.ABS-CSL-1PL.IO-BEN-3.SG.IO-LOC:on-3PL.IO-DAT-3SG.ERG-CAUS-put-PST
 ‘He made them put asphalt on our street for us.’

⁵ The 3rd person absolutive and 3rd person singular indirect object prefixes in Circassian are zero. They won’t be marked in the examples unless necessary.

Polypersonalism in NWC is facilitated by a rich system of valency-increasing derivations, in particular, by the numerous semantically specialized applicatives that introduce indirect objects expressing various peripheral participants (see Arkadiev et al. forthcoming and references therein), e.g. the benefactive *z-* in (20) and the benefactive *fə-* and the locative *tər-* in (21). This possibility to freely add peripheral participants to the verbal core by means of productive morphology was called “open head-marking” by Nichols (2017), and can be considered one of the hallmarks of “true” polysynthesis (Zúñiga 2019: 12). Applicatives in NWC are very numerous (from about twenty in Circassian to several dozens in Abaza and Abkhaz) and range semantically from underspecified, as e.g. the “dative” in (20) and (21), to highly specialised, as e.g. the comitative (22a) or the numerous locative applicatives (22b).

(22) Besleney Kabardian (texts)

a. *s-a-də-če-č'-a*

1SG.ABS-3PL.IO-COM-LOC:under-go_out-PST

‘I went away with them.’

b. *š'abe qə-s-č'-jə-ʒ-t-jə*

soft CSL-1SG.IO-LOC:under-3SG.ERG-throw-IPF-ADD

‘She would put something soft under me.’

NWC applicatives combine with intransitive and transitive verbs alike and introduce indirect objects. These are expressed by a special series of pronominal prefixes in dedicated slots in the prefixal chain and their presence normally does not affect the otherwise ergative expression of agents and patients, consider examples in (23).

(23) West Circassian (Letuchiy 2009: 331)

a. *txəl-xe-r s-š'efə-ke-x*

book-PL-ABS 1SG.ERG-buy-PST-PL

‘I bought books.’

b. *č'ale-xe-m txəl-xe-r a-fe-s-š'efə-ke-x*

boy-PL-OBL book-PL-ABS 3PL.IO-BEN-1SG.ERG-buy-PST-PL

‘I bought books for the boys.’

Applicatives allow stacking, see (21) above and even limited recursion (24), testifying to a high degree of productivity and semantic transparency.

(24) West Circassian (Lander & Letuchiy 2010: 269)

s-a-fə-Ø-f-e-txe

1SG.ABS-3PL.IO-BEN-3SG.IO-BEN-DYN-write

‘I write to him for their benefit / to them for his benefit.’

On the other hand, many combinations of applicatives with verbal roots are lexicalised, as e.g. the comitative in (25), and some applicatives, particularly the “dative” one, express arguments required by the lexical root, as e.g. the addressee of ‘say’ in (20).

(25) Abaza (texts)

šə-r-c-graʕa-ra

a-taqə-p

2PL.ABS-3PL.IO-COM-help-MSD 3SG.N.IO-need-NPST.DCL

‘We have to help them.’

The extensive polypersonalism of NWC and the valency-increasing mechanisms of “open head-marking” behind it raise important questions about the cross-linguistic variation in argument structure and the argument-adjunct distinction.

4.2. Lexical affixes and traces of incorporation

NWC languages possess lexical affixes from most of the ontological domains described by Mattissen (2006), e.g. degree (26), situational modality (27), phasal (28), repetitive (29) and focus (30), as well as from domains not listed by Mattissen, e.g. subjective evaluation (31) and similitude (43) in section 4.3.

Abaza (texts)

- (26) *j-s-g^wapχa-ʒa-wá-ta*
3SG.N.ABS-1SG.ERG-love-INT-IPF-ADV
‘I liked it *very much*, and...’
- (27) *j-g'-zǎ-na-m-ʒa-t*
3PL.ABS-NEG.EMP-POT-TRL-NEG-reach-DCL
‘They *could* not reach it.’
- (28) *j-gála-rk^wa-ztən-g'əj ...* *h-tʒə*
3SG.N.ABS-stand-CNT-COND-ADD 1PL.PR-house
‘If our house *still* exists...’
- (29) *awa-ʔa h-ata-də-r-ca-χ-wa-n*
DIST-LOC 1PL.ABS-REP-3PL.ERG-CAUS-go-RE-IPF-PST
‘They used to make us go there *again*.’
- (30) *awasa j-h^wa-ʕ^waca* *aža-zažəḵ*
but 3SG.N.ABS-say-RSTR word-one
‘But say *just* one word.’ (published texts, Luke 7:7)
- (31) West Circassian (Rogava & Kerasheva 1966: 306)
k^wa-ʁe-g^wəʒ^ʔ-ep
go-PST-DPR-NEG
‘*Unfortunately*, he didn’t go.’

In principle, chains of affixes are possible, as in (32), however, forms with more than one affix are only rarely attested in texts.

- (32) Ubykh (Fenwick 2011: 127, transcription and glosses adapted)
jǎ-Ø-s-t^w-aj-le-f-ew-mə-t
3SG.ABS-3SG.IO-1SG.ERG-give-RE-CMPL-POT-FUT-NEG-FUT
‘I won’t *be able* to give it *back* to him *completely*.’

The affixes just shown do not fit well into the traditional classification of morphological phenomena into derivation vs. inflection (Spencer 2013, Haspelmath to appear). On the one hand, they are derivation-like in being optional and expressing rather concrete semantic content; on the other, they resemble inflection by being (at least in principle) highly

productive, compositional and able to derive clearly ad-hoc forms. Their properties are thus close to De Reuse's (2009) PNC, even if their combinatorics is not as impressive as that found in Yupik.

The largest set of "lexical" affixes in NWC express spatial meanings. These include simplex and complex locative prefixes ("preverbs"), ranging in number from ca. 30 in Circassian to ca. 150 in Abkhaz and Abaza, as well as much less numerous directional suffixes. Importantly, all locative preverbs in Circassian and many in Ubykh, Abaza and Abkhaz are applicatives introducing the landmark as indirect object, as in (22b) in section 4.1 above; in many cases this is the only way to express the landmark. Locative preverbs are obligatory with some verbal roots denoting position and directed motion, the choice of the preverb depending on the spatial configuration and the type of the landmark (Paris 1995), cf. (33).

(33) Standard Kabardian (Kumakhov 1964: 165)

- a. *tjepšeč'ə-m* *jə-λə-n*
 plate-OBL LOC:container-lie-MSD
 'to be on a plate'
- b. *škampə-m* *de-λə-n*
 cupboard-OBL LOC:enclosure-lie-MSD
 'to be in a cupboard'
- c. *wəne-m* *še-λə-n*
 room-OBL LOC:under-lie-MSD
 'to be in a room'

Directional suffixes express such meanings as 'inside', 'outside', 'around', 'up' and 'down'. They normally combine with locative preverbs introducing the landmark (34a); in some cases such combinations are fixed, the landmark being implicit (34b).

(34) Besleney Kabardian (texts)

- a. *šə-m* *tje-d-be-təs-ha-ne*
 horse-OBL LOC:on-1PL.ERG-CAUS-sit-LAT-FUT
 'We shall make him sit on a horse.'
- b. *pšeχ^wə-r* *d-a-hə-je-ž'-a*
 chain-ABS LOC:enclosure-3PL.ERG-carry-VERT-RE-PST
 'They carried the chain up.'

Some directional suffixes can occur as independent roots (taking the appropriate locative preverbs) with the same meaning, hence, their combinations with verbal roots can be considered verb-root serialisation (cf. the discussion of "incorporation" in Circassian by Kumakhov 1964: 139–146), see (35).

(35) Abaza

- a. *š-sə-d-thawsəχ'a-l-χ-əj-t*
 2PL.ABS-1SG.IO-LOC:close-complain-go_in -RE-PRS-DCL
 'You come to me with complaints.' (Klychev 1972: 96)
- b. *d-a-k^wə-z-ga-ša-t*

3SG.H.ABS-3SG.N.IO-LOC-1SG.ERG-carry-go_around-DCL

‘I carried it around.’ (Klychev 1995: 138)

Diachronically, most locative preverbs in NWC stem from incorporated nouns, mainly denoting parts of the body or of other objects (Kumakhov 1964: 139–146, 164–182; Kumakhov 1989: 200–228; Klychev 1994; Avidzba 2017; Arkadiev & Maisak 2018: 121–127). In those cases when the preverb synchronically coexists with the noun, the former still shows some semantic link with the latter, even if the meaning of the preverb is more abstract (36).

(36) West Circassian (Kumakhov 1964: 177–179)

a. *k^weçə* ‘intestines’: *k^weçə-λhe-n* ‘put *inside* sth’
k^weçə-rə-čə-n ‘run *through* sth’

b. *ʔ^wə* ‘mouth’: *ʔ^wə-c^we-n* ‘stand *near* sth’
ʔ^wə-š’ə-n ‘lead *away* from sth’

Some preverbs can even be viewed synchronically as incorporated nouns, given that they retain their lexical meaning and can host pronominal prefixes referring to the possessor. Such cases are marginal in Circassian, only involving a few body-part nouns (37), but are more widespread in Abaza and Abkhaz, where both body-part (38) and non-relational (39) nouns can incorporate. However, even in the latter languages this type of compounding is not productive, with only a limited number of nouns co-occurring with a limited number of verbs.

(37) Standard Kabardian (Kumakhov 1964: 181–182)

a. *še* ‘head’: *še-rə-xə-n* ‘take from one’s *head*’
b. *ʔe* ‘hand’: *ʔe-še-xə-n* ‘take from one’s *hands*’

(38) Abaza (Klychev 1995: 154)

a-saba *ʃa-rə-lakta-pl-əw-n*
DEF-dust CSL-3PL.IO-face-pour.powder-IPF-PST
‘Dust was pouring onto *their faces*.’

(39) Abkhaz (Avidzba 2017: 99)

a-raχ^w *ʋə-çə-h-ga-ra.wə-p̄*
DEF-cattle winter-LOC:under-1PL.ERG-carry-DEB-NPST.DCL
‘We have to keep the cattle during the winter.’

Thus, in terms of Mattissen’s typology, NWC polysynthesis is transitional between “compounding” and “affixal”, with both noun incorporation and verb-root serialisation being attested, but applying to closed classes of roots and tending to yield grammaticalised elements patterning with affixes. While many of the NWC “lexical” affixes have cognate roots and some of them even retain their original semantics, the majority of them are highly grammaticalised and fully integrated into the morphological system, pointing towards “older” polysynthesis in Fortescue’s terms.

Simultaneously, NWC morphology shows clear signs of numerous layers of expansion and renewal of polysynthetic structures, including some clearly recent formations. The latter comprise, for instance, a number of TAM suffixes derived from former auxiliaries (see Arkadiev & Maisak 2018: 127–132 on Circassian), or a remarkable case of “dependent-head synthesis” (Mattissen 2003) in Abaza exemplified above in (18). These forms involve the verb *ʒə-š’a* ‘seem, think’ compounded with the head of its sentential complement, the two

verbal stems being furnished each with their own indexing prefixes and temporal/aspectual suffixes and admitting separate modification by temporal adverbials (40).

- (40) Abaza (Panova 2020a: 98)
sara jacə [wara *wax'ça* *χabez*
 1SG yesterday 2SG.M today Khabez
wə-c-əw-š']-zə-s-š'-əw-n
 2SG.M.ABS-go-IPF-FUT-LOC-1SG.IO-seem-IPF-PST
 'Yesterday I thought you would go to Khabez today.'

At the same time, as shown by Panova (2020a,b), such forms behave as coherent words, as evidenced by their inseparability and unpermutability as well as by the bipartite negation marker, whose prefixal part occurs to the left of the dependent verbal stem even when the main verb is negated (41).

- (41) Abaza (Panova 2020b: 291)
d-g'-ʒa-j-zə-s-š'-əw-m
 3SG.H.ABS-NEG.EMP-CSL-go-PVB-1SG.IO-seem-IPF-NEG
 'I don't think he came.'

These patterns of “morphologically bound complementation” (Maisak 2016: 837) or “multiclausal polysynthesis” (Zúñiga 2019: 5–6) clearly go against Fortescue’s (2017: 119–121) conjecture that the complexity of a polysynthetic predicate is always limited by the so-called “macro-event property” (Bohnemeyer et al. 2011). However, the formal and semantic transparency of these constructions as well as their absence in Abkhaz suggest their fairly recent origin; one might hypothesise that diachronic development of such verbal compounds should lead to tighter semantic integration.

4.3. Morphological organisation

NWC languages present an intricate and complex interplay of templatic and scope-ordered principles in their morphological make-up (Korotkova & Lander 2010, Arkadiev & Letuchiy 2011, Lander 2016 on West Circassian, Panova 2018 on Abaza). Table 3 schematically presents the general verbal structure distilled from much more detailed and expanded templates of each particular language.

Table 3: The general structure of the NWC verb (Arkadiev & Lander 2020: 404)

prefixes				root		suffixes				
argument structure zone			pre-stem elements	stem			endings			
absolutive	subordinators	applicatives and indirect objects	ergative	preradical negation	causative	root	aspectual, modal and evaluative operators	temporal operators	suffixal negation	illocutionary operators or subordinators
1	1	>1	1	1	1 or 2	may be complex	>1	>1	1	>1

The verbal complex is divided into several zones each of which includes a number of slots and follows its own organising principles. The distinction between the zones of “stem” and “endings” is most robust in Circassian, where it manifests itself in stress assignment and

application of certain morphophonological processes. The ordering of suffixes is largely scope-driven, so that actional and evaluative modifiers normally precede markers of TAM, which are followed by clause-typing markers (42). Reordering of suffixes is also attested, although it is fairly limited (43).

- (42) West Circassian (Korotkova & Lander 2010: 307)
w-je-bewə-ž'ə-š'ə-š't-a?
 2SG.ABS-DAT-kiss-RE-POT-FUT-Q
 ‘Will you be able to kiss her again?’ (question > future > possibility > again)
- (43) West Circassian (Lander 2016: 3523)
 a. *g^wəš^we-š^we-ž'ə-β*
 be.happy-SML-RE-PST
 ‘s/he pretended again that s/he was happy’ (again > pretend)
- b. *g^wəš^we-ž'ə-š^wa-β*
 be.happy-RE-SML-PST
 ‘s/he pretended that s/he was happy again’ (pretend > again)

The ordering of prefixes is more intricate. On the one hand, some slots, e.g. those where the absolutive and ergative arguments are indexed, as well as the causative and the preradical negation, are fixed. On the other hand, the “intermediate” prefixes are at least partly scope-ordered (subordinators > prefixal potential > compositional applicatives > lexicalised applicatives) (44). Semantically-driven reordering of prefixes, however, is very rare.

- (44) Abaza (elicited)
d-šə-z-wə-c-nə-m-χ-əw-š *l-h^wa-t*
 3SG.H.ABS-SBD-POT-2SG.M.IO-COM-PVB-NEG-work-IPF-FUT 3SG.F.ERG-say-DCL
 ‘She said that she won’t be able to work with you.’

At the same time, there are clear cases of counter-scopal ordering of prefixes, so that e.g. the preradical negation and the causative, which occur closest to the root, often take scope over the prefixes located farther from it (45); the Circassian cislocative prefix, despite being often lexicalised, always occurs in the position to the left of subordinators scoping above it (46).

- (45) Abaza (texts)
a-wandər h-a-k^w-də-r-çə-t
 DEF-cart 1PL.ABS-3SG.N.IO-LOC:top-3PL.ERG-CAUS-sit-DCL
 ‘They put us in the cart (lit. made us sit on it).’ (causative > locative)
- (46) Besleney Kabardian (texts)
qə-š'ə-k^we-m
 CSL-TEMP-go-OBL
 ‘when he came to her’ (when > cislocative)

Dependencies between non-adjacent slots, characteristic of templatic morphology (Stump 2006), are also attested, see e.g. Arkadiev & Letuchiy (2011) on prefix-suffix interaction in West Circassian. Thus, in Abkhaz and Abaza the choice of the past tense suffix on the right edge of the word is sensitive to the choice of the personal vs. relative absolutive prefix in the leftmost position (47), while in Ubykh the number of the absolutive argument influences the

shape of affixes in several non-adjacent positions and governs suppletion of certain roots (48) (a case of multiple exponence), see Smeets (1997).

(47) Abkhaz (Chirikba 2003: 44)

a. *də-r-ga-wá-n*
3SG.H.ABS-3PL.ERG-carry-IPF-PST.DCL
'they were taking him/her'

b. *jə́-r-ga-wa-z*
REL.ABS-3PL.ERG-carry-IPF-PST.NFIN
'whom they were taking'

(48) Ubykh (Dumézil & Esenc 1975: 173, transcription adapted, glosses added)

a. *sə-w-də-q^we.t^w-qe*
1SG.ABS-2SG.ERG-CAUS.SG-stop.SG-PST
'You (sg) made me stop.'

b. *š'ə-w-ve-q^we.χe-qe-n*
1PL.ABS-2SG.ERG-CAUS.PL-stop.PL-PST-PL
'You (sg) made us stop.'

Finally, the position of some affixes is simply variable without any discernible difference in meaning, cf. the "floating" 3PL indirect object prefix in Circassian (49).

(49) Besleney Kabardian (elicited)

a. *sə-q-a-de-k^w-a*
1SG.ABS-CSL-3PL.IO-COM-go-PST

b. *s-a-qə-de-k^w-a*
1SG.ABS-3PL.IO-CSL-COM-go-PST
a=b 'I came with them.'

Thus, the morphological organisation of NWC verbs is not uniform and defies any straightforward analysis aiming to reduce affix combinatorics and ordering either to semantic scope or to a rigid template. This complexity and heterogeneity obviously reflects a long and non-trivial historical development.

4.4. Morphology-syntax interface

One of the most remarkable aspects of the NWC polysynthetic morphology is the fact that a large part of it serves the purposes of syntax. Head-marking, applicatives and other valency-changing mechanisms are deployed for the expression of core and peripheral participants of the clause; the nominal complex briefly mentioned above is the main means of encoding head-modifier relations in the nominal domain. There are numerous other constructions where morphology plays a crucial role, e.g. reflexives and reciprocals. Not only are binding relations between co-arguments normally expressed within the verb in NWC, as in many languages of the world, but the way they are encoded is significant, see (50)–(51).

(50) West Circassian (Letuchiy 2012: 342)

a. *wə-sə-wəpsə-β*
2SG.ABS-1SG.ERG-shave-PST
'I shaved you.'

- b. *zə-sə-wəpsə-ɸ*
RFL.ABS-1SG.ERG-shave-PST
'I shaved (myself).'

(51) West Circassian

- a. *sə-š^wə-š'-e-g^wəɸə*
1SG.ABS-2PL.IO-LOC-DYN-rely
'I rely on you (pl).' (adapted from AdCorp)
- b. *tə-ze-š'-e-g^wəɸə-ž'ə-x*
1PL.ABS-REC.IO-LOC-DYN-rely-RE-PL
'We rely on each other.' (Letuchiy 2007: 788)

As is clear from these examples, in Circassian the reflexive and reciprocal markers occupy the same slots as the corresponding personal prefixes (Letuchiy 2007) and do not reduce the verbs' valency (Lander & Letuchiy 2017). They can be considered morphologically bound reflexive and reciprocal pronouns subject to syntactic binding (Ershova 2019), just like personal prefixes are morphologically bound referential pronominals (Kibrik 2011: 92–97).

Another domain where morphology is employed for the purposes of syntax in NWC is clause combining. These languages possess elaborated systems of morphological encoding of inter-clausal relations including nominalisations, converbs, and relativisation. The latter is the least trivial and has received considerable attention in the literature, see e.g. Hewitt (1979a, 1979b), O'Herin (2002), Caponigro & Polinsky (2011), Lander (2012), Ershova (2021). The morphological expression of relativisation is illustrated in (52).

(52) Abaza (elicited)

- a. *a-ph^wəspa* *çə* *lə-s-t-t*
DEF-girl apple [3SG.N.ABS]3SG.F.IO-1SG.ERG-give-DCL
'I gave an apple to the girl.'
- b. [*a-ph^wəspa* *j-lə-s-tə-z*] *a-çə*
DEF-girl REL.ABS-3SG.F.IO-1SG.ERG-give-PST.NFIN DEF-apple
'the apple that I gave to the girl'
- c. [*a-çə* *zə-s-tə-z*] *a-ph^wəspa*
DEF-apple REL.IO-1SG.ERG-give-PST.NFIN DEF-girl
'the girl whom I gave the apple'
- d. [*a-ph^wəspa* *çə* *lə-z-tə-z*] *ə-ç'k^wən*
DEF-girl apple 3SG.F.IO-REL.ERG-give-PST.NFIN DEF-boy
'the boy who gave an apple to the girl'

Again, like reflexivity and reciprocity, relativisation is expressed by a special series of prefixes occupying the same positions as the corresponding personal markers. This suggests that NWC relative verbal forms are not "participles", as traditional grammars sometimes dub them (cf. Shagal 2019: 28), but rather involve morphologically bound resumptive pronouns (Lander & Daniel 2019).

Relativisation in NWC is employed not only for encoding clausal modifiers of nouns, but for other functions as well, see Caponigro & Polinsky (2011), Lander (2012). Headless relative clauses can be used for reference (53), in pseudo-clefts marking focus (Sumbatova 2009) (54), for adverbial subordination (55) and sentential complementation (56); in the latter two functions special relativisation markers are employed.

(53) West Circassian (published texts)

zə-z-ʁe-bəλə-ž'ə-ʁe-m *λəχ^wə-ʁ-ep*
 RFL.ABS-REL.ERG-CAUS-hide-RE-PST-OBL search-PST-NEG
 'He did not look for the one who hid.'

(54) Besleney Kabardian (texts)

[*jə-pe* *nahə-b-əw* *wə-z-ʁe-g^wəmeč'ə-r*] *bze-ra*
 POSS-before more-much-ADV 2SG.ABS-REL.ERG-CAUS-worry-ABS language-PRED
 'What worries you most of all is the language.'

(55) Abaza (texts)

[*ápχ'arta* *s-an-šá-lga*] *a-institút* *s-cá-ʧ*
 DEF+school 1SG.ABS-REL.TEMP-CSL-finish DEF-college 1SG.ABS-go-DCL
 'When I finished school I went to college.'

(56) Besleney Kabardian (texts)

[*mew-bə-m* *λəʁe* *zerə-xe-mə-λ-r-jə*] *qə-g^wəraʧ^w-a*
 DIST-OBL-OBL courage REL.FACT-LOC.mass-NEG-lie-ABS-ADD CSL-understand-PST
 'She realised that he didn't have courage.'

The most peculiar development of relative verbal forms is attested in Abaza and Abkhaz, where they serve as bases for finite forms expressing matrix content questions (Arkadiev 2020, Arkadiev & Caponigro 2021). In these forms, the relative prefix indicates the role of the question variable while dedicated interrogative markers encode its ontological class (human vs. non-human argument vs. adjunct), see (57).

(57) Abaza

a. *j-wá-c-k^wa-z-da* *l-h^wa-n*
 REL.ABS-2SG.M.IO-be.with-PL-PST.NFIN-Q.H 3SG.F.ERG-say-PST.DCL
 'Who were with you? she asked.' (texts)

b. *wə-z-pš-wá-ja?*
 2SG.M.ABS-REL.IO-look-IPF-Q.N
 'What are you looking at?' (elicited)

c. *aráj* *áč'á* *n-bá-ša-z-ʁəč'*
 PROX DEF+money REL.TEMP-Q.ADV-CSL-1SG.ERG-steal
 'When did I steal this money?' (texts)

Such forms, probably going back to univerbation of pseudocleft constructions (Arkadiev 2020: 245–247), testify to the intricate ways syntactic patterns can feed morphological structures in polysynthesis, leading to morphology “usurping” the functions most languages express syntactically.

5. Conclusions and prospects

As the exposition above has shown, the Northwest Caucasian languages adhere to the cross-linguistic “prototype” of polysynthesis, but show a number of specific features. These include extraordinary polypersonalism and “open head-marking” enhanced by a rich system of applicatives and highly developed syntactic functions of morphology, including a cross-linguistically rare pattern of relativisation forming one of the core mechanisms of NWC grammar. In terms of Mattissen’s typology (and contrary to her own characterisation of Abkhaz, e.g. Mattissen 2004: 206), NWC languages belong to the “transitional” type showing vestiges of different types and diachronic layers of incorporation feeding the system of lexical affixes, as well as a complex mixture of scope-ordered and templatic organisation.

An important implication from the Northwest Caucasian material is that polysynthesis represents a specific way of drawing boundaries between morphology and syntax. As we have seen, in NWC verbal (as well as nominal) morphology is employed to express and manage syntactic relations both intra- and inter-clausally to a remarkable extent, with processes of affixation and compounding fulfilling such functions as adnominal modification, encoding of core and peripheral participants, coreference relations, relativisation, and, through the latter, marking of focus and questions. In turn, the productive subsystems of polysynthetic morphology can display syntax-like properties like additivity, compositionality and recursion, rendering the boundary between “inflection” and “derivation” blurred or even altogether irrelevant.

To conclude, polysynthesis is a composite notion not reducible to a single feature such as high syntagmatic complexity of morphology, head-marking or incorporation, and languages traditionally described as polysynthetic considerably vary on such parameters as availability and types of compounding, semantic types of affixation, morphological makeup and many others. Moreover, some polysynthetic traits are found in non-polysynthetic languages as well, suggesting that the boundaries of polysynthesis are fuzzy rather than sharp. Neither internal diversity nor permeability of the class of polysynthetic languages should, however, be considered as an embarrassment, rather, they are a logical necessity, since polysynthetic properties emerge from various sources and gradually accumulate over time. Whether the notion “polysynthesis” itself is useful for morphological typology remains a matter of perspective. It has certainly been helpful in allowing linguists to map the blank spots of linguistic diversity, discover typologically rare and exceptional structures and free themselves of the Eurocentric preconceptions about the divisions between lexicon and grammar and between morphology and syntax. However, to facilitate further progress, the notion surely has to be made more precise by carefully disentangling its various components and exploring their distribution and mutual correlations, also in languages that we are not used to treat as “polysynthetic”.

Acknowledgments

I thank all the native speakers of West Circassian, Kabardian and Abaza with whom I have worked, as well as the participants of our collective fieldtrips. This paper is based on the invited talks given at the 18th International Conference on Typology and Grammar for Young Researchers (Saint-Petersburg, November 2021) and the 13th Mediterranean Morphology Meeting (Rhodes, May 2022). I thank the organisers of MMM13 for inviting me and to the audience of the conference for their questions. Besides that, I am grateful to Jenny Audring, Rainer Feer, Arnaud Fournet, Rupert Hartung, Dmitry Zelensky and Fernando Zúñiga for their feedback on the draft version of this article. All faults and shortcomings remain mine.

References

- AdCorp = Adyghe Corpus, http://adyghe.web-corpora.net/index_en.html
- Arkadiev, Peter. 2005. Grammaticalization of polysynthesis (with special reference to Spoken French). Talk at the 4th Typological School. Tsakhkadzor, Armenia, September 2005.
- Arkadiev, Peter. 2010. Notes on the Lithuanian restrictive. *Baltic Linguistics* 1. 9–49.
- Arkadiev, Peter. 2020. Syntax in morphological guise: Interrogative verbal morphology in Abaza. *Linguistic Typology* 24(2). 211–251.
- Arkadiev, Peter. 2021. Is Lithuanian a polysynthetic language? In Peter Arkadiev, Jurgis Pakerys, Inesa Šeškauskienė & Vaiva Žeimantienė (eds.), *Studies in Baltic and other Languages. A Festschrift for Axel Holvoet on the occasion of his 65th birthday*, 28–44. Vilnius: Vilnius University Press. Available at: <https://www.zurnalai.vu.lt/open-series/article/view/24481>
- Arkadiev, Peter & Ivano Caponigro. 2021. Conveying content questions without *wh*-words: evidence from Abaza. In Patrick Grosz, Luisa Martí, Hazel Pearson, Yasutada Sudo & Sarah Zobel (eds.), *Proceedings of Sinn und Bedeutung 25*, 73–94. Available at: <https://ojs.ub.uni-konstanz.de/sub/index.php/sub/issue/view/29>
- Arkadiev, Peter & Yury Lander. 2020. The Northwest Caucasian languages. In Maria Polinsky (ed.), *The Oxford Handbook of the Languages of the Caucasus*, 369–446. Oxford: Oxford University Press.
- Arkadiev, Peter, Yury Lander & Irina Bagirokova. Applicative constructions in the Northwest Caucasian languages. Submitted to Fernando Zúñiga & Denis Creissels (eds.), *Applicative Constructions. A Comparative Handbook*.
- Arkadiev, Peter & Alexander Letuchiy. 2011. Prefixes and suffixes in the Adyghe polysynthetic wordform: Types of interaction. In Vittorio Springfield Tomelleri, Manana Topadze & Anna Lukianowicz (eds.), *Languages and Cultures in the Caucasus*. 495–514. München, Berlin: Otto Sagner.
- Arkadiev, Peter & Timur Maisak. 2018. Grammaticalization in the North Caucasian languages. In Heiko Narrog & Bernd Heine (eds.), *Grammaticalization from a Typological Perspective*, 116–145. Oxford: Oxford University Press.
- Avidzba, Asmat V. 2017. *Lokal'nye preverby v abxazskom i abazinskom jazykax*. [Local preverbs in Abkhaz and Abaza]. PhD dissertation, Abkhaz Institute for the Humanities, Sukhum.
- Baker, Mark C. 1996. *The Polysynthesis Parameter*. Oxford, New York: Oxford University Press.
- Bickel, Balthasar. 2011. Grammatical relations typology. In Jae-Jung Song (ed.), *The Oxford Handbook of Typology*, 399–444. Oxford: Oxford University Press.
- Bickel, Balthasar & Johanna Nichols. 2007. Inflectional morphology. In Timothy Shopen (ed.), *Language Typology and Syntactic Description*. Vol. 3. *Grammatical Categories and the Lexicon*, 169–240. 2nd ed. Cambridge: Cambridge University Press.
- Bickel, Balthasar & Fernando Zúñiga. 2017. The ‘word’ in polysynthetic languages: Phonological and syntactic challenges. In: Fortescue et al. (eds.), 158–185.
- Bohnemeyer, Jürgen, Nicholas J. Enfield, James Essegbey & Sotaro Kita. 2011. The macro-event property. The segmentation of causal chains. In Jürgen Bohnemeyer & Eric Pederson (eds.), *Event Representation in Language and Cognition*, 43–67. Cambridge: Cambridge University Press.
- Broadbent, Sylvia M. 1964. *The Southern Sierra Miwok Language*. Berkeley, Los Angeles: University of California Press.
- Caponigro, Ivano & Maria Polinsky. 2011. Relative embeddings: A Circassian puzzle for the syntax/semantics interface. *Natural Language and Linguistic Theory* 29(1). 71–122.
- Chamoreau, Claudine. 2017. Purépecha, a polysynthetic but predominantly dependent-marking language. In Fortescue et al. (eds.), 667–695.
- Charitonidis, Chariton. 2008. Polysynthetic tendencies in Modern Greek. *Linguistik Online* 33(2). <https://doaj.org/article/6c5927ad2a1042f384cf9ae9ff136f55>
- Chirikba, Vyacheslav. 2003. *Abkhaz*. München, Newcastle: LINCOM Europa.
- CCL — Corpus of Contemporary Lithuanian, <http://tekstynas.vdu.lt/tekstynas/>
- Dahl, Östen. 2017. Polysynthesis and complexity. In Fortescue et al. (eds.), 19–29.

- de Reuse, Willem. 2009. Polysynthesis as a typological feature. An attempt at a characterization from Eskimo and Athabaskan perspectives. In Marc-Antoine Mahieu & Nicole Tersis (eds.), *Variations on Polysynthesis: The Eskaleut languages*, 19–34. Amsterdam, Philadelphia: John Benjamins.
- Dorais, Louis-Jacques. 2017. The lexicon in polysynthetic languages. In Fortescue et al. (eds.), 135–157.
- Drossard, Werner. 1997. Polysynthesis and polysynthetic languages in comparative perspective. In: Bohumil Palek (ed.), *Typology: Prototypes, Item Orderings and Universals. Proceedings of the Conference held in Prague August 20–22 1996*, 251–264. Prague: Charles University Press.
- Dumézil, Georges & Tefvik Esenç. 1975. *Le verbe oubykh. Études descriptives et comparatives*. Paris: Klincksieck.
- Duponceau, Peter S. 1819. Report of the corresponding secretary to the committee, of his progress in the investigation committed to him of the general character and forms of the languages of the American Indians. *Transactions of the Historical & Literary Committee of the American Philosophical Society* 1. xvii–xivii.
- Ershova, Ksenia. 2019. *Syntactic Ergativity in West Circassian*. PhD Dissertation, University of Chicago. Available at: <https://knowledge.uchicago.edu/record/1848?ln=en>
- Ershova, Ksenia. 2020. Two paths to polysynthesis. Evidence from West Circassian nominalizations. *Natural Language and Linguistic Theory* 38(2). 425–475.
- Ershova, Ksenia. 2021. Diagnosing clause structure in a polysynthetic language: Wh-agreement and parasitic gaps in West Circassian. *Linguistic Inquiry* 52(1). 1–38.
- Evans, Nicholas. 2017a. Polysynthesis in Dalabon. In Fortescue et al. (eds.), 759–781.
- Evans, Nicholas. 2017b. Polysynthesis in Northern Australia. In Fortescue et al. (eds.), 313–335.
- Evans, Nicholas & Hans-Jürgen Sasse. 2002. Introduction: problems of polysynthesis. In Nicholas Evans & Hans-Jürgen Sasse (eds.), *Problems of Polysynthesis*, 1–13. Berlin: Akademie Verlag.
- Evans, Nicholas & Hans-Jürgen Sasse (eds.). 2002. *Problems of Polysynthesis*. Berlin: Akademie Verlag.
- Feldman, Harry. 1986. *A Grammar of Awtuw*. Canberra: Australian National University.
- Fenwick, Rohan S. H. 2011. *A Grammar of Ubykh*. München: LINCOM Europa.
- Foley, William A. 1991. *The Yimas Language of New Guinea*. Stanford: Stanford University Press.
- Fortescue, Michael. 1994. Polysynthetic morphology. In R. A. Asher (ed.), *Encyclopaedia of Language and Linguistics*, Vol. 5, 2600–2602. Oxford: Pergamon Press.
- Fortescue, Michael. 2007. The typological position and theoretical status of polysynthesis. *Tidsskrift for Sprogforskning* 5(2). <https://tidsskrift.dk/tfs/article/view/530>
- Fortescue, Michael. 2016. Polysynthesis: A diachronic and typological perspective. In *Oxford Research Encyclopedia of Linguistics*.
- Fortescue, Michael. 2017. What are the limits of polysynthesis? In Fortescue et al. (eds.), 115–134.
- Fortescue, Michael, Marianne Mithun & Nicholas Evans (eds.) 2017. *Oxford Handbook of Polysynthesis*. Oxford: Oxford University Press.
- Greenberg, Joseph H. 1960 A quantitative approach to the morphological typology of language. *International Journal of American Linguistics* 26(3). 178–194.
- Haspelmath, Martin. 2011. The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica* 45(1). 31–80.
- Haspelmath, Martin. 2018. The last word on polysynthesis: A review article of the *Oxford Handbook of Polysynthesis*. *Linguistic Typology* 22(2). 307–326.
- Haspelmath, Martin. To appear. Inflection and derivation as traditional comparative concepts. To appear in *Linguistics*.
- Hewitt, B. George. 1979a. The relative clause in Abkhaz (Abzui dialect). *Lingua* 47. 151–188.
- Hewitt, B. George. 1979b. The relative clause in Adyghe (Temirgoi dialect). *Annual of Ibero-Caucasian Linguistics* 6. 134–162.
- Hewitt, B. George. 2005. North West Caucasian. *Lingua* 119. 91–145.
- Istratkova, Vyara. 2004. On multiple prefixation in Bulgarian. *Nordlyd Special Issue on Slavic Prefixes* 32(2). 301–321.

- Kageyama, Taro. 2016. Verb-compounding and verb-incorporation. In: Taro Kageyama & Hideki Kishimoto (eds.), *Handbook of Japanese Lexicon and Word-Formation*, 273–310. Berlin, Boston: De Gruyter Mouton.
- Kibrik, Andrej A. 2011. *Reference in Discourse*. Oxford: Oxford University Press.
- Klychev, Rauf N. 1972. *Glagol'nye suffiksy napravlenija v abazinskom i abxazskom jazykax*. [Verbal suffixes expressing direction in Abaza and Abkhaz]. Cherkessk: Stavropol'skoe knižnoe izdatel'stvo, Karačaevo-Čerkesskoe otdelenie.
- Klychev, Rauf N. 1994. *Lokal'no-preverbnoe obrazovanie glagolov abazinskogo jazyka*. [The locative preverbal derivation of verbs in Abaza]. Cherkessk: Adž'pa.
- Klychev, Rauf N. 1995. *Slovar' sočetaemosti lokal'nyx preverbov s suffiksoidami i glagol'nymi kornjami v abazinskom jazyke*. [The collocational dictionary of locative preverbs with suffixoids and verbal roots in Abaza]. Cherkessk: Karačaevo-čerkesskoe knižnoe izdatel'stvo.
- Korotkova, Natalia & Yury Lander. 2010. Deriving suffix ordering in polysynthesis: Evidence from Adyghe. *Morphology* 20. 299–319.
- Kumakhov, Mukhadin A. 1964: Кумахов М. А. *Morfologija adygskich jazykov. Sinxronno-diaxronnaja xarakteristika. I. Vvedenie, struktura slova, slovoobrazovanie častej reči*. [Morphology of Circassian languages. A synchronic and diachronic characteristic. I. Introduction, word structure, derivation of parts of speech.]. Nal'čik: Kabardino-balkarskoe knižnoe izdatel'stvo.
- Kumakhov, Mukhadin A. 1989. Кумахов М. А. *Sravnitel'no-istoričeskaja grammatika adygskich (čerkesskich) jazykov*. [Historical-comparative grammar of the Circassian languages]. Moscow: Nauka.
- Kumakhov, Mukhadin A. (ed.) 2006. *Kabardino-čerkesskij jazyk* [The Kabardian Language]. Vol. I. Nalchik: El-Fa.
- Kumakhov, Mukhadin A. & Vamling, Karina. 2009. *Circassian Clause Structure*. Malmö: Malmö University.
- Lambrecht, Knud. 1981. *Topic, Antitopic and Verb Agreement in Non-Standard French*. Amsterdam, Philadelphia: John Benjamins.
- Lander, Yury. 2012. *Reljativizacija v polisintetičeskom jazyke: adygejskie odnositel'nye konstrukcii v tipologičeskoj perspektive*. [Relativization in a polysynthetic language: West Circassian relative clauses in typological perspective]. PhD Dissertation, Institute of Oriental Studies of the Russian Academy of Sciences, Moscow.
- Lander, Yury. 2016. Word formation in Adyghe. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds), *Word-Formation. An International Handbook of the Languages of Europe*. Vol. 5, 3508–3526. Berlin: Mouton de Gruyter.
- Lander, Yury. 2017. Nominal complex in West Circassian: Between morphology and syntax. *Studies in Language* 41(1). 76–98.
- Lander, Yury & Michael Daniel. 2019. West Caucasian relative pronouns as resumptives. *Linguistics* 57(6). 1239–1279.
- Lander, Yury & Alexander Letuchiy. 2010. Kinds of recursion in Adyghe morphology. In Harry van der Hulst (ed.), *Recursion in Human Language*, 263–284. Berlin, New York: Mouton de Gruyter.
- Lander, Yury & Alexander Letuchiy. 2017. Decreasing valency-changing operations in a valency-increasing language? In Albert Álvarez González & Iñaki Navarro (eds.), *Verb Valency Changes. Theoretical and typological perspectives*, 286–304. Amsterdam, Philadelphia: John Benjamins.
- Letuchiy, Alexander. 2007. Reciprocals, reflexives, comitatives, and sociatives in Adyghe. In Vladimir P. Nedjalkov (ed.), *Reciprocal constructions*, 773–812). Amsterdam, Philadelphia: John Benjamins.
- Letuchiy, Alexander. 2009. Affiksy benefaktiva i malefaktiva: sintaksičeskie osobennosti i krug upotreblenij [Benefactive and malefactive markers: Syntactic peculiarities and range of uses]. In: Yakov Testelets (ed.), *Aspekty polisintetizma: Očerki po grammatike adygejskogo jazyka* [Aspects of polysynthesis: Studies in the grammar of West Circassian], 329–371. Moscow: Russian State University for the Humanities.

- Letuchiy, Alexander. 2012. Ergativity in the Adyghe system of valency-changing derivations. In: Gilles Authier & Katharina Haude (eds.), *Ergativity, Valency and Voice*, 323–354. Berlin, New York: Mouton de Gruyter.
- Mahieu, Marc-Antoine & Nicole Tersis (eds.). 2009. *Variations on Polysynthesis. The Eskaleut Languages*. Amsterdam, Philadelphia: John Benjamins.
- Maisak, Timur. 2016. Morphological fusion without syntactic fusion: The case of the “verificative” in Agul. *Linguistics* 54(4). 815–870.
- Mattissen, Johanna. 2003. *Dependent-Head Synthesis in Nivkh. A Contribution to a Typology of Polysynthesis*. Amsterdam, Philadelphia: John Benjamins.
- Mattissen, Johanna. 2004. A structural typology of polysynthesis. *Word* 55(2). 189–216.
- Mattissen, Johanna. 2006. The ontology and diachrony of polysynthesis. In Dieter Wunderlich (ed.), *Advances in the Theory of the Lexicon*, 287–353. Berlin, New York: Mouton de Gruyter.
- Mattissen, Johanna. 2017. Sub-types of polysynthesis. In: Fortescue et al. (eds.), 70–98.
- Mithun, Marianne. 1988. System-defining structural properties in polysynthetic languages. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 41(4). 442–452.
- Mithun, Marianne. 1997. Lexical affixes and morphological typology. In John Haiman, Joan Bybee & Sandra Thompson (eds.), *Essays on Language Function and Language Type. Dedicated to T. Givón*, 357–372. Amsterdam, Philadelphia: John Benjamins.
- Mithun, Marianne. 1998. Yup'ik roots and affixes. In Osahito Miyaoka & Minoru Oshima (eds.), *Languages of the North Pacific Rim*, Vol. 4, 63–76. Kyoto: Kyoto University Graduate School of Letters.
- Mithun, Marianne. 2010. The search for regularity in irregularity: Defectiveness and its implications for our knowledge of words. In Matthew Baerman, Greville Corbett & Dunstan Brown (eds.), *Defective paradigms: Missing Forms and What they Tell us*, 125–149. Oxford: Oxford University Press.
- Mithun, Marianne. 2016. Affix ordering: Motivation and interpretation. In Andrew Hippisley & Gregory Stump (ed.), *Cambridge Handbook of Morphology*, 149–185. Cambridge: Cambridge University Press.
- Mithun, Marianne. 2017. Polysynthesis in North America. In: Fortescue et al. (eds.), 235–259.
- Moreno Cabrera, Juan Carlos. 2014. El español hablado como lengua aglutinante y polisintética. In José María Santos Rovira (ed.), *Ensayos de Lingüística Hispánica*, 177–216. Lisboa: Faculdade de Letras, Universidade de Lisboa.
- Nichols, Johanna. 1986. Head-marking and dependent-marking grammar. *Language* 62(1). 56–119.
- Nichols, Johanna. 2017. Polysynthesis and head marking. In Fortescue et al. (eds.), 59–69.
- O’Herin, Brian. 2002. *Case and Agreement in Abaza*. Arlington: SIL International & University of Texas Press.
- Olthof, Marieke. 2020. *Incorporation: Constraints on Variation*. Amsterdam: LOT Publications. <https://www.lotpublications.nl/incorporation-constraints-on-variation>
- Osborne, C. R. 1974. *The Tiwi Language. Grammar, Myths and Dictionary of the Tiwi Language spoken on Melville and Bathurst Islands, Northern Australia*. Canberra: Australian Institute for Aboriginal Studies.
- Panova, Anastasia. 2018. Derivational verbal suffixes in Abaza. Higher School of Economics Working Papers. Series: Linguistics. No. 70. <https://publications.hse.ru/en/preprints/229493762>
- Panova, Anastasia. 2020a. Morfoložičeski svjazannaja komplementacija v abazinskom jazyke [Morphologically bound complementation in Abaza]. *Voprosy jazykoznanija* 4. 87–114.
- Panova, Anastasia. 2020b. A case of morphologically bound complementation in Abaza: an LFG analysis. In Miriam Butt & Ida Toivonen (eds.), *Proceedings of the LFG’20 Conference*, 289–306. Stanford: CSLI Publications.
- Paris, Catherine. 1995. Localisation en tcherkesse: forme et substance du référent. In André Rousseau (éd.), *Les Préverbes dans les langues d’Europe. Introduction à l’étude de la préverbativité*, 345–379. Lille: Presse Universitaires de Septentrion.
- Rogava, Giorgi V. & Zejnab I. Kerasheva. 1966. *Grammatika adygejskogo jazyka* [A grammar of West Circassian]. Krasnodar: Krasnodarskoe knižnoe izdatel’stvo.
- Sadock, Jerrold. 2017. The subjectivity of the notion of polysynthesis. In Fortescue et al. (eds.), 99–114.

- Saltarelli, Mario. 1988. *Basque*. London etc.: Croom Helm.
- Shagal, Ksenia. 2019. *Participles. A Typological Study*. Berlin, Boston: De Gruyter Mouton.
- Siewierska, Anna. 2005. Verbal person marking. In Matthew S. Dryer, Martin Haspelmath, David Gil & Bernard Comrie (eds.), *The World Atlas of Language Structures*, 414–417. Oxford: Oxford University Press.
- Skorik, Petr Ya. 1961. *Grammatika čukotskogo jazyka. Část' I. Fonetika i morfologija imennyx častej reči* [A grammar of Chukchi. Part I. Phonetics and morphology of nominals]. Moscow: Izdatel'stvo Akademii Nauk SSSR.
- Spencer, Andrew. 2013. *Lexical Relatedness*. Oxford: Oxford University Press.
- Stump, Gregory. 2006. Template morphology. In Keith Brown (ed.), *Encyclopedia of Language and Linguistics*, Vol. 12, 559–563. Oxford: Elsevier.
- Sumbatova, Nina. 2009. Constituent questions and argument-focus constructions: some data from the North-Caucasian languages. In: Johannes Helmbrecht, Yoko Nishima, Yong-Min Shin, Stavros Skopeteas & Elisabeth Verhoeven (eds.). *Form and Function in Language Research. Papers in Honour of Christian Lehmann*, 313–328. Berlin, New York: Mouton de Gruyter.
- Tallman, Adam J. R. 2020. Beyond grammatical and phonological words. *Language and Linguistics Compass*, 2020;e12364.
- Tallman, Adam J. R. 2021. Constituency and coincidence in Chácobo. *Studies in Language* 45(2). 321–383.
- Testelets, Yakov & Yury Lander. 2017. Adyghe (Northwest Caucasian). In Fortescue et al. (eds.), 948–970.
- Trudgill, Peter. 2017. The anthropological setting of polysynthesis. In: Fortescue et al. (eds.), 186–202.
- Vaida, Edward. 2017. Polysynthesis in Ket. In: Fortescue et al. (eds.), 906–929.
- Vinyar, Alexey. 2021. *Noun Incorporation in Chukchi and Beyond: A Force-Dynamic Constructionist Approach*. MA Thesis, Higher School of Economics, Moscow. <https://www.academia.edu/50267979>
- Woodbury, Anthony C. 2017. Central Alaskan Yupik (Eskimo-Aleut). A sketch of morphologically orthodox polysynthesis. In: Fortescue et al. (eds.), 536–559.
- Zúñiga, Fernando. 2017. On the morphosyntax of indigenous languages of the Americas. *International Journal of American Linguistics* 83(1). 111–139.
- Zúñiga, Fernando. 2019. Polysynthesis: a review. *Language and Linguistics Compass* 13:e12326.

Motivating a Morphome: Albanian case syncretism as a case study

Dorđe Božović
University of Belgrade
djordje.bozovic@fil.bg.ac.rs

1. Introduction

The notion of *morphomes*, going back to Aronoff (1994), figures prominently in the debates about the autonomous status of morphology and the nature of its interfaces with other modules of grammar, syntax in particular. A morphological pattern is said to be *morphomic* when no explicit motivation for it can be found outside of morphology itself (Corbett 2015, 2016), that is, when its existence cannot be explained away by phonological (e.g. shape of the stem) or semantic conditioning (i.e. feature composition)—what Aronoff (1994) has called “pure morphology”. Although for Aronoff *all* of morphology is ultimately morphomic (i.e. “unnatural”; cf. now also Aronoff 2016), morphomic patterns (or splits, in Corbett’s terms) contrast with (externally) *motivated* ones, and the issue at hand in much of the literature on morphomes so far is how to distinguish the two kinds of phenomena.¹

However, it has been observed that the distinction between morphomic and motivated may actually be a more fine-grained one, or a scale rather than a dichotomy (see Smith 2013 and other contributions to that volume; now also Herce 2020a). Here, I will present one case study in that vein, with data from Albanian (IE Balkan language), in which apparently morphomic patterns of case syncretism in noun inflection have been produced in fact by an external (syntactic) motivation. In that sense, I will argue that the Albanian noun inflection is both morphomic *and* motivated at the same time, thus being a curious case of a “motivated morphome” (sic!).

The paper is organized as follows. In Section 2, the theoretical background, concepts and their definitions are briefly introduced and explained. This is to set out the foundation for our case study, which is detailed in Section 3. Finally, some tentative conclusions are drawn from there in Section 4.

2. Motivated vs. morphomic splits

In a discussion of what constitutes a canonical morphome, O’Neill (2011a, 2011b, 2013) gives the following definition for the concept: a “regular distribution of identical form, usually an allomorphic root/stem, which does not correspond to any coherent generalization or function, phonological, semantic or syntactic” (O’Neill 2013: 221 et seq.). As a negative definition, then, the definition of a morphome depends crucially on our understanding of what does constitute a “coherent generalization or function”, that is, what counts as motivation for a morphological pattern. According to Corbett (2016), paradigmatic splits are motivated if they “correspond to morphosemantic, morphosyntactic or phonological specifications” and ultimately “motivation is justified by appeal to a natural class” (Corbett 2016: 85). A morphome (morphomic pattern

¹ In addition to the works already referred to here, cf. also the other contributions in Luís & Bermúdez-Otero (2016) for a more recent discussion.

or morphomic split), therefore, is any regular pattern that does not form a natural class, which is usually defined in terms of featural makeup:

“[...] motivation is most easily seen by reference to natural classes in the feature system. In a reasonable feature system, perfective forms versus imperfective constitute natural classes, as do past versus nonpast, singular versus plural, and so on. Motivated segments of a paradigm are sometimes called ‘subparadigms’. By contrast, first-person plural is not a natural class, since it requires reference both to person and to number. Anything beyond natural classes requires an extra step, and so needs additional justification” (Corbett 2015: 163).

A number of such morphomic patterns have been identified in the literature and argued to be psycholinguistically real and diachronically persistent; albeit mostly for Romance languages, following Maiden (2005; see also Maiden 2018, 2021 for the most recent surveys).² One such pattern in Romance verb inflection is the so-called “L-pattern”, identified by Maiden (2018), in which stem allomorphy splits the verbal paradigm in two halves in such a way that only 1SG present indicative and all persons of the subjunctive regularly feature a palatalized allomorph, while all the remaining persons of the present indicative have a non-palatalized stem. Identical forms, in this case allomorphs of the stem, are thus regularly distributed in a way that fails to form a natural class, because neither the combination 1SG.PRS.IND+PRS.SBJV nor PRS.IND minus 1SG qualify as valid subparadigms; they both require an “extra step” to be defined. Interestingly though, in spite of that, this is a stable pattern throughout all of Romance. A subset of examples from Portuguese, with the L-shaped morphomic pattern marked in bold, is reproduced here in Table 1.

Table 1: The “L-pattern” in Portuguese verbal inflection (from Maiden 2018: 86)

	1SG	2SG	3SG	1PL	2PL	3PL
PRS.IND	tenho ‘have’	<i>tens</i>	<i>tem</i>	<i>temos</i>	<i>tendes</i>	<i>têm</i>
PRS.SBJV	tenha	tenhas	tenha	tenhamos	tenhais	tenham
PRS.IND	vejo ‘see’	<i>vês</i>	<i>vê</i>	<i>vemos</i>	<i>vedes</i>	<i>vêem</i>
PRS.SBJV	veja	vejas	veja	vejamos	vejais	vejam
PRS.IND	faço ‘do’	<i>fazes</i>	<i>faz</i>	<i>fazemos</i>	<i>fazeis</i>	<i>fazem</i>
PRS.SBJV	faça	faças	faça	 façamos	façais	façam
PRS.IND	venho ‘come’	<i>vens</i>	<i>vem</i>	<i>vimos</i>	<i>vindes</i>	<i>vêm</i>
PRS.SBJV	venha	venhas	venha	venhamos	venhais	venham
PRS.IND	meço ‘measure’	<i>medes</i>	<i>mede</i>	<i>medimos</i>	<i>medis</i>	<i>medem</i>
PRS.SBJV	meça	meças	meça	meçamos	meçais	meçam
PRS.IND	caibo ‘fit’	<i>cabes</i>	<i>cabe</i>	<i>cabemos</i>	<i>cabeis</i>	<i>cabem</i>
PRS.SBJV	caiba	caibas	caiba	caibamos	caibais	caibam

² Even when they do discuss the data from outside the Romance family, most of the analyses of morphomic patterns in the literature focus exclusively on single languages or language branches. A wider typological survey has not been undertaken until Herce (2020b), where as much as 110 morphomic structures have been identified across the world’s languages, in addition to several important cross-linguistic generalizations on morphemes.

Recently, Round (2015) has identified three kinds of possible morphomic phenomena, termed *rhizomorphomes*, *meromorphomes*, and *metamorphomes*. Rhizomorphomes are morphomic patterns realized at the level of inflectional classes of words, insofar as they are lexically determined, i.e. unmotivated from outside of morphology itself. Meromorphomes are “categories which mediate between morphosyntactic feature structures and the phonological operations by which individual pieces of individual word forms are composed” (Round 2015: 48). Metamorphomes, in turn, are realizations of meromorphomes in specific paradigms which consist of regular patterns of formal identity between pieces of a paradigm (like the L-pattern in Table 1 above), that are similarly unmotivated or “purely morphological”.

In addition to stem allomorphy, another typical instance of a metamorphomic pattern, in the sense of Round (2015), is syncretism. Following the Jakobsonian tradition of featural decomposition of Russian case forms (Jakobson 1962, 1984), syncretism is often represented via feature underspecification (Caha 2019). However, when a syncretism pattern lacks such motivation in terms of featural makeup, as for instance, when it splits the paradigm into unnatural classes, it has been often used as an evidence that morphological structures are autonomous, even outside of the literature on morphemes (cf. Baerman 2004; Baerman, Brown & Corbett 2005, *inter alia*). In the following section, I will describe in more detail such apparently metamorphomic patterns of case syncretism in Albanian noun inflection.

3. Case study: Albanian case syncretism

In this section, I focus on Modern Standard Albanian (MSA) noun inflection as a case study of an externally motivated (meta)morphomic pattern. First I will argue that MSA syncretism patterns are indeed morphomic, in the sense that they form unnatural classes which cannot be possibly defined in terms of feature composition. Then I will provide a synchronic motivation in the syntax for precisely such a morphomic distribution of Albanian case forms, arguing that the MSA metamorphome under investigation is in fact both motivated *and* morphomic in the relevant sense.

MSA nouns have three genders which roughly correspond to three inflectional classes in the singular, traditionally termed masculine, feminine and neuter. Masculines (M) take the NOM.SG.DEF suffix *-i* or (phonologically conditioned) *-u*, while feminines (F) have the NOM.SG.DEF in *-a* and neuters (N) in *-t* (with phonologically conditioned variants *-it* and *-të*). Although N is a productive class for deverbal and deadjectival substantivizations with the prepositive article *të* (e.g. *të ardhur-it* ‘arrival’ ← participle *ardhur* ‘to arrive’, *të ftohtë-t* ‘coldness’ ← adjective *i/e ftohtë* ‘cold’), other than those it has lost most of the inherited neuters from Old Albanian, which are inflected as M instead in the modern language (e.g. *vaj-i* ‘oil’, *mish-i* ‘meat’ for the older *vaj-të* and *mish-të* etc.), so it is often said to be in decline (cf. Agalliu 2002; Buchholz & Fiedler 1987; Newmark, Hubbard & Prifti 1982).

MSA has two numbers, singular (SG) and plural (PL). Formation of the plural stems is highly irregular for most nouns and more derivation-like than inflection-like (Bozhoviq 2021, with references therein). All nouns inflect the same in the plural, however, regardless of their gender, taking the same set of case suffixes and the definiteness suffix *-t* (or its phonologically conditioned variants *-it* and *-të*). In addition, in some cases, gender agreement in the plural may differ from the pattern of the corresponding singular noun (as in *shtet-i* ‘state’ vs. *Shtetet e Bashkuar-a* ‘United-F States’), showing that gender is truly an inherent property of the plural stems rather than lexemes. Therefore, counting SG and PL inflections separately, there are a total of four inflectional classes in MSA, marked traditionally according to the NOM.DEF suffix: M.SG (*-i/u*), F.SG (*-a*), N.SG (*-t*) and a PL (also *-t*) class.

In both the SG and the PL, MSA nouns inflect for case and definiteness. Indefinite forms (INDEF) are unmarked, the definite ones (DEF) take special suffixes. Nonetheless, due to

pervasive mergers throughout the paradigm, the exact number of cases is often debated in the Albanological literature (see e.g. Përnaska 2003). At most four morphologically distinct case forms may be identified, however. These are NOM, ACC, DAT and ABL. MSA noun inflection is summarized in Table 2.

Table 2: Modern Standard Albanian noun inflection

	M.SG		F.SG		N.SG		PL	
	INDEF	DEF	INDEF	DEF	INDEF	DEF	INDEF	DEF
NOM	-Ø	-i/u	-Ø	-a	-Ø	-t/ -it/-të	-Ø	-t/ -it/-të
ACC		-in/-un		-n/-në				-it/-të
DAT	-i/u	-it/-ut	-e	-s/-së	-i	-it/-të	-ve	
ABL							-sh	

3.1. Evidence for morphomic splits

Let us now focus on the patterns of syncretism in Table 2. First, it is obvious that the paradigm is split along the lines of the core/non-core (i.e. structural/inherent) case distinction, while both are syncretic: there are two major mergers in the paradigm, viz. NOM/ACC merger on the one and DAT/ABL merger on the other hand. This is a motivated split, which can have morphomic splits nested inside, according to Corbett (2016). Neither of these two mergers is actually complete, though. ACC is still kept formally distinct from NOM in the M.SG.DEF and F.SG.DEF inflections, by virtue of the ACC.SG.DEF suffix *-n* (and its phonological variants), and the ABL.PL.INDEF form in *-sh* remains the single non-syncretized cell in the entire DAT/ABL subparadigm. In addition to that, NOM.INDEF and ACC.INDEF forms in the M and F classes trigger different case agreement on their agreement probes despite formal identity; cf. the shape of the ezafe-like linker morpheme (LNK) in (1a) and (2a) versus (1b) and (2b), respectively.³

- (1) a. *Ky është një djalë i mirë.*
 this.M be.3SG.PRS a boy.NOM.SG.INDEF LNK good
 ‘This is a good boy.’
- b. *E=pashë një djalë të mirë.*
 3SG.ACC=see.1SG.AOR a boy.ACC.SG.INDEF LNK good
 ‘I saw a good boy.’
- (2) a. *Kjo është një vajzë e mirë.*
 this.F be.3SG.PRS a girl.NOM.SG.INDEF LNK good
 ‘This is a good girl.’
- b. *E=pashë një vajzë të mirë.*
 3SG.ACC=see.1SG.AOR a girl.ACC.SG.INDEF LNK good
 ‘I saw a good girl.’

As regards the noun form, though, NOM and ACC are both systematically unmarked and regularly merged throughout the indefinite, as well as N.DEF and PL.DEF paradigms. In other

³ It should be noted that the Albanian linkers themselves represent a closed system with only four forms: *i*, *e*, *të* and *së*, alternating as exponents of the various combinations of gender, number, case and definiteness features. At moments this system seems to make more distinctions than noun inflection does, as in the examples (1) and (2) above, but for the most part linkers are even more underspecified than the corresponding noun forms. No morphomic patterns can be identified, as the shape of the linkers falls out entirely from their feature composition.

words, MSA NOM/ACC merger is a *metasyncretism* of a kind identified by Williams (1994) as a (meta)pattern pertaining to different paradigms (or in Williams's terms, a metaparadigm). If one agrees with Aronoff (1994: 25) and Corbett (2016: 72) that even single cells may be morphomic, in the sense that, as singletons, both they and the remainder of the paradigm minus that one cell, form unnatural classes, it may be argued that the single non-syncretized cell in this metapattern, viz. the ACC.SG.DEF one, is also a morphomic split of a kind, nested within a motivated one.

The other merger, the one of DAT and ABL, also has an apparent morphomic split nested inside. That is the L-shaped syncretic pattern in DAT/ABL.PL. Syncretism here, too, regularly affects DAT.PL and ABL.PL cells, but with the exclusion of a single cell, viz. ABL.PL.INDEF in *-sh*, thus forming unnatural class consisting of DAT.PL.INDEF, DAT.PL.DEF and ABL.PL.DEF, to the exclusion of ABL.PL.INDEF.

In addition to this, there is also a formal identity between DAT/ABL.INDEF and NOM.DEF in the M.SG and the F.SG inflections. In M.SG, both of these forms end in *-i/u*, while in F.SG the formal identity is obscured by a phonological change that has affected the original NOM.SG.DEF suffix **-e* for F nouns (still preserved as the corresponding form of the agreeing LNK morpheme, as in (2a)) in hiatus formed with the stem-final vowel, coalescing them both into *-a* (cf. Topalli 2009: 207-208). This may seem as a purely accidental syncretism, if only it wasn't fully regular and of a metasyncretic character (i.e. unifying the paradigms of M.SG and F.SG underlyingly, regardless of the exact surface form of their suffixes that make up the pattern). Needless to say, as a split involving NOM.DEF and DAT/ABL.INDEF, it forms a very unnatural class.⁴

Another possible metasyncretism in the MSA noun inflection in Table 2 could be the one involving N.SG.DEF and PL.DEF, which are both marked with the suffix *-t* (*-it/të*).⁵ A connection between N and PL is semantically plausible in Albanian (with N typically covering various abstract and mass nouns). Unifying N.SG and PL (and conversely, M.SG and F.SG) into a single metaparadigm can also be corroborated by the almost mirror-like distribution of forms more generally in the N.SG and PL paradigms on the one, and the M.SG and F.SG paradigms (with distinct ACC and the syncretism of DAT/ABL.INDEF and NOM.DEF) on the other hand. If so, this would be another motivated split, inside which however the aforementioned metasyncretism of DAT/ABL.INDEF and NOM.DEF in the SG metaparadigm is nested as a (minor) morphomic one.

Leaving clearly motivated syncretisms aside,⁶ the remaining candidates for morphomic splits in MSA noun inflection that have been discussed so far are summarized visually in Table 3, by shading all the cells that form a particular pattern.

⁴ Compare a similarly odd syncretism involving GEN.SG and NOM.PL, which was often considered accidental in the literature, but is nonetheless notoriously recurring in different language families and eventually even turns out to be motivated, as shown by Caha (2016; cf. also Caha 2019 and references therein).

⁵ Note also that there are Tosk Albanian dialects in which the suffix *-t* is generalized across the PL.DEF inflection (so that DAT/ABL.PL.DEF has the desinence *-vet*), and the inherited neuters such as *vaj-të*, *mish-të* etc. are also better preserved there (cf. Çerpja 2017 for an overview).

⁶ Alternatively, given the all-pervading mergers in MSA noun inflection, one could say that motivated syncretisms are shown in Table 3 as well, only by lack of any specific shading. This is actually significant for fully comprehending the nature of MSA system of inflection: it is truly a (*meta*)system of syncretisms, as will be discussed in what follows shortly (§ 3.2).

Table 3: Morphomic patterns in MSA noun inflection

	M.SG		F.SG		N.SG		PL	
	INDEF	DEF	INDEF	DEF	INDEF	DEF	INDEF	DEF
NOM								
ACC								
DAT								
ABL								

As has already been said, none of the shaded patterns in Table 3 form a natural class. If motivation for a paradigmatic split is understood to mean “reference to natural classes in the feature system” (Corbett 2015, 2016), then the (meta)syncretism patterns in Table 3 cannot be motivated and therefore must be considered morphomic.

3.2. Evidence for external motivation

While it is true that the paradigmatic splits in Table 3 are unnatural, I have argued in Božović (2021), however, that MSA case syncretisms of the kind shown in Table 3 are not just “purely formal”, in the sense that they actually play a role at the morphology-syntax interface.

Namely, it should be noted that the morphomic distributions in MSA noun inflection are a product of a specific interaction of two categories, viz. case and definiteness. Both case and definiteness are categories of contextual (i.e. required by the syntax, as per Booij 1994, 1996) inflection in MSA. This may be seen in the following examples (3-4).

- (3) a. *Vajz-a* *është* *e* *mençur*.
 girl-NOM.SG.DEF be.3SG.PRS LNK smart
 ‘(A/the) girl is smart.’
- b. **Vajzë* *është* *e* *mençur*.
 girl.INDEF be.3SG.PRS LNK smart
- (4) a. *Vjollc-a* *është* *studente*.
 Vjollca-NOM.SG.DEF be.3SG.PRS student-F.SG.INDEF
 ‘Vjollca [a female personal name] is a student.’
- b. **Vjollc-a* *është* *student-ja*.
 Vjollca-NOM.SG.DEF be.3SG.PRS student-F.SG.DEF

A subject NP has to be definite in addition to bearing the NOM case, as in (3a) vs. (3b), and this is true for both common and proper nouns, for which cf. (4), i.e. regardless of their inherent semantics, showing that definiteness in MSA is truly a category of contextual (that is to say, bound to marking syntactic relations) rather than inherent inflection. Therefore, in principle, a morphologically definite form in MSA may be ambiguous with respect to the referential or non-referential readings, as is also shown by the English translation of (3a).⁷ A predicative NP, on the other hand, has to be indefinite; cf. (4a) vs. (4b). Now compare this with the ACC forms marking various kinds of objects in (5-7).

⁷ For a more detailed contrastive study of noun definiteness in Albanian and English, cf. Backus Borshi (2015).

- (5) a. (*E=*)*kam* *punë-n* *e* *rëndë*.
 3SG.ACC=have.1SG.PRS work-ACC.SG.DEF LNK heavy
 ‘I have a lot of work to do.’
- b. *(*E=*)*kam* *punë* *të* *rëndë*.
 3SG.ACC=have.1SG.PRS work-INDEF LNK heavy
- (6) a. *A* *ke* *uri?*
 Q have.2SG.PRS hunger-ACC.SG.INDEF
 ‘Are you hungry?’
- b. **A* *ke* *uri-në?*
 Q have.2SG.PRS hunger-ACC.SG.DEF
- (7) a. *Jetoj* *në* *Evropë* (*Jugor-e*).
 live.1SG.PRS in Europe-ACC.SG.INDEF Southern-F
 ‘I live in (Southern) Europe.’
- b. *Jetoj* *në* *Evropë-n* *(*Jugor-e*).
 be.1SG.PRS from Europe-ACC.SG.DEF Southern-F

While here it is possible in principle to have either a definite or an indefinite object phrase, depending on its semantics and pragmatics, specific syntactic configurations, such as those involving optionality vs. obligatoriness of object clitic doubling (5), certain phraseological constructions (6), various noun modification strategies and prepositions governing the ACC (7), actually systematically disallow one of the options. In other words, there is a specific *division of labour* between case and definiteness, making use of this additional distinction provided by the morphology in order to signal some of the syntactic relations in the functional domain of cases, which in turn reduces the overall number of necessary distinct forms in the (singular) metaparadigm to just three: two of them marked, non-syncretic ones, viz. NOM.DEF (for marking subjects) and ACC.DEF (for objects made either semantically or pragmatically specific), and the third “elsewhere” (i.e. syncretic) form.

This similarly holds for the DAT/ABL merger, as well. The only syntactic position in which DAT (typically marking indirect objects) and ABL (typically marking complements of various prepositions) systematically contrast with each other (only this time in the PL paradigm) is that of a modifier/complement of a DEF vs. INDEF head noun; cf. (8-9).

- (8) a. *dru* *lisa-sh*
 tree.INDEF oaks-ABL.PL.INDEF
 ‘oak tree(s)’
- b. **drur-i* *lisa-sh*
 tree-DEF oaks-ABL.PL.INDEF
- (9) a. *drur-i* *i* *lisa-ve* *qindra-vjeçare*
 tree-DEF LNK oaks-DAT/ABL.PL.DEF hundred-year-olds
 ‘the tree of the hundred-year-old oaks’

b. <i>dru</i>	<i>i</i>	<i>lisa-ve</i>
tree.INDEF	LNK	oaks-DAT/ABL.PL.DEF
‘oak tree [e.g. as a material]’		

While the syncretic DAT/ABL modifier (with a linker) can combine with both an INDEF and a DEF head noun in different syntactic configurations, as in (9), the non-syncretic ABL.INDEF is reserved for INDEF contexts only (8a) and cannot modify a DEF noun (8b). Here too, a division of labour between case and definiteness has worked to produce a reduced number of distinct forms, delegating some of the functions of cases to the distinction in definiteness, resulting thus in an L-shaped morphomic (meta)paradigm, with just two distinct forms: a single non-syncretized ABL.PL.INDEF one, and an “elsewhere” one, oblivious as regards the case, but contrasting in definiteness.

According to Božović (2021), the division of labour between case and definiteness in MSA has led thus to a specific *complementary distribution* of the syncretized and non-syncretized forms with respect to their syntactic functions. Namely, forms such as ACC.SG.DEF and ABL.PL.INDEF are kept formally distinct only in (morpho)syntactically ambiguous contexts, as in (5a), where the object clitic is syntactically optional, in (7) with an overt modifier, or in (8a) and (9b), with an INDEF head noun. Only in such contexts, the case/definiteness distinction has to be formally maintained, because it remains the only overt signal of a syntactic relation. If there is, however, any other strategy of syntactic function coding available, such as the obligatory object clitic doubling in (5b), a phraseologically fixed VO construction as in (6), and the like, then the noun (NP) need not mark a case distinction overtly; instead, it can revert to the syncretic “elsewhere” form, and thus maintain a laudably high level of language economy.

This equally holds for the motivated as well as morphomic mergers in MSA. Crucially, however, it is precisely this kind of merging forms that are in a complementary syntactic distribution, so as to reduce the number of necessary distinct forms to an “optimal” minimum, that as an effect produces in turn unnatural classes of the kind we have observed in Section 3.1 above. Recall, for instance, the DAT/ABL.SG.INDEF+NOM.SG.DEF morpheme. There is not a single syntactic context in which the exponents of these values would ever compete for the same position. This is, however, exactly what allows them to formally syncretize, as instead of having to mark all the relevant contrasts formally on the noun, speakers can rely on the specific syntactic configurations to distinguish the necessary functions. In return, the necessary number of distinct inflectional forms is maximally reduced, but the resulting distribution of identical forms within a paradigm necessarily produces unnatural classes, i.e. morphomic splits. In that sense, the incomplete mergers of NOM/ACC and DAT/ABL in MSA, as well as the apparent formal identity of DAT/ABL.SG.INDEF and NOM.SG.DEF, are disturbingly both relevantly morphomic *and* motivated by language economy.

4. Conclusion

In this paper, I have argued that the distribution of syncretized and non-syncretized forms in Albanian noun inflection is motivated by mechanisms of language economy, driven by the division of labour between case and definiteness in the syntax, in such a way that syncretism is used as a means to produce an “optimal” (that is, maximally economical) distribution of formally distinct case/definiteness forms for each (sub)paradigm. In turn, this creates several metamorphomic patterns, in which, as a rule, syncretized forms never make up a natural class (e.g. NOM and ACC, but with the exception of ACC.SG.DEF, or DAT and ABL with the exception of ABL.PL.INDEF, or a rather strange merger of DAT/ABL.SG.INDEF and NOM.SG.DEF).

In other words, it is precisely the morpheme that, far from being “useless” and “arguably increas[ing] the complexity of the system with no obvious corresponding return” (Corbett 2016:

64), actually plays a crucial role in the organization of forms in the language. In that way, Albanian noun inflection, with its pervasive case syncretisms, features non-trivial splits that are both morphomic *and* (externally) motivated at the same time.

It is important to note that syncretism, as a means of maintaining this maximal economy, couldn't work this way if it *didn't* produce unnatural classes, such as those discussed here; in that case, its power to maximally economically organize the system of forms would be significantly reduced, if not lost. This is why, in the end, *morphomic* (in the sense of forming an unnatural class) and (externally) *motivated* should not be understood as a total dichotomy: here we have seen that, in the case of Albanian case syncretism, a syntactic (functional) drift may actually feed and itself rely on morphomic distributions of forms.

The analysis provided here for Albanian, therefore, may contribute to the “morpheme debate” in morphology, which is still almost exclusively dominated by the data from Romance, and to a better understanding of the morphology-syntax interface in general, as well as to the literature on (meta)syncretism patterns and the morphosyntax of the Balkan Sprachbund noun phrase in particular.

On a final (side) note, it was already pointed out by Newmark (1962), some sixty years ago, that the Albanian case system is in fact a “combinatorial” one, in which case and definiteness interact so as to reduce the number of necessary distinct forms; in what was essentially a proto-derivational account of inflection:⁸

“In traditional descriptions of Albanian the essential simplicity of the case system is obscured by mixing together information about the morphological structure, the syntactic distribution, and the semantic functioning of the case form. By treating these aspects of linguistic structure separately but in relation to one another, a combinatorial description may reveal underlying regularities of structure in each aspect, without sacrificing a view of the complex integrity of the language itself” (Newmark 1962: 321).

In so many aspects this short article resonates with the present issues.

Acknowledgements

In addition to the *13th Mediterranean Morphology Meeting*, various parts of this research have been presented over the years in the conferences *Mapping Balkan Routes 2: Literary, Cultural and Linguistic Landscapes*, held at the University of Belgrade in October, 2018 (under the title “Roli i sinkretizmit në sistemin rasor të shqipes”, eventually published in Božović 2021) and *Belgrade Linguistics Days (BeLiDa)*, a conference to mark the 30th anniversary of the Department of General Linguistics at the University of Belgrade in December, 2021 (under the title “What Is Syncretism Good for? A Case Study”). I wish to thank the audiences and the organizers of each of these events for their valuable feedback to this research, as it was gradually evolving into its present form. All shortcomings remain my own.

References

- Agalliu, Fatmir. 2002. Emri [The noun]. In Shaban Demiraj (ed.), *Gramatika e gjuhës shqipe, Vëllimi 1: Morfologjia* [Grammar of the Albanian language, Volume 1: Morphology], 81-152. Tiranë: Akademia e Shkencave e Shqipërisë.
- Aronoff, Mark. 1994. *Morphology by itself: Stems and inflectional classes*. Cambridge, MA: MIT Press.
- Aronoff, Mark. 2016. Unnatural kinds. In Ana R. Luís & Ricardo Bermúdez-Otero (eds.), *The morpheme debate*, 11-32. Oxford: Oxford University Press.
- Backus Borshi, Orkida. 2015. *Formal expression of definiteness in Albanian: A description based on comparison with English*. Prague: Charles University PhD thesis.

⁸ As such, Newmark's analysis represents an important early alternative to the Jakobsonian feature-based decompositional approach to case syncretism, which has had a significant influence on later researchers. Sadly, this paper has remained practically unknown.

- Baerman, Matthew, Dunstan Brown & Greville G. Corbett. 2005. *The syntax-morphology interface: A study of syncretism*. Cambridge: Cambridge University Press.
- Baerman, Matthew. 2004. Directionality and (un)natural classes in syncretism. *Language* 80(4). 807-827.
- Booij, Geert. 1994. Against split morphology. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 1993*, 27-49. Dordrecht: Kluwer.
- Booij, Geert. 1996. Inherent versus contextual inflection and the split morphology hypothesis. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 1995*, 1-16. Dordrecht: Kluwer.
- Bozhoviq, Gjorgje. 2021. From Number Cycle to Split Plurality and Back in Albanian. *Balcania et Slavia* 1(2). 157-180.
- Božović, Dorđe. 2021. Roli i sinkretizmit në sistemin rasor të shqipes [The role of syncretism in the Albanian case system]. In Mirjana Marinković & Merima Krijezi (eds.), *Mapping Balkan routes 2: Literary, cultural and linguistic landscapes*, 59-68. Belgrade: Faculty of Philology.
- Buchholz, Oda & Wilfried Fiedler. 1987. *Albanische Grammatik*. Leipzig: Verlag Enzyklopädie.
- Caha, Pavel. 2016. GEN.SG = NOM.PL: A mystery solved? *Linguistica Brunensia* 62. 25-40.
- Caha, Pavel. 2019. Syncretism in morphology. In Mark Aronoff (ed.), *Oxford Research Encyclopedia of Linguistics*, 1-28. Oxford: Oxford University Press.
- Corbett, Greville G. 2015. Morphosyntactic complexity: A typology of lexical splits. *Language* 91(1). 145-193.
- Corbett, Greville G. 2016. Morphomic splits. In Ana R. Luís & Ricardo Bermúdez-Otero (eds.), *The morphome debate*, 64-88. Oxford: Oxford University Press.
- Çerpja, Adelina. 2017. Çintegrimi i asnjansit dhe mbetje të tij në të folme të shqipes [Disintegration of the neuter and its remnants in Albanian dialects]. *Seminari Ndërkombëtar për Gjuhën, Letërsinë dhe Kulturën Shqiptare* 35(1). 145-156.
- Herce, Borja. 2020a. On morphemes and morphomes: Exploring the distinction. *Word Structure* 13(1). 45-68.
- Herce, Borja. 2020b. *A typological approach to the morphome*. Guildford: University of Surrey PhD thesis.
- Jakobson, Roman. 1962. Beitrag zur allgemeinen Kasuslehre: Gesamtbedeutungen der russischen Kasus. In *Selected writings, Volume 2*, 23-71. The Hague: Mouton.
- Jakobson, Roman. 1984. Morphological observations on Slavic declension (the structure of Russian case forms). In Linda R. Waugh & Morris Halle (eds.), *Russian and Slavic Grammar: Studies 1931-1981*, 105-133. Berlin: Walter de Gruyter.
- Luís, Ana R. & Ricardo Bermúdez-Otero (eds.). 2016. *The morphome debate*. Oxford: Oxford University Press.
- Maiden, Martin. 2005. Morphological autonomy and diachrony. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 2004*, 127-175. Dordrecht: Springer.
- Maiden, Martin. 2018. *The Romance verb: Morphomic structure and diachrony*. Oxford: Oxford University Press.
- Maiden, Martin. 2021. The morphome. *Annual Review of Linguistics* 7. 89-108.
- Newmark, Leonard, Philip Hubbard & Peter Prifti. 1982. *Standard Albanian: A Reference Grammar for Students*. Stanford: Stanford University Press.
- Newmark, Leonard. 1962. An Albanian case system. *Lingua* 11. 313-321.
- O'Neill, Paul. 2011a. *The Ibero-Romance verb: Allomorphy and the notion of the morphome*. Oxford: University of Oxford PhD thesis.
- O'Neill, Paul. 2011b. The notion of the morphome. In Martin Maiden, John Charles Smith, Maria Goldbach & Marc-Olivier Hinzelin (eds.), *Morphological autonomy: Perspectives from Romance inflectional morphology*, 70-94. Oxford: Oxford University Press.
- O'Neill, Paul. 2013. The morphome and morphosyntactic/semantic features. In Silvio Cruschina, Martin Maiden & John Charles Smith (eds.), *The boundaries of pure morphology: Diachronic and synchronic perspectives*, 220-246. Oxford: Oxford University Press.
- Përnaska, Remzi. 2003. A ka rasë gjinore shqipja? [Does Albanian have a genitive case?] *Seminari Ndërkombëtar për Gjuhën, Letërsinë dhe Kulturën Shqiptare* 21(2). 107-115.
- Round, Erich. 2015. Rhizomorphemes, meromorphemes and metamorphemes. In Matthew Baerman, Dunstan Brown & Greville G. Corbett (eds.), *Understanding and measuring morphological complexity*, 29-52. Oxford: Oxford University Press.
- Smith, John Charles. 2013. The morphome as a gradient phenomenon: Evidence from Romance. In Silvio Cruschina, Martin Maiden & John Charles Smith (eds.), *The boundaries of pure morphology: Diachronic and synchronic perspectives*, 247-261. Oxford: Oxford University Press.
- Topalli, Kolec. 2009. *Sistemi rasor i emrave të shqipes [The system of cases in Albanian nouns]*. Tiranë: Akademia e Shkencave e Shqipërisë.
- Williams, Edwin. 1994. Remarks on lexical knowledge. *Lingua* 92. 7-34.

Locative forms in Nakh-Daghestanian as an example of a transcategorial paradigm

Marina Chumakina
Surrey Morphology Group
M.Chumakina@surrey.ac.uk

The definition of a word class relies on the notion that different word classes have different paradigms: “canonically, lexemes in different syntactic categories exhibit different morphology,...inflect for different morphosyntactic property sets, and ... have different exponents” (Stump 2015: 229). One possible deviation from this is **transcategorial polyfunctionality**: cases in which “distinct but related content is systematically expressed by the same morphology in different syntactic categories” (Stump 2015: 230).

Word class division in the languages of Nakh-Daghestanian (North-East Caucasian) family in general is rather straightforward: nouns, pronouns, verbs, adjectives (for those languages that have this class) and adverbs have distinct paradigms with the familiar mixed categories deviations: verbs normally have sub-paradigms of participles and verbal nouns which employ nominal paradigms.

However, there is one area in nominal morphology which can be classified as instantiating categorial polyfunctionality: the situation where the same content is expressed by the same morphology in distinct word classes. In many Nakh-Daghestanian languages, both nouns and adverbs employ the same case endings. Examples (1) to (3) from a Lezgian language Archi can serve as a starting point.

- (1) **tus:əl-l-a-k¹** sak:u-qi
bag(III)-SG.OBL-IN-LAT 1PL.look-FUT
‘We will look inside the bag.’
- (2) χit:a **kana-ki** <w>di-muχur eχni-li oq^ʕa-li
then there-LAT <I.SG>be.PST-when [IV.SG]forget.PFV-CVB [IV.SG]leave.PFV-EVID
‘Then, when he was there, he forgot.’

In (1) the noun *tus:əllak* ‘into the bag’ has a locative ending *-k*. In (2) the same ending attaches to the adverb *kana* ‘there’. Note that when used with the adverb, the ending loses its meaning, and the form *kanak* does not mean an expected ‘towards there’. Example (3) from an Archi text collected in 2006 contains two locative nouns and two locative adverbs:

- (3) i<w>χ:u-li i<w>di-li iši-š te:n-ši uq^ʕa-t:u
<I.SG>remain.PFV-CVB <I.SG>be.PST-EVID here-EL there-ALL I.SG.come.PFV-ATTR.I.SG
ħaž-li-t:i-k rume-k-l-a-š os haman-nu i<w>χ:u-li
Hajj(IV)-SG.OBL-SUP-LAT rume-k-SG.OBL-IN-EL one Lak-ATTR.I.SG <I.SG>remain.PFV-EVID
‘There, in Mecca remained a Lak (person) from Rume-k, who went there to do Hajj.’

The adverbs *išiš* ‘from here’ and *te:nši* ‘to there’ employ the relative ending *-š* and the allative ending *-ši* respectively (note that this time both locative endings retain their

¹ I use IPA sign for length (:) to denote fortis consonants.

semantics). The nouns *ħaʒlit:ik* ‘to (do) Hajj’ and *ʁumeklaʒ* ‘from Rumeĸ’ contain the locative ending *-k* and the relative ending *-ʒ* respectively.

These examples reflect a situation typical for a Nakh-Daghestanian language: the paradigm of locative cases straddles the otherwise clearly marked border between nouns and adverbs. To my knowledge, this situation has never been the focus of a theoretical discussion despite the fact that the locative paradigms in Nakh-Daghestanian languages attracted the attention of linguists before (Bokarev 1954, Creissels 2009, Daniel & Ganenkov 2009). The purpose of this paper is to describe the issue and define the questions that needed answers providing the road map for more detailed investigation.

1. Noun paradigm structure in Nakh-Daghestanian languages

Nouns in Nakh-Daghestanian languages are famous for having large and complex paradigms. Two features distinguish Nakh-Daghestanian nominal systems: the opposition between direct and oblique cases, and the division into non-spatial and spatial subparadigms.

The first division runs through both non-spatial and spatial subparadigms. It opposes the unmarked (at least in the singular) absolutive case to all other cases, which are produced from the oblique stems. The oblique stems are often homophonous with the form of the ergative case. Example (4) shows a paradigm of non-spatial cases of a noun from the Lezĸic language Archi. We can see that the absolutive case is opposed to the form of the ergative which serves as a base for all the other case forms. In the singular this opposition is irregular (*baʿkʹ* vs *beʿkʹiri*) whereas in the plural it is expressed by a regular suffix *-čej*.

(4) Non-locative paradigm of *baʿkʹ* ‘ram’ (Archi, Lezĸic)

	SG	PL
ABSOLUTIVE	baʿkʹ	baʿkʹ-ur
ERGATIVE	beʿkʹiri	baʿkʹ-ur-čej ²
GENITIVE	beʿkʹiri-n	baʿkʹ-ur-če-n
DATIVE	beʿkʹiri-s	baʿkʹ-ur-če-s
COMITATIVE	beʿkʹiri-ł:u	baʿkʹ-ur-če-ł:u
SIMILATIVE	beʿkʹiri-qʹdi	baʿkʹ-ur-če-qʹdi
CAUSAL	beʿkʹiri-š:i	baʿkʹ-ur-če-š:i
COMPARATIVE	beʿkʹiri-ħur	baʿkʹ-ur-če-ħur
PARTITIVE	beʿkʹiri-qʹiʒ	baʿkʹ-ur-če-qʹiʒ
SUBSTITUTIVE	beʿkʹiri-klʹena	baʿkʹ-ur-če-klʹena

Large part of a Nakh-Daghestanian noun paradigm is taken up by the locative subparadigm; the forms in the locative subparadigm are based on the same oblique stem as the non-locative forms, but involve the addition of (at least) two elements: a localization suffix and another element, which in some languages (such as Archi) functions as an ending and in some (such as Dargwa) – as a suffix (i.e. can attach further morphological material). The term for this element varies between linguistic traditions within the family, thus, the Archi element is called ‘a directional ending’, whereas Dargwa descriptions call it ‘category of orientation’.

² Here, I break away from the tradition to spell the affix of the plural Ergative case as *-čej* (see, for example, Kibrik 1977) and choose the spelling *-čej* as it reflects the actual pronunciation better and makes clear that the form of the ergative case serves as the oblique stem (the final *j* disappears in the case forms as a result of a phonological rule).

The latter term is easier to use across the family as in several languages there is also a category of directive, which makes the term ‘directional ending’ confusing.

Example (5) from a spontaneous Archi text illustrates usage of locatives: the form *bošormirak* ‘to the husband’ of the noun *bošor* ‘man, husband’ contains a regular suffix of the oblique singular stem *-mi*, a suffix of a CONT(act) localization *-ra* and a lative case ending *-k*:

- (5) tu-w bošor-mi-ra-k kaḁər t’alau-na
 that-I.SG man(I)-SG.OBL-CONT-LAT letter(III)[SG.ABS] <III.SG>send.PFV-CVB
 ‘By sending a letter to this husband (we’ll bring him here)...’

Archi locative subparadigm involves five forms of localization and six forms of orientation; the localizations (LOC) distinguish contact (CONT), inside hollow space (IN), inside filled space (INTER), under (SUB) and on (SUPER) surfaces. Orientation distinguishes ESSive (being in the LOC), ELative (moving from LOC), LATive (moving towards LOC), ALLative (moving to the area of LOC), TERMinative (moving to LOC and no further), and TRANSlative (moving through LOC):

(6) Archi locative affixes

localization		orientation	
CONT	-r-	ESSIVE	ZERO
IN	-a-	ELATIVE	-š
INTER	-q ^c -	LATIVE	-k
SUB	-kl’i-	ALLATIVE	-ši
SUPER	-t:i-	TERMINATIVE	-kəna
		TRANSLATIVE	-χut

Archi represents an average Nakh-Daghestanian locative paradigm. Both larger and smaller paradigms (involving just these two features, localization and orientation) are possible. Thus, Khwarshi, a language belonging to the Tsezic group, has the same number of orientation values as Archi but six rather than five localizations:

(7) Locative suffixes in Khwarshi (Tsezic)

	ESSIVE	DIRECTIVE	VERSATIVE	ABLATIVE	TRANSLATIVE	TERMINATIVE
AD	-ho	-ho-l	-ho-ḁol	-ho-žo	-ho-jža	-ho-q’a
CONT	-l	-l-el	-l-ḁol	-l-žo	-l-ejža	-l-q’a
IN	-ma	-ma-l	-ma-ḁol	-ma-žo	-ma-jža	-ma-q’a
POSS	-qo-	-qo-l	-qo-ḁol	-qo-žo	-qo-jža	-qo-q’a
SUB	-λ	-λ-el	-λ-ḁol	-λ-žo	-λ-ejža	-λ-q’a
SUPER	-λ’o	-λ’o-l	-λ’o-ḁol	-λ’o-žo	-λ’o-jža	-λ’o-q’a

Khalilova, Testelets ms.

Localization and orientation are not the only possible locative categories. Tsezic languages add a third feature, that of the proximity to the speaker (van den Berg 1995, Testelets 1980/2019, Radkevich 2008). As the following example from Bezhta demonstrates, this category can be optional. In (8b) it is realised by the suffix *-da* which goes between localization suffix and the orientation:

- (8) a. do roso-ka-s λ'alo ježeč
 1SG.ERG wall-AD-EL stone carry.PRS
 'I am carrying the stone (away) from the wall.'
- b. do roso-ka-da-s λ'alo ježeč
 1SG.ERG wall-AD-APPROX-EL stone carry.PRS
 'I am carrying the stone (away) from somewhere near the wall.'
- (Testelet 1980/2019 via Lyutikova 2022)

Another optional category, the directive, is registered in Avar (Bokarev 1954), Dargwa (Sumbatova 2003, Lander 2011) and Tabassaran (Khanmagomedov 1958). As (9) from Tanty Dargwa shows, this is an optional category realised by the suffix which attaches to the orientation suffix. This category normally involves direction up or down and sometimes also include meanings hither and thither. In (9) the directive expresses the meaning 'down':

- (9) čet:i-d-at-ur q'uš-me-ra qa'b-li-ja-r-kale
 put-NPL-LV:PF-PRET foot-PL(ABS)-ADD neck-OBL-SUPER-EL-DOWN
 '(Literally:) And he put his feet from the top of the (other's) neck down.'
- (Lander 2011: 2)

Thus, the locative paradigms of the noun in Nakh-Daghestanian languages have common structural properties: the locative forms are based on the same oblique stems as non-locative forms and consist of at least two elements: localization and orientation. The actual values of these features and the usages of locative forms (such as coding the verbal arguments) vary across the languages but this is not the focus of this paper; here, I concentrate on formal elements of locative subparadigm shared by nouns and other parts of speech, namely, adverbs and postpositions.

2. Locative paradigms of adverbs

Adverbs in Nakh-Daghestanian languages present a typologically familiar picture: it is a heterogeneous class encompassing words with different etymological sources: some adverbs clearly derive from case forms of nouns, some originate in converbs and some are non-derived. Mostly, adverbs do not inflect, although in every language there are some adverbs which allow inflection for directional cases and a (much smaller) number of adverbs which have agreeing forms. I am interested in the former type of adverbial inflection here.

If we take Archi as the first example, there are two classes of adverbs: locative and non-locative. Example (10) presents examples from both classes:

- (10) Two adverbial classes in Archi

locative adverbs	non-locative adverbs
<i>jašul</i> 'inside'	<i>jasqi</i> 'today'
<i>jak</i> 'to.inside'	<i>kelaw</i> 'than'
<i>q'on</i> 'between'	<i>χit:a</i> 'then'
<i>l'arak</i> 'under'	<i>jonsaw</i> 'again'
<i>emik</i> 'there'	<i>nessen</i> 'now'
<i>harak</i> 'in front'	<i>o:k'ur</i> 'slowly'

Only locative adverbs can inflect for location, but not every adverb with locative semantics

does so: the grammar of Archi gives examples of inflecting adverbs but never states that the list is exhaustive. Adverbial locative paradigm is smaller than that of a noun: nouns have six values for orientation (6) whereas adverbs have four:

(11)	Inflection of adverbs in Archi			
		‘there’	‘in front’	‘down there’
	ELATIVE	emi-š	hara-š	kl’ara-š
	LATIVE	emi-k	hara-k	kl’ara-k
	ALLATIVE	emi-ši	hara-ši	kl’ara:-ši
	TRANSLATIVE	emi-χut	hara-χut	kl’ara-χut

Compared to noun paradigm, the adverbs lack the essive and the terminative case. The orientation cases employed in both nominal and adverbial paradigms have the identical realizations. While the form is identical, the meanings are not: the form of the lative in the adverbial paradigm does not denote the meaning ‘towards’ but rather means ‘be somewhere’, so the lative case in adverbs functions as the essive case in nouns. Compare the adverb *emik* ‘there’ in (12) which does not mean ‘towards there’ and the noun *dux:ʰat:ak* ‘towards the mill’ in (13):

(12)	emi-k	ħurmat	q’imat	au-li
	there-LAT	respect(III)[SG.ABS]	esteem(III)[SG.ABS]	<III.SG>do.PFV-EVID
	‘...and there they were shown all the respect and esteem’ (that was due to them)			

(13)	q ^w a-li	dux:ʰat:-a-k	tuw
	come.I.SG.PFV-EVID	mill(IV).OBL.SG-IN-LAT	he
	‘He came to the mill’		

Since the lative in adverbial paradigm has the essive meaning, the form of the allative expresses the meaning ‘towards’; in the nominal paradigm the lative and allative have, according to the grammar, the meanings ‘towards’ and ‘towards the area of’, although if we look in the texts, both of cases tend to mean ‘towards’ but are used with different lexical items, and the allative is most frequently used to code the speech addressee.

From a purely morphological point of view, the adverbial inflection for locative cases seems less regular than the nominal inflection: two adverbs are registered to be defective and two are overabundant. The defective ones are shown in (14)³.

(14) Locative paradigms of the Archi adverbs *imik* ‘there’ and *jak* ‘inside’

ELATIVE	imi-š	ja-š
LATIVE	imi-k	ja-k
ALLATIVE	no form	no form
TRANSLATIVE	imi-χut	no form

For two adverbs an overabundant paradigm was registered: the adverbs *kana* ‘there’ and *jat* ‘above’ have the essive form that the other adverbs lack:

³ It has to be noted that we cannot draw the parallel with nouns with the absolute certainty as no systematic study has been done on the nominal lexicon to check whether every noun has the full locative paradigm; such study is almost impossible on a language with only a small corpus, and one can never be sure that the apparent lack of a form (arising in a situation where the linguist offers a form to the speaker) is not a by-product of the fact that the context was not set correctly.

(15) Locative paradigms of the Archi adverbs *kana* ‘there’ and *jat* ‘above’

ESSIVE	<i>kana</i>	<i>jat</i>
ELATIVE	<i>kana-š</i>	<i>jat:i-š</i>
LATIVE	<i>kana-k</i>	<i>jat:i-k</i>
ALLATIVE	<i>kana:-ši</i>	<i>jat:i-ši</i>
TRANSLATIVE	<i>kana-χut</i>	<i>jat:i-χut</i>

However, the essive form of *kana* does not have the expected meaning of ‘there’; rather, it means ‘look!’; in (16) the semantic contrast with *išik* ‘here’ used in the same sentence highlights this:

- (16) *kana, kana, zon išk w-i bo-li*
 there there 1SG.ABS here I.SG-be say.PFV-EVID
 ‘Look, look, I am here, – he said.’

- (17) *kana, bo-li zon wiš lo bo-li*
 there say.PFV-EVID 1SG.ABS your child.SG.ABS say.PFV-EVID
 ‘Look, she said, I am your daughter, — she said.’

While in Archi the locative paradigm for adverbs is smaller than that of the nouns, the adverbs and nouns in Khwarshi have exactly the same set of locative endings:

(18) Locative paradigm for adverbs in Khwarshi

	ESSIVE	DIRECTIVE	VERSATIVE	ABLATIVE	TRANSLATIVE	TERMINATIVE
‘there’	<i>iⁿgo</i>	<i>iⁿgo-l</i>	<i>iⁿgo-koł</i>	<i>iⁿgo-žo</i>	<i>iⁿgo-jža</i>	<i>iⁿgo-q’a</i>
‘here’	<i>idi</i>	<i>idi-l</i>	<i>idi-koł</i>	<i>idi-žo</i>	<i>idi-jža</i>	<i>idi-q’a</i>

(Khalilova, Testelets, ms)

Unlike adverbs in Archi, the adverbs in Khwarshi can attach to both the localization affix and the orientation one:

- (19) *žid-a łona biton-no l-eča-na, iⁿgo-ho-l l-ez-na...*
 they-GEN1 three can-ADDIV-be.CVB.PFV there-AD-LAT IV-take-CVB.PFV
 ‘They had three cans, they took (them) there...’ (Fox, Bear and Wolf, 2)

(Ljutikova 2022)

Bagwalal, a language from Andic group of Nakh-Daghestanian family, has a locative subparadigm with seven localizations but only four orientations:

(20) Noun locative paradigm in Bagwalal

localization		orientation	
AD	-x-	ESSIVE	
CONT	-č'-	ELATIVE	-s:
IN	-ini-, -ni-	LATIVE	-a
INTER	-hi-	TRANSLATIVE	-s:ini
LOCPOSS	-la-		
SUB	-kl'i-		
SUPER	-la-, -lla-		

(Kibrik et al. 2001: 141)

There are adverbs in Bagwalal which inflect for orientation but, similar to what we saw in Archi, the adverbial locative paradigm is smaller than the locative sub-paradigm of the noun. In Bagwalal only two values of the orientation feature are used, the essive and the elative. Example (21) shows an inflecting adverb form Bagwalal.

(21) Inflection of the adverb č'ihī 'above'

ESSIVE č'ihī
ELATIVE č'ihī-s:

These examples drawn from three languages of different branches of the family show us that there is considerable variation in the locative paradigms and in the number of transcategorial elements, i.e. elements shared between adverbial and nominal paradigm, but at the moment we do not have enough data on the adverbial inflection across the family to make any significant conclusions.

3. Locative paradigms of postpositions

If our data on locative inflection of adverbs is sketchy, we know even less of the locative inflection of the postpositions. But the glimpses we get from the descriptions of individual languages are interesting enough to prompt further studies.

The difficulty to distinguish between adverbs and postpositions is a known issue in Nakh-Daghestanian linguistics. Every language in the family has postpositions, i.e. function words which head PPs and select a nominal complement in certain case. However, almost every such word can also be used as an adverb, i.e. without the complement in the initial as well as final position in the clause. Because of this fact, some grammatical descriptions do not distinguish two classes but say that there is a class of adverbs (a lexical class much larger than that of the postpositions) which includes a sub-class of adverbs-postpositions. However, at least one grammar, that of Archi, makes a point that when it comes to taking locative morphology, the postpositions demonstrate some specific properties: while the adverbs take four orientations, the postpositions can only take three: elative, lative and translative. Example (22) shows which parts of locative paradigm is shared between nouns, adverbs and postpositions in Archi. The noun is shown in the form of SUPER localization as this localization demonstrates the least amount of idiosyncrasies when combining with various forms of orientation.

(22) Locative paradigm sharing in Archi

	NOUN ‘ram’	ADVERB ‘under’	POSTPOSITION ‘under X’
ESSIVE	be ^s k’iri-t		
ELATIVE	be ^s k’iri-t:i-š	kɫ’ara-š	kɫ’ara-š
LATIVE	be ^s k’iri-t:i-k	kɫ’ara-k	kɫ’ara-k
ALLATIVE	be ^s k’iri-t:i-ši	kɫ’ara:-ši	
TERMINATIVE	be ^s k’iri-t:i-kəna		
TRANSLATIVE	be ^s k’iri-t:i-χut	kɫ’ara-χut	kɫ’ara-χut

This is a possible test to distinguish adverbs and postpositions, but more data is needed both for Archi and other languages. Thus, we do not even know how the locative forms of postpositions are used in Archi; there are no examples in the texts or in the grammatical descriptions.

Bagwalal also has some inflecting postpositions; like adverbs, they take two values of orientation, but the values are different from those taken by the adverbs.

(23) Inflection of the postposition *la* ‘above X’

ELATIVE	-la-s:
LATIVE	-la-a

Because of the difference in the values of the orientation taken by different parts of speech, the shared paradigm for Bagwalal has a different configuration from that of Archi; only one cell is shared across all three word classes:

(24) Locative paradigm sharing in Bagwalal

	NOUN ‘ram’	ADVERB ‘above’	POSTPOSITION ‘above X’
ESSIVE	miq’a-la	č’ihi	
ELATIVE	miq’a-la-s:	č’ihi-s:	la-s:
LATIVE	miq’a-la-a		la-a
TRANSLATIVE	miq’a-la-s:ini		

The grammatical description of Bagwalal gives examples of the inflected postposition usage: if the locative form of the noun is governed by a postposition, the orientation ending attaches to the postposition and not to the noun:

- (25) *istolla č’ihi* ‘on the table’ — *istolla č’ihi-s:* / **istolla-s: č’ihi-s:* ‘from the table’
(Kibrik et al. 2001: 144)

4. Conclusions

The locative forms in Nakh-Daghestanian family distort the otherwise canonical division of the lexicon into lexical classes. In general, the languages of the family have easily distinguishable lexical classes, each with its own set of morphosyntactic features. The locative paradigms present a clear case of paradigm with shared forms, rather than an instance of ‘borrowing’ forms (in contrast with, for example, participles which make use of nominal case endings when used as headless attributives). In the case of locative forms of nouns, adverbs and postpositions, it is not clear which lexical class got the locative endings ‘originally’ and which only make use of them as a result of some sort of transposition. Very little is known of

the diachrony of these forms and the fact that nouns sometimes have larger locative paradigms cannot, I think, be viewed as an indication of the diachronic path for these forms.

To the best of my knowledge, there has been only one theoretical account for this situation: Lander (2011) proposes to consider locative forms in Dargwa as a specific lexical class. He believes that the appearance of the localization marker on a nominal stem derives a member of a special word class – locatives – with its own syntactic distribution and morphological properties. Besides locative forms of nouns, this class also includes locative adverbs/postpositions and some toponyms. This class has specific syntactic and morphological properties: all locatives normally appear as adjuncts and they all have a special inflectional category – orientation.

Lander (2011) also proposes to consider the production of locative forms to be an instance of incorporation rather than suffixation: locative forms result from incorporation of nominal stems into locative adverbs/postpositions. Like incorporation in many other languages, the formation of locative forms is quite regular and productive and to a large extent lexically determined. Finally, an incorporating element determines the syntactic category of the whole.

At the moment, it is unclear to me whether this analysis can scale up to account for Nakh-Daghestanian languages in general: as we have seen, while it can potentially work for the situations where the adverbs, nouns and postpositions have the same set of orientation values (as we have seen in Tsezic and as it is in Dargwa as well), the instances like Archi and Bagwalal, where the nouns, adverbs and postpositions do not share the whole of the locative paradigm but just some cells, seem to be more challenging.

At the moment, there is no systematic description of the morphosyntactic behaviour of the locative forms across Nakh-Daghestanian languages and therefore I will end with a set of questions for the future:

- Which word classes can participate in the paradigm sharing?
- Are there predictable lexical / semantic classes participating in paradigm sharing?
- How much variation is there in the size of shared paradigm?
- What are the diachronic path(s) resulting in shared locative paradigms?

References

- Baerman, Matthew, Irina Monich & Tatiana Reid. 2019. Nominal inflection classes in verbal paradigms. *Morphology* 29. 317–335.
- van den Berg, Helma. 1995. Grammar of Hunzib. Lincom Europa.
- Bokarev, E.A. 1954. O kategorii padeža (primeritel'no k dagestanskim jazykam.) [On case feature (in relation to Daghestanian languages)]. *Voprosy Jazykoznanija* 1.
- Creissels, Denis. 2009. Spatial cases. In Andrej Malchukov & Andrew Spencer (eds.), *The Oxford handbook of case*, 609–625. Oxford: Oxford University Press.
- Daniel, Michael & Dmitri Ganenkov. 2009. Case marking in Daghestanian. In Andrej Malchukov & Andrew Spencer (eds.), *The Oxford handbook of case*. Oxford: Oxford University Press.
- Kibrik, A. E. 1977. *Opyt strukturnogo opisanija arčinskogo jazyka: II: Taksonomičeskaja grammatika. (Publikacii otdelenija strukturnoj i prikladnoj lingvistiki, 11)*. Moscow: Izdatel'stvo Moskovskogo universiteta.
- Kibrik, A.E. (ed.) 2001. *Bagvalinrskij Jazyk. Grammatika. Teksty. Slovare*. Moskva.
- Khalilova, Zaira & Yakov Testelets (ms.). Khwarshi. To be published in: *The Caucasian languages. An international handbook*. De Gruyter Mouton.
- Khanmagomedov, B. G. 1958. *Sistema mestnyx padežej v tabasaranskom jazyke* [Locative case system in Tabassaran]. Makhachkala: Dagučpedgiz.
- Lander, Yury. 2011. Tanti Dargwa locative forms: Ordering the spatial categories. Handout of the talk given at *3rd Vienna Workshop on affix order: Advances in affix order research*, 15 January 2011.
- Ljutikova, E.A. 2022. Lokalizacija i aproximacija v xvaršinskem jazyke [Localization and approximation in Khwarshi]. Paper presented at the *Smaller languages in big linguistics* conference (МЯБЛ), 20.04.2022.

- Nikolaeva, Irina & Andrew Spencer. 2019. *Mixed categories: The morphosyntax of noun modification*. Cambridge: CUP.
- Radkevich, Nina. 2010. The geometry of local case affixes. University of Connecticut dissertation.
- Sumbatova, Nina. 2003. *Icari Dargwa*. Lincom Europa.
- Stump, Gregory. 2016. *Inflectional paradigms: Content and form at the syntax-morphology interface*. Cambridge University Press.
- Testelet, Yakov. 1980/2019 (ms) Imennye lokativnye formy v dagestanskix jazykax [Nominal locative forms in Daghestanian languages]. Diploma. MGU.

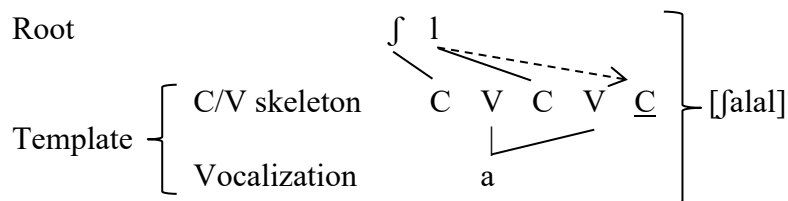
Root reduplication and alignment overcome three challenges to the biradical, OCP-based analysis of Semitic QaTaT stems

Noam Faust
Université Paris 8, CNRS SFL
noam.faust@univ-paris8.fr

1. Introduction

Greenberg (1950) reported on a major asymmetry in Semitic verbs: stems with identical final and penultimate consonants – henceforth QaTaT – are ubiquitous, whereas stems with identical initial and peninitial consonants – henceforth QaQaT – are almost non-existent. In order to explain this asymmetry, McCarthy (1981) famously proposed that the Obligatory Contour Principle (OCP), banning adjacent identical units, holds at the level of the Semitic root. Both \sqrt{QTT} and \sqrt{QQT} roots are illicit. Instead, QaTaT verbs are based on biradical roots \sqrt{QT} matched with a tri-positional template. As illustrated in (1) for the Modern Hebrew verb [ʃalal] ‘he invalidated’, the root is mapped to the template from left to right. When the final C-slot of the template (underlined) is left empty, the closest root consonant spreads to occupy it, in what McCarthy termed “template satisfaction”. Given these premises, a biradical root can never derive a QaQaT verb.

- (1) Bipartite root meets tripartite template => “Template Satisfaction” (McCarthy 1981)



McCarthy’s analysis undeniably constitutes one of the most important events in autosegmental phonology. In the years that passed since its publication, it faced off many challenges.¹ I will concentrate here on three challenges: **(i)** full reduplication of biradicals, **(ii)** the unsatisfied template of vowel-final stems (the QaTaT-QaTa problem from the title), and **(iii)** Amharic templatic intrusion and the purported violation of the OCP in this language (Broselow 1984). The latter two have not been taken up to the best of my knowledge.

Below I propose a solution to challenge **(i)** which follows Marantz (1982), and necessitates the specific alignment principle in (2).

¹ Especially interesting in my opinion is the debate around the existence of QaQaT verbs in Ethiosemitic, for which see Lowenstamm (2010, 2022) (the former also summarizes other arguments for biradical roots). Also interesting are the adaptations of the original analysis into Optimality Theory (Ussishkin 2000, Bat-El 2006), as well as those using representations without a skeletal tier (McCarthy & Prince 1996); but these are irrelevant for the present purpose.

(2) *Misalignment

A non-final root element must not be template-final.

The principle in (2) is shown to underlie the problems posed by challenges (ii) and (iii), such that they are no longer challenges.

In section 2, the three challenges are presented in further details. Section 3 shows how (2) resolves the problems raised by the three challenges.

I end this introduction with a disclaimer. Under the influence of McCarthy & Prince (1996) and Optimality Theory, mainstream work on Semitic templates has seemingly moved away from the skeletal tier, and indeed autosegmental representations. Instead, efforts were concentrated on deriving the form of templates from universal constraints (e.g. Bat-El 2002, 2003; Ussishkin 2005). Nevertheless, as argued in Faust (2015) and Faust & Lampitelli (to appear), templates with arbitrary, lexical shapes have not been argued against convincingly.² In this paper, I maintain a definition of templates using C and V slots; consequently, and for reasons of brevity, work in Optimality Theory is not engaged with directly.

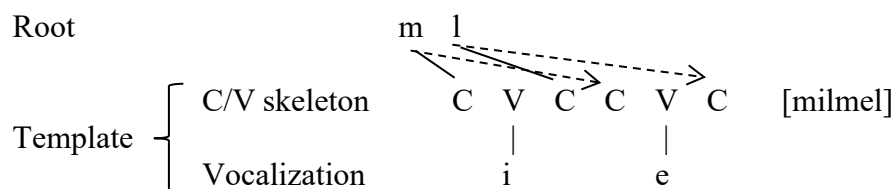
2. Three challenges

2.1. First challenge: fully reduplicated biradicals QaTQaT

The logic behind the mapping in (1) is that spreading is local. \underline{C} is empty, and therefore the closest segment spreads to fill it. The first segment cannot spread to fill \underline{C} , because that can only be achieved through line-crossing, which is disallowed.

All Semitic languages exhibit quadriradical stems, e.g. Modern Hebrew [tikgem] ‘he translated’. Many of these involve fully reduplicated biradicals, like [milmel] ‘mutter’ from the same language, related to [mila] ‘word’. However, assuming a quadri-positional template, left-to-right association and a biradical root \sqrt{ml} , [milmel] can only be derived through line crossing:

(3) Line crossing in left-to-right + spreading account of fully reduplicated biradicals



Within autosegmental phonology, Broselow & McCarthy (1983) propose a solution to this challenge, which I will show below runs into a principled difficulty. See also Bat-El (2006) for an account without autosegmental representations.

2.2. Second challenge: the QaTaT – QaTa problem

Consider the three Modern Hebrew verbs in (4). They are all of the same type known as “qal” or “paʕal”, as attested by the shared vocalization <a,a> in the PST3MSG and the use of the same action noun and passive participle templates QTilLa and QaTuL. (4b) involves identical penultimate and final consonants, and would be derived from a biradical root through template satisfaction. However, (4c) lacks a third consonant. If one wants to argue that the three verbs

² But see the interesting proposal in Golston (1996), where templates are defined by the markedness constraints they *violate*.

share a template, one must explain why the template is not satisfied in (4c) or, in other words, why the third C-slot of (4c) may remain empty.

(4) Three Modern Hebrew verbs

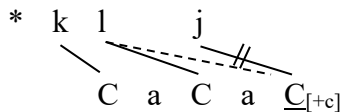
	PST.3MSG	ACTION N	PASS.PRTC.3MSG	
a.	kalat	klita	kalut	‘receive’
b.	kalal	klila	kalul	‘include’
c.	kala	klija	kaluj	‘roast’

Triplets like the one in (4) are found in most, if not all Semitic language. Any Semiticist knows the beginning of the solution to the challenge posed by them: the root of (4c) is not biradical \sqrt{kl} , but triradical \sqrt{klj} . The final /j/ is even apparent in the action noun and passive participle.

However, this view does not immediately answer the question. In the verbal form, the final /j/ is clearly absent from the final position; let us assume that the final C-slot of the verbal template is specified [+c(onsonantal)].³ Why then is the template allowed to remain unsatisfied?

The question is posed in graphic form in (5). Assuming that the root-final /j/ cannot be associated to \underline{C} , the situation is identical to that in (2) above. Why is the position allowed to remain empty, instead of the second radical /l/ satisfying the template as in (2) above, to derive [kalal]?

(5) Template satisfaction wrongly predicts [kalal] for $\sqrt{klj}+CaCaC_{[+c]}$



The existence of QaTaT forms with a satisfied template alongside QaTa forms with an unsatisfied template is what I call the QaTaT – QaTa problem. I am unaware of this challenge having been raised in the past.

2.3. Third challenge: Amharic (Broselow 1984)

Broselow (1984) claims that the analysis of QaTaT verbs as derived from \sqrt{QT} does not hold for the Ethiosemitic language Amharic. In this language, QaTaT verbs are based on OCP-violating \sqrt{QTT} roots.

The argument begins with the comparison of the paradigms in (6) (the data are slightly altered, based on Leslau 1995’s reference grammar). (6a) shows the basic stems of a type A verb with an unremarkable root (three different consonants, always surface-true). (6b) shows that stems with identical final and penultimate consonants adhere to the same templates. (6c) shows a third paradigm, also of type A, which differs from the other two in several respects. Two are crucial: **i.** the PFV, IPFV and JUSS involve one less consonant than (6a,b); and **ii.** the GRD and the INF in (6c) feature the same number of consonants as (6a,b) because an additional [t] (in bold) occurs in the final consonantal position (the L position in the TEMPLATE column).

³ Indeed, no verbal form in Hebrew ever features [j] in this position. But the reason for the non-realization of /j/ cannot be phonological, as [aj] is a legitimate sequence in Modern Hebrew, e.g. [banaj] ‘builder’. See Aïm (2004) for more cases of featural specification for templatic positions.

(6) Three verbal paradigms in Amharic

	TEMPLATE	a. ‘break’	b. ‘like’	<i>but</i>	c. ‘scorch’
PFV3MSG	QäTTäL-ä	säbbär-ä	wäddäd-ä		fädǰǰ-ä
IPFV3MSG	jiQäTL-all	jisäbr-all	jiwädd-all		jifädǰ-all
JUSS3MSG	jiQTäL	jisbär	jiwdäd		jifidǰ
GRND3MSG	QäTL-o	säbr-o	wäddo		fädǰto
INF	mäQTäL	mäsbar	mäwdäd		mäfdǰät

Broselow argues that verbs like (6c) are based on biradical roots ($\sqrt{fdǰ}$ for 6c). The [t] in the GRND and INF of (6c) is then a “default consonant inserted in order to satisfy the tri-positional template.” But if this is so, why is this strategy not used for the seemingly biradical (6b)? Broselow concludes that the verb in (6b) cannot be based on a biradical root \sqrt{wd} . Instead, such verbs are based on OCP-violating \sqrt{QTT} roots (\sqrt{wdd} for 6b). Thus, for Broselow, Amharic roots may violate the OCP.

Three aspects of the data in (6) nevertheless remain unexplained. First, what brings about the difference in the JUSS templates between (6a,b) on the one hand and (6c) on the other? Second, Leslau reports that all verbs of the type in (6c) – of which there are quite a few – involve a second palatalized consonant (with two exceptions, only one of which involves a second palatalizable consonant). Finally, if the non-radical [t] is inserted to fill a templatic position, why is it only used in the GRND and INF?

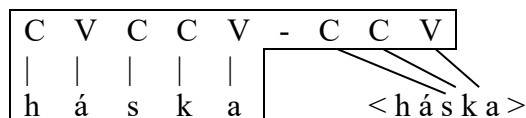
I am unaware of any published (or unpublished) response to the challenge posed by Broselow.

3. Analysis

In a classic paper about reduplication, Marantz (1982) proposed the following analysis. Reduplicants are specified only at the skeletal level, not at the segmental one. In order to satisfy the template of the reduplicant, the segmental material of the base is reduplicated. Depending on the template of the reduplicant, all or only part of the base can reappear in the reduplicant.

This is illustrated by the Dakota example [háska] ‘be tall’, whose reduplicated form is [háska-ska]. The inner frame represents the first step, wherein a skeletally-specified reduplicant /-CCV/ is added to the base. In the second stage (outer frame), the segmental material is reduplicated in its entirety and then used to satisfy the template of the reduplicant. Importantly for the present purpose, the satisfaction of the reduplicant proceeds from right to left.

(7) Suffixed reduplicant and its template satisfaction (Marantz 1982)



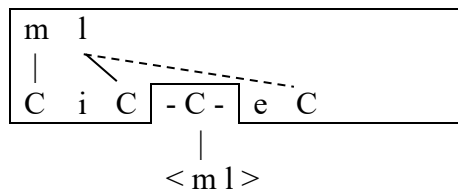
In the overwhelming majority of the cases Marantz surveyed, right-to-left association correlated with the reduplicant appearing to the right of the base. This follows from the generalization proposed in the introduction:

- (8) *Misalignment (repeated from 2)⁴
 A non-final root element is also final in the derived form.

In order to make sure that the final root element is also final in the derived form, association proceeds from right to left.

Using Marantz’s approach, Broselow & McCarthy (1983) account for QaTQaT cases like [milmel] with the notion of “infixed reduplicants”. The first stage (framed) is identical to the left-to-right template satisfaction in (1) above. The second stage (unframed) inserts an infix skeletal slot C. As in (9), the root is reduplicated. It is associated from left-to-right again, and [milmel] is yielded.⁵

- (9) [milmel] is a case of an infix reduplicant C (Broselow & McCarthy 1983)



The authors argue for the validity of this analysis by showing that, in Levantine Arabic, some triradical roots exhibit a surprising reduplication pattern 123 => 1213, e.g. [barad] ‘he shaved’ [barbad] ‘he shaved unevenly’. Such a pattern would work exactly as in (9); the only difference would be the number of radicals.

A crucial point that Broselow & McCarthy seem to miss is that, across Semitic, the 1213 pattern is extremely rare, whereas the biradical 1212 pattern is ubiquitous. Modern Hebrew, for instance, lacks the former altogether, but exhibits many verbs of the latter. It cannot be the case that C infixation applies only for biradicals.

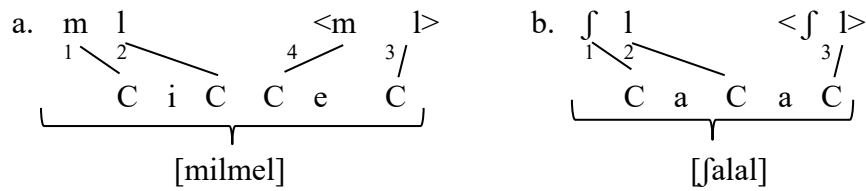
An alternative to this account appears in (10a). The root \sqrt{ml} is matched with the template CiCCeC. Association proceeds from left-to-right. The mismatch between root and template triggers root reduplication as in Marantz’s account. Accordingly, the reduplicated root is associated from right to left (the order of operations is referred to by numbering). Crucially, this approach also applies to biradical roots with tripositional templates, as shown in (10b): the root is reduplicated and associated edge-in. The only difference between (10a) and (10b) is that in the latter, just like in the Dakota case in (7) above, one of the segments of the reduplicated root remains unassociated for lack of a C-slot.⁶

⁴ Alignment considerations do feature prominently in the aforementioned accounts in Optimality Theory; *Misalignment is certainly not an innovation of the present proposal. Having said that, Bat-El (2006) argues, along with Nelson (2003) and in contrast to the present proposal, that right-anchoring is not part of Universal Grammar.

⁵ Henceforth, for graphic reasons, vowels are represented instead of V-slots with associated segments; the tier labels “root, skeleton, vocalization” are also absent.

⁶ Edge-in association has been argued for by Yip (1988) and Buckley (1990). However, the application of the principle in those papers is different from the one proposed here. Verbs like [jalal] (10b) are not treated as reduplicated; instead, first the final radical associates to the final C-slot and then it spreads leftwards. This approach and the one championed here encounter difficulties in accounting for different forms ([milmel] in (10a) poses a challenge to Yip 1988, too), and this is not the place to compare them. Both approaches adhere to *Misalignment.

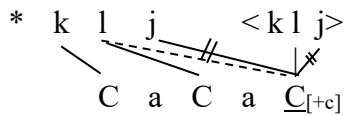
(10) Reduplication followed by edge-in association



This way of regarding reduplication and template satisfaction avoids the line-crossing mentioned in (3) above, does not claim internal infixation only for biradical roots and is in conformity with the typological generalizations in Marantz (1982).

We may consider this challenge overcome. But crucially, overcoming it involved admitting the alignment principle above. In (10b), why is /l/ and not /f/ associated to the final slot? The answer is that such an association would violate the principle of *Misalignment by deriving [falaf] from √fl. Both (10a) and (10b) abide by *Misalignment.

The ban on misalignment also sheds light on the QaTaT – QaTa problem. Recall that QaTa verbs involved an unsatisfied template, which is expected to be satisfied and yield QaTaT. The configuration is given again in (11), with the reduplication and right-to-left association I now claim is general: /j/ cannot attach to C_[+c], the root is reduplicated, but even the reduplicant's /j/ can't associate to C_[+c]. Why doesn't the next consonant of the reduplicant associate to C_[+c]?

 (11) Template satisfaction wrongly predicts [kalal] for √klj+CaCaC_[+c]


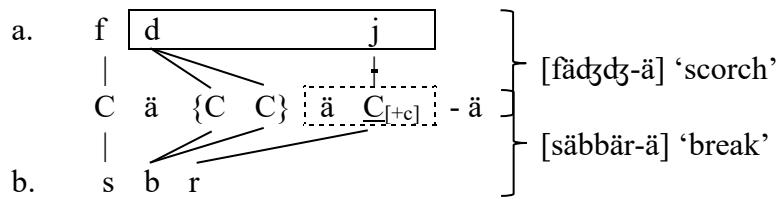
The impossibility of (11) can now be attributed to the violation of *Misalignment which would result from positioning the penultimate radical at the right edge of the template. Since the correct alignment is impossible in this case (because of the specification [+c]), the template may remain unsatisfied.

Returning to the Amharic cases, we may now better understand some of facts, on the basis of a comparison to Modern Hebrew. What if the set in (6) above represented the Amharic instantiation of the QaTaT-QaTa problem? That is, what if seemingly biradical verbs like [fädʒdʒ-ä] were based on *triradical* roots whose final consonant cannot associate to the final templatic slot?

The identity of the missing final radical is already hinted at by the palatal nature of all of the second surface consonants of these verbs. Like in Modern Hebrew, the missing final radical is the palatal /j/. The palatality of this missing consonant ends up on the preceding consonant. In other words, what surfaces as two consonants [f,dʒ], originates in a triconsonantal set /f,d,j/.

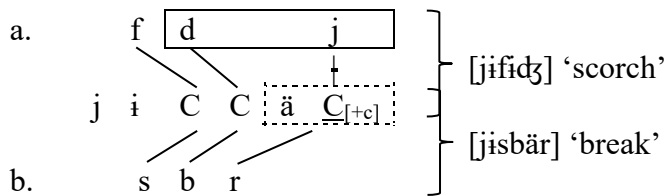
The analysis is made explicit in (12), showing both the similarity to and the difference from Modern Hebrew. The first template examined is that of the perfective, with the prespecified gemination of the second consonant (signaled by {CC}). As in Modern Hebrew, the final C-slot of the template is specified [+c], and so in (12a), the final radical cannot access it (cf. 12b). Unlike in Hebrew, the radical is joined to the preceding consonant and palatalizes it. Also unlike in Hebrew, the non-satisfaction of the template leads to its truncation: both the second /ä/ vowel of the template and the final C-slot are deleted (the sequence deleted is framed in a broken contour). At no point is *Misalignment violated.

(12) /j/-final vs. regular verbs in Amharic – PFV



The possibility for templates to truncate explains a second issue I raised with respect to Broselow’s (1984) analysis, namely the reason that different templates seem to be used in triconsonantal JUSS [jɨsbär] and its allegedly biconsonantal parallel [jɨfidʒ]. As shown in (13), the two forms in fact do share a template; but since the root √fdj cannot satisfy the template fully, its template is truncated. This leads to epenthesis appearing between the last two Cs (Amharic does not tolerate [fdʒ] as a final cluster; in verbs of the same class that result in a licit cluster, no epenthesis occurs):

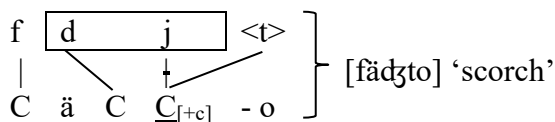
(13) /j/-final vs. regular verbs in Amharic – JUSS



More importantly than the specifics of the analysis of verbs such as ‘scorch’, once their roots are identified as triradical, there is no longer reason to regard verbs like [wäddäd-ä] ‘he liked’ as based on OCP-violating roots. The Semitic-wide analysis of such verbs as based on biradicals can be maintained for Amharic, too. The second consonant of the root is [+c] and therefore can associate to the final slot through reduplication and right to left association, as in the Hebrew case in (4) above.

What of the main issue of Broselow’s paper, namely the insertion of default [t] in the GRND and INF? I argue, with Broselow, that [t] is inserted in order to satisfy the template. Importantly for the present purpose, this template satisfaction strategy *does not violate* *Misalignment: [t] is a non-root consonant, and so the root is not misaligned – all non-final root consonants are also not template-final.

(14) /j/-final verbs in Amharic – GRND



Of course, one no longer expects [t] insertion in paradigms based on true biradical roots, since, in these, reduplication and right-to-left association can satisfy the template. To summarize, [t]-insertion emerges as a template satisfaction strategy which allows the grammar to adhere to the *Misalignment principle.

One question remains: why does [t]-insertion occur in the GRND and INF of /j/-final verbs, but not in the PFV, IMPF and JUSS forms? This question is answered in detail in Faust (to appear, b), based on Faust (to appear, a). In the latter, a parallel case of [t]-intrusion from Modern Hebrew is shown to employ not a default consonant, but the feminine suffix /-t/. Faust (to

appear, b) then argues for the same analysis in Amharic. Following Leslau (1995), it is claimed that the Amharic GRND and INF are morphologically nominal; they are therefore the only bases in the verbal paradigm that can employ an external feminine suffix in order to satisfy the template (since only nouns can carry non-agreement gender suffixes). In other words, the intrusive [t] in Amharic is not a “default consonant” but a feminine suffix. I leave the issue at that, as it is elaborated upon elsewhere.

4. Conclusion

This short paper discussed three challenges to the OCP-based account of QaTaT verbs in Semitic. I argued that these challenges can be overcome using *Misalignment and template satisfaction through reduplication. QaTa and QaTQaT verbs no longer require specific mechanisms, and Amharic does not have OCP-violating \sqrt{QTT} roots. In some cases, an intrusive [t] can be employed in order to satisfy the template without violating *Misalignment.

In the last decades, there has been a debate around the cognitive reality of the Semitic root (see survey in Faust & Hever 2010). It has been repeatedly suggested (e.g. Bat-el 2002, 2003; Laks 2018, 2022; Ussishkin 2000, 2005) that this notion is unnecessary: there is no sense in which [säbbära] is “derived from” or “based on” a unit \sqrt{sbr} . The analysis in this paper adds another argument in favor of the necessity of admitting the root as a morph. The proposed constraint *Misalignment refers directly to this level, and cannot be reformulated without it; specifically, [t]-insertion is a possible solution to misalignment precisely *because* /t/ is not a root element. Time will tell if surface-oriented approaches which do not recognize roots can cover the phenomena discussed in this paper.

References

- Aim, Emmanuel. 2004. Aramaic & Hebrew metathesis. *Israeli Association for Theoretical Linguistics (IATL)* 20. Available online. <http://linguistics.huji.ac.il/IATL/20/Aim.pdf>
- Bat-El, Outi. 2002. Semitic verb structure within a universal perspective. In Joseph Shimron (ed.), *Language processing and acquisition in languages of Semitic, root-based, morphology*, 29–59. Amsterdam: John Benjamins.
- Bat-El, Outi. 2003. The fate of the consonantal root and the binyan in optimality theory. *Recherches Linguistiques de Vincennes* 32. 31–60.
- Bat-El, Outi. 2006. Consonant copying and consonant identity: The segmental and prosodic structure of Hebrew reduplication. *Linguistic Inquiry* 37. 179–210.
- Broselow, Ellen. 1984. Default consonants in Amharic morphology. In Margaret Speas & Richard Sproat (eds.), *MIT Working Papers in Linguistics*. 15–32.
- Broselow, Ellen & John J. McCarthy. 1983. A theory of internal reduplication. *The Linguistic Review* 3. 25–88.
- Buckley, Eugene. 1990. Edge-in association and OCP ‘violations’ in Tigrinya. *West Coast Conference on Formal Linguistics (WCCFL)* 9. 75–90.
- Faust, Noam. 2015. A novel, ncombined approach to Semitic word-formation. *Journal of Semitic Studies* 60(2). 287–316.
- Faust, Noam. To appear, a. Templatic in[t]ruders in Modern Hebrew. In Ronny Meyer & Chris Reintges (eds.), *Issues in templatic morphology*.
- Faust, Noam. To appear, b. Intrusion as template satisfaction and the QaTaT-QaTa problem in Semitic. *Linguistic Inquiry*.
- Faust, Noam & Nicola Lampitelli. To appear. Root and pattern in Semitic – and beyond. In Peter Ackema, Sabrina Bendjaballah, Eulàlia Bonet & Antonio Fábregas (eds.), *The Wiley Blackwell companion to morphology*.
- Faust, Noam & Yaar Hever. 2010. Empirical and theoretical arguments for the discontinuous root in Semitic. *Brill’s Annual of Afroasiatic Languages and Linguistics* 2. 80–118.
- Golston, Chris. 1996. Representation as pure markedness. *Language* 72(4). 713–748.
- Greenberg, Joseph H. 1950. The patterning of root morphemes in Semitic. *Word* 6. 162–181.
- Laks, Lior. 2018. Verb innovation in Hebrew and Palestinian Arabic. *Brill’s Journal of Afroasiatic Languages and Linguistics* 10(2). 238–284.

- Laks, Lior. 2022. Adjective-derived verbs and verb-derived adjectives in Hebrew. *Radical: A Journal of Phonology* 4. 249–291.
- Leslau, Wolf. 1995. *Reference grammar of Amharic*. Weissbaden: Harrassowitz.
- Lowenstamm, Jean. 2010. An introductory note to Noam Agmon's 'Materials and language', with special attention to the issue of biliteral roots. *Brill's Annual of Afroasiatic Languages and Linguistics* 2. 1–22.
- Lowenstamm, Jean. 2022. Oh to understand the impish behaviour and temperament of Ethiopian left-geminators! *Radical: A Journal of Phonology* 4. 453–490.
- Marantz, Alec. 1982. Re Reduplication. *Linguistic Inquiry* 13. 435–482.
- McCarthy, John J. 1981. A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry* 12(3). 373–418.
- McCarthy, John J. & Alan Prince. 1996. Prosodic morphology 1986. *Linguistics Department Faculty Publication Series* 13.
- Nelson, Nicole Alice. 2003. *Asymmetric anchoring*. New Brunswick, NJ: Rutgers University dissertation.
- Ussishkin, Adam. 2000. *The emergence of fixed prosody*. Santa Cruz, CA: University of California dissertation.
- Ussishkin, A. 2005. A fixed prosodic theory of nonconcatenative templatic morphology. *Natural Language and Linguistic Theory* 23. 169–218.
- Yip, Moira. 1988. Template morphology and the direction of association. *Natural Language and Linguistic Theory* 6. 551–577.

The productivity of adverbs and adverbials in Modern Hebrew

Malka Muchnik
Bar Ilan University, Israel
malka.muchnik@biu.ac.il

1. Introduction

In different languages, the definition of adverbs and adverbials is very problematic and controversial. In their research on adverb classes in European languages (German, English, Dutch, French and Italian), Pittner, Elsner & Barteld (2005) state that these classes are very heterogeneous and therefore difficult to define.

Discussing Spanish, Salazar García (2007) reminds us that adverbs are usually considered very complex and heterogeneous, and this is the reason why we encounter enormous difficulties in defining them, both theoretically and descriptively. Since their semantic value and syntactic uses are highly divergent, it is hard to suggest a common definition and a coherent and systematic classification. Also, Maienborn & Schäfer (2011) comment that clear-cut definitions of adverbs and adverbials are difficult to formulate, as we define the word class adverb on the basis of the adverbial syntactic function.

One of the most argued questions among scholars is whether adverbs are an open or closed class. For Talmy (2000), they are a closed category, meaning that there is a limited number of such words and the class cannot be productive. Salazar García (2007) proposes to divide adverbs into two sub-categories. Adverbs of manner would be considered content words, i.e. an open class, while other adverbs, such as those of degree or negation, are function words or grammatical particles, i.e. a closed class.

Some scholars try to answer the question whether adverbs represent a special type of morphology. According to Giegerich (2012), English adverbs are not different from adjectives and have no morphology of their own, but share it with adjectives. Moreover, he claims that the adverb in English is not a lexical category but merely a specific modifier with a function performed by members of the category adjective, associated with contexts other than those traditionally associated with adjectives.

Pounder (2001) shows that, although German and English were historically similar regarding the use of adverbs, they differ from each other. Diepeveen & van de Velde (2010) state that, in contrast to English, which, according to them (and contrary to Giegerich's aforementioned statement), marks the distinction between adjectives and adverbs with an adverbial suffix, Dutch and German allow adjectives to be used adverbially without extra morphology.

Following some of these claims, we may expect that not only the grammatical category, but also individual adverbs, would tend to become unrecognizable as a specific part of speech in a language and be integrated with adjectives. However, it appears that this is not the case for Modern Hebrew.

With respect to the frequency of the different parts of speech in Hebrew, Schwarzwald (2019) points out that adverbs are among the frequent words that are stable, meaning that they do not change or disappear from one period to another. Moreover, she adds that even new processes in the language do not influence them. Similarly, Muchnik (2000) found that

Hebrew adjectives used in slang are formed according to known patterns or common phonological structures.

The purpose of the present study is to examine the presence of adverbs and adverbials in Modern Hebrew according to their morphological formation, and see whether they are stable or have changed from the classical to the modern language. I will try to prove that they are productive, meaning that they not only remain in the language as lexical items, but also expand in known or similar patterns.

2. Adverb formation and productivity

In many European languages, adverbs are clearly distinguished by the suffixes added to adjectives. For instance, we find adverbs ending in *-ly* in English, *-lich* in German, *-ment* in French, and *-mente* in Spanish, Italian or Portuguese. However, these suffixes are not imperative, as we can find suffix-less adverbs like *well*, *super*, *genial*, etc.

In a diachronic study on British English, Tagliamonte & Ito (2002) state that the use of adverbs with zero suffix increased over time, when compared with those with the *-ly* suffix. While this process was even greater in American English, in British English the use of zero-suffix adverbs was considered an informal, colloquial, familiar, and even vulgar style, but nonetheless it did not disappear. They further add that the longitudinal linguistic change regarding the use of zero-suffix adverbs is attributed to social class or education. They show that less educated males used more zero-suffix adverbs, and claim that this is an example of the social and historical development.

When dealing with grammaticalization as an adverbial creator, Killie (2015) shows that the *-ly* suffix has come to be used in a number of contexts and functions where it was not originally used, because in Old English most adverbs did not present any suffix. During the nineteenth and twentieth centuries, the adverbs with *-ly* suffix became extremely productive. The suffix was also attached to present participles in adverbial functions. Most adverbs ending in *-ly* are manner adverbs or intensifiers, such as *perfectly*, *completely*, *totally*, *absolutely*. Due to their increasing productivity, Killie (ibid.) states that the term ‘adverbialization’ seems most appropriate.

In a study on English literature, Killie (2000, 2022) found that the drift from more literate to more oral styles led to an increase of adverbials. The spread of *-ly* suffixes, including the development and diversification of stative adverbs, is bound up with new genres. In addition, many of the adverbs in her corpus had a manner or a manner-like function. Moreover, the process had a snowball effect, meaning that the more *-ly* adverbs in the language, the more such adverbs we are likely to get. She adds that, psychologically, people get so used to adverbs, that they prefer using them instead of adjectives or other alternatives. This trend was attested in popular, non-expository registers.

It is possible that not only in English, but also in other languages, the use of adverbs will increase over time, meaning that we may witness their productivity. Van Marle (1985, 1992) defines productivity as a process by means of which the lexicon of a language can systematically or regularly be extended. He further states that new coined words must have parallel forms in the language. Creative formations typically have special connotations, such as elements of humor, irony or contempt.

3. Hebrew adverbs and adverbials

As stated by Berman (1987), Modern Hebrew is a particularly good case for the analysis of lexical productivity, because the language represents a sort of “diglossia” between the puristic

requirements of prescriptive or official norms compared with the colloquial usage manifested by native speakers of different levels of education.

Nir & Berman (2010) and Bolozky & Berman (2020) maintain that Modern Hebrew adverbs represent an intermediate category between the open class of content words and closed class of function words, and typically lie between the two extremes of lexicon and grammar. Regarding morphological and syntactic aspects, Ravid & Shlesinger (2000) show that Hebrew adverbs are fuzzy and very diverse. They argue that they present an atypical character, as they do not resemble any other content word. All Hebrew verbs and many nouns and adjectives are formed by a stem or a consonantal root and a vocalic pattern and can be inflected, while adverbs do not make extensive productive use of morphological structure, and do not inflect.

Therefore, Ravid & Shlesinger (ibid.) describe Hebrew adverbs as a peripheral lexical category in a language that defines its content words by both derivational and inflectional markers. They emphasize that Modern Hebrew does not really present a productive morphological class of adverbs, despite its synthetic Semitic character.¹ They add that Hebrew nouns, verbs and adjectives can be included in morpho-lexical classes, while the function of adverbials of manner cuts across the lexicon, morphology, semantics, and syntax.

While all Hebrew adjectives can be inflected according to gender and number, most adverbs do not present this possibility. For example, the adjective *tov* 'good' (SING, MASC) can be inflected into *tova* (SING, FEM), *tovim* (PL, MASC) and *tovot* (PL, FEM). The parallel normative adverb would be *heitev* 'well', but it is not regularly used in colloquial language, and the non-inflected adjectival form *tov* is preferred. It is possible then, that this morphological differentiation will prevent adverbs from disappearing in Hebrew.

Note that Modern Hebrew adverbs may derive into adjectives, by adding the suffix *-i*, like in *mamaši* 'real', *axšavi* 'current', *pit'omi* 'sudden', *hinámi* 'gratuitous', and, according to Bolozky (1999), this process is quite productive. In rare cases, adverbs can be used as nouns by adding to them the plural suffix *-im*, such as *etmolim* 'yesterdays' and *émešim* 'last nights' in literary language.

As in other languages, we should distinguish between Hebrew adverbs and adverbials (or adverbial clauses), since adverbs constitute a lexical class, whereas adverbials are a functional and syntactic class, generally formed by a preposition followed by a noun.

Ravid (2020) claims that there is not a productive class of morphologically derived adverbs in Hebrew. Instead, they are expressed by prepositional phrases, zero-derived adjectives in colloquial usage, or inflected feminine suffixes attached to adjectives in very high register or literary style. To the contrary, Kogut (2002) points out that modern languages, among them Hebrew, contain formation patterns that enrich adjectives and adverbs, which contributes to the stylistic diversification.

In what follows, I will try to prove that Modern Hebrew adverbs are an open and productive class. The same is true for adverbials, which are composed of existing content words joined with function particles, mostly prepositions. In both cases, they are productive in recent years, particularly in colloquial language and in Israeli slang.

For this purpose I have used two dictionaries, *Rav-Milim* [Many Words] (Choueka 2010), which is updated online, and *Milon HaSleng HaMakif* [Dictionary of Israeli Slang] (Rosenthal 2005).

¹ Muchnik (2004) shows that the synthetic character of Hebrew is changing into a more analytic way.

4. Adverbs

Many adverbs were found in the present survey, and they will be exemplified here according to their formation categories. In some cases, their form continues patterns already found in Classical Hebrew, such as unmarked monosyllabic adverbs, while in other instances they consist of grammaticalized words or are formed in totally new ways.

4.1. Unmarked adverbs

Basic Hebrew adverbs already found in Classical Hebrew and still used to this day are morphologically unmarked and underived. Here are some examples:

- (1) a. *kan* 'here'
- b. *šam* 'there'
- c. *kax* 'so'
- d. *az* 'then'
- e. *po* 'here'
- f. *'od* 'yet'
- g. *me'od* 'very'
- h. *le'at* 'slowly'
- i. *levad²* 'alone'
- j. *stam³* 'just'

The aforementioned adverbs are peculiar, because they are monosyllabic and are not ruled by any typical pattern, like the combination of a consonantal root with a known vocalic pattern. Non-derived stems like these are also used in some nouns, considered ancient words (Schwarzwald 2001). No new adverbs were found in a similar form.

4.2. Discontinuous versus linear formation

Besides these unmarked adverbs, which are a small minority, Modern Hebrew adverbs are found in two different formation styles, the classical discontinuous form, meaning the combination of consonantal roots with vocalic patterns, and the linear formation attaching prefixes or suffixes to a base or stem (Nir 1993). The productivity of each of them can sometimes distinguish between classical and modern formation styles (Muchnik 2004). In what follows, I will first present adverbs found in discontinuous formation (Section 4.3), and afterwards those in linear formation (Section 4.4).

4.3. Discontinuous formation

Not many adverbs formed by consonantal roots and vowel patterns were found in the present study, and all of them actually represent a secondary use of existing parts of speech, such as nouns, absolute infinitives, adjectives, nominal forms and present participles, as we can see in the next sections.

² In the words *me'od*, *le'at* and *levad*, /e/ represents a shwa and is not counted as a vowel.

³ The word *stam* is also used in humorous language prolonging the vowel *a* and becoming *staaam*, meaning 'just kidding'.

4.3.1. Nominal patterns

Most Hebrew nominal patterns are disyllabic. Some adverbs are formed in the pattern CVCV(C), which is known in many words (Cohen-Gross 1997; Schwarzwald & Cohen-Gross 2000). Here are some examples of adverbs formed in this pattern and used to this day:

- (2) a. *maḥar* ‘tomorrow’
 b. *ḥaval* ‘it is a pity’
 c. *mamaš* ‘really’
 d. *vaday* ‘certainly’

Some disyllabic nouns are used as adverbs in Modern Hebrew, and particularly in slang. For instance:

- (3) a. *‘anak* ‘giant[ly]’
 b. *hamon* ‘multitude = plenty’
 c. *ḥalom* ‘dream[ily]’

In all these cases, the original nouns appear in the Bible, and were later derived into adjectives by adding the suffix *-i*, namely *‘anaki*, *hamoni* and *ḥalomi*. In recent years, the nouns were adopted as adverbs. The form *‘anaki* is actually unnecessary, because the noun *‘anak* is also used as an adjective, similarly to *nora‘i* shown in (7c).

Special disyllabic adverbs were found, where the whole word is repeated, like in the next examples:

- (4) a. *kaxa-kaxa* ‘so-so’
 b. *rega-rega* ‘a moment-a moment’
 c. *para-para* ‘cow [after] cow’
 d. *‘ehad-‘ehad* ‘one [by] one’
 e. *turki-turki* ‘Turk [after] Turk’
 f. *nora-nora* ‘terribly-terribly’

4.3.2. Absolute infinitives

Another form of adverbs, regularly used in Classical Hebrew, is identical to absolute infinitives, which are rarely used nowadays (see Schwarzwald 1989). However, some of these adverbs remain in use, among them:

- (5) a. *harbe* ‘many’
 b. *hayšer* ‘directly’
 c. *heitev* ‘properly’
 d. *harḥek* ‘distantly’
 e. *halox vašov* ‘back and forth’
 f. *halox veḥazor* ‘back and forth’

The idiomatic expressions *halox vašov* and *halox veḥazor* (5e,f) contain two absolute infinitives each, and are used in an adverbial-aspectual sense in literary writing, where the meaning is ‘doing something repeatedly’, but also in colloquial language meaning ‘round trip’ (Saydon 2018). Absolute infinitives were also found by Muchnik (1994) in a very popular gossip section in the 1990’s, although they were not used as adverbs but as verb constructions, like these:

- (6) a. *halox halxa* 'she went'
 b. *šalom šilem* 'he payed'
 c. *baroz hibriza* 'she shirked'⁴

Example (6a) above is still used, mostly in children's literature. The other examples (6b,c) are only typical in the gossip genre or humorous or ironical speaking and writing.

4.3.3. Adjectival form

Adverbs may also present identical forms as basic masculine adjectives, as mentioned by Amir Coffin & Bolozky (2005) and Schwarzwald (2001). For instance:

- (7) a. *yafe* 'nice[ly]'
 b. *na'im* 'pleasant[ly]'
 c. *nora* 'awful[ly]'
 d. *gadol* 'big, great[ly]'
 e. *ħazak* 'strong[ly]'
 f. *male* 'ful[ly]'
 g. *ħazuy* 'hallucinatory, odd'

In these cases, the difference between the words is that all adjectives can be inflected according to gender and number, whereas adverbs have only one unchangeable form. All these words are known in Classical Hebrew, but they were used there only as adjectives, while in Modern Hebrew they are also used as adverbs. In the case of *nora* (7c), it appears that the use as an adverb is preferred, since in popular language a parallel form was coined, *nora'i*, using the typical form of a derived adjective and allowing it to be inflected. The use of (7d-g) as adverbs is only known in colloquial language.

4.3.4. Present Participles

In Classical and Modern Hebrew, we find adverbs using active and passive present participle forms. The same patterns are used as verbs or adjectives inflected according to gender and number, whereas adverbs are only used in the singular masculine form. These patterns are: CoCeC, CaCuC, meCaCeC, maCCiC, niCCa, muCCaC, meCuCaC.

For example:

- (8) a. *holex* 'it goes = agreed'
 b. *sagur* 'closed = agreed'
 c. *nifla* 'wonderful[ly]'
 d. *mukdam* 'early'
 e. *me'uħar* 'late'
 f. *metsuyan* 'excellent[ly]'

Many adverbs were coined in these patterns in recent years in colloquial and slang language. Below are some instances.

- (9) a. *hores* 'destroying'
 b. *madhim* 'amazing'
 c. *meħamem* 'stunning'

⁴ In examples (6b,c) and in other cases found in the same corpus, the absolute infinitive is used in a wrong form, probably because this is the most known construction.

- d. *matrif* ‘maddening’
- e. *mešagea* ‘making crazy’
- f. *metamtem* ‘making stupid’

All of the words in (9) have a positive connotation when used in slang, whether as adjectives or as adverbs. These cases are similar to the new *-ly* adverbs in English used as intensifiers, as mentioned by Killie (2015).

4.4. Linear formation

As mentioned before, beside discontinuous forms, adverbs may present linear formation, meaning prefixed or suffixed adverbs, as shown below.

4.4.1. Prefixed adverbs

In this case, we find only one type of prefix, *ha-*, which is actually the definite article added to a noun. Although we are dealing with formations with an added particle, they are considered here as adverbs, meaning lexical items, and not adverbials. Here are some examples:

- (10) a. *hayom* ‘today’
- b. *haboker* ‘this morning’
- c. *ha’erev* ‘this evening’
- d. *halayla* ‘this night’
- e. *hašavua* ‘this week’
- f. *haḥodeš* ‘this month’
- g. *hašana* ‘this year’
- h. *harega* ‘this moment’
- i. *hašniya* ‘this second’

The translation into English can be confusing, as there are different Hebrew expressions using *ha-* in the meaning of the definite article. For instance, *hayom haze* ‘this day’, *haboker haze* ‘this morning’. Note that in these cases the article is used before both words. Indeed, in Classical Hebrew, these words were used as article + noun, and not as adverbs. The examples in (10h,i) are typical of colloquial language.

4.4.2. Suffixed adverbs

In Classical Hebrew, some adverbs were marked by the suffixes *-am* or *-om*, added to a base or stem⁵, and some of them are still used to this day. For instance:

- (11) a. *hinam* ‘gratis’
- b. *dumam* ‘quietly’
- c. *yomam*⁶ ‘during the day’
- d. *reikam*⁷ ‘empty’
- e. *omnam* ‘truly’
- f. *haumnam?* ‘indeed?’

⁵ Schwarzwald (2001) calls them pseudo-base stems.

⁶ The word *yomam* is formed by the noun *yom* ‘day’, and is only used in the expression *yomam valeyl* or *yomam valayla* ‘day and night’.

⁷ This is the only case found, where the base *reik* is used as an adverb, generally in spoken language, while *reikam* is only used in literary language.

- g. *pit'om* 'suddenly'
- h. *šilšom* 'the day before yesterday'

In some cases, adverbs are derived from adjectives, adding to their stem the suffix *-ot*, which is normally a plural feminine suffix. However, this sort of formation is not productive, and the examples presented below are only used in literary language.

- (12)
- a. *ketsarot* 'shortly'
 - b. *arukot* 'longly'
 - c. *berurot* 'clearly'
 - d. *yešivot* 'directly'
 - e. *'amukot* 'deeply'
 - f. *gevohot*⁸ 'highly'
 - g. *kašot* 'hardly'
 - h. *kalot* 'slightly'

A very productive and popular adverbial formation in Modern Hebrew is adding the suffix *-it*⁹ to an existing noun. For instance:

- (13)
- a. *klalit* 'generally'
 - b. *išit* 'personally'
 - c. *'ekronit* 'in principle'
 - d. *sofit* 'finally'
 - e. *yehasit* 'relatively'
 - f. *zmanit* 'temporarily'
 - g. *rišmit* 'formally, officially'
 - h. *telefonit* 'by phone'

All the adverbs in (13) were coined in Modern Hebrew, probably based on two words found in Classical Hebrew, *rešit* 'firstly' and *šenit* 'secondly'. These two adverbs are still used today, and in popular language people say *rešit kol* 'first of all' and *šenit kol* 'second of all'. This even expanded to another pair of expressions, *alef kol* 'a of all' and *bet kol* 'b of all'.¹⁰

In colloquial Hebrew we find some cases where the suffix *-it* is added to an already existing adverb, such as *pit'omit* (from *pit'om*) 'suddenly' or *miyadit*¹¹ (from *miyad*) 'immediately'.

Note that the Hebrew suffix *-i* may transform a noun into an adjective. Therefore, we could say that the aforementioned adverbs were derived from adjectives (by adding the consonant *-t*).

Another suffix used to create adverbs is *-ayim*, probably from the number *šnayim*¹² 'two', as seen in these examples:

- (14)
- a. *pa'amayim* 'twice'
 - b. *kiflayim* 'twofold'
 - c. *šiv'atayim* 'sevenfold'

⁸ A similar expression is used in literary language, *gevoha-gevoha* 'high-high', based on a feminine adjective.

⁹ On the expanded use of the suffix *-it* in Modern Hebrew see Muchnik (1996).

¹⁰ Compare with the popular redundant adverbial *besax hakol haklali* 'on the whole generally'.

¹¹ See in section 6 (26d) about the use of *bamiyadit* '[in the] immediately'.

¹² In Modern Hebrew, we also use the expression *pi šnayim* 'twice as much'.

- d. *bentayim* ‘meanwhile’
- e. *mohrotayim* ‘the day after tomorrow’

The origin of all of the aforementioned adverbs is in Classical Hebrew, and they are still used today, but no new items were found formed with this suffix. It would appear, then, that this category of adverbs is no longer productive. However, Schwarzwald (1996) found new items in literature, particularly for children. Most of these words were nouns, and only a few of them were adverbs, among them *kiflayim* ‘twofold’ and *šiv’atayim* ‘sevenfold’, mentioned in (14), as well as *šloštayim* ‘threefold’, *arba’atayim* ‘fourfold’ and *me’odotayim* ‘lit. twice very’, which are not regularly used.

5. Foreign words

Many adverbs added to Modern Hebrew were taken from foreign languages, mostly in their original form. As in other foreign words borrowed into Hebrew, their origin is not only in different languages, but even in different language families.¹³ Below are some examples of foreign adverbs.

- (15) a. *fiks* ‘perfectly’ [English (fix)]
- b. *revers* ‘backward’ [English]
- c. *punkt* ‘exactly’ [German/Yiddish]
- d. *de-lux* ‘extra quality’ [French]
- e. *fanan* ‘enjoyably’ [Arabic]
- f. *sababa* ‘great, cool’ [Arabic]
- g. *aškara* ‘really, truly’ [Arabic]

In the next section (6, 29) we will see that not only foreign independent words are used in Modern Hebrew as adverbs, but also adverbial clauses.

6. Adverbials

Hebrew adverbials are generally formed by prepositions followed by a noun. However, in some cases the suffix *-a* is added to a noun, with the same meaning as the prefix *-le*, both indicating direction.¹⁴ In Classical Hebrew we find the suffix *-a* attached to names of places. For instance:

- (16) a. *efráta* ‘to Efrat’
- b. *ħarána* ‘to Ĥaran’
- c. *yotbáta* ‘to Yotbat’
- d. *mitsráyma* ‘to Egipt’
- e. *yerušaláyma* ‘to Jerusalem’

The adverbial formation related to names of places is only used in humorous speech nowadays. Nevertheless, it remains in use in words like these:

- (17) a. *(le)má’la* ‘up’
- b. *(le)máta* ‘down’

¹³ On the influence of foreign languages in Hebrew see Nir (1993) and Schwarzwald (1998).

¹⁴ Note that, while other suffixes are always stressed, in this case the stress is on the syllable before *-a*.

- c. *smóla* ‘to the left’
- d. *yamína* ‘to the right’
- e. *habáyta* ‘to the house = back home’
- f. *kadíma* ‘forward’
- g. *aḥóra* ‘backward’
- h. *tsafóna*¹⁵ ‘to the north’
- i. *daróma* ‘to the south’

The words in (17a, b) are used in classical and literary language also without the prefix *-le*, i.e. *má'la* and *máta*, but the longer form is used today. In popular and humorous speech, the words in (17f, g) are transformed into *kadímanit* and *aḥóranit*, which can allude that the suffix *-a* is not always felt as referring to direction.¹⁶ This can also explain the use of the word *šáma* instead of *šam* ‘there’ and the popular redundant use of *lešáma* ‘to there’.

It should be noted that Hebrew function words, and among them prepositions, are always enclitically used before content words.¹⁷ These prepositions are never stressed, and the stress remains in the original place of the main word.

The most popular preposition added to existing nouns is *be-* ‘in’, found in Classical and Modern Hebrew. For instance:

- (18) a. *besimḥa* ‘joyfully’
- b. *beratson* ‘with pleasure’
- c. *bexavana* ‘on purpose’
- d. *berogez* ‘angrily’
- e. *bekalut* ‘easily’
- f. *bekalei kalut* ‘very easily’

Following this formation, we find many new coined adverbials in popular Hebrew, and particularly in slang, like these:

- (19) a. *behikon* ‘on call’
- b. *bešlifa* ‘unsheathing’
- c. *besratim* ‘in movies = confused’
- d. *beketa' tov* ‘in a good matter’
- e. *beketa' ra'* ‘in a bad matter’
- f. *beramot* ‘in heights = extremely’
- g. *beramot 'al* ‘in super heights = extremely’

In some cases, the prefix *be-* is added to an adjective to form an adverbial. It can also include the article *ha-* becoming *ba-*. For example:

- (20) a. *begadol* ‘in big = generally’
- b. *be'anak* ‘[in] gigantic[ally]’
- c. *baprati* ‘[in the] private[ly]’
- d. *baragua* ‘[in the] calm[ly]’
- e. *baninoaḥ* ‘[in the] relaxed[ly]’

¹⁵ The words *tsafóna* and *daróma* are used in colloquial language also referring to time, i.e. ‘before’ and ‘after’ respectively.

¹⁶ Compare this to the use of *mikadíma* and *meaḥóra* below.

¹⁷ The only exception is *et*, the preposition that marks the accusative case.

f. *baktana* ‘[in the] small (FEM) = no big deal’

The preposition *–be* not only serves in the formation of one-word adverbials, but can also be added to four specific nouns denoting manner, *ofen* [MASC], *oraḥ* [MASC], *derex* [FEM] and *tsura* [FEM], attached to an adjective, and forming a great number of adverbials. Here are some possibilities:

- (21) a. *be’ofen iši* ‘in a personal mode’
 b. *be’oraḥ ḥelki* ‘in a partial manner’
 c. *bederex tipšit* ‘in a silly way’
 d. *betsura klalit* ‘in a general form’

Another way to create new adverbials in popular Modern Hebrew is by adding the preposition ‘*al*’ on’ to adjectives, such as these:

- (22) a. *’al ḥam* ‘on hot = red handed’
 b. *’al batuaḥ* ‘on secure = securely’
 c. *’al ratuv* ‘on wet = for real’
 d. *’al yaveš* ‘on dry = not for real’
 e. *’al reik* ‘on empty = groundless’

The expressions in (22c,d) are used in the military, referring to maneuvers done with loaded or unloaded weapon respectively.

In some cases the preposition ‘*al*’ is followed by the article *ha-* before a noun:

- (23) a. *’al ha’okem* ‘on the curve = crooked’
 b. *’al hamazal* ‘on the luck = haphazardly’
 c. *’al hapanim* ‘on the face = lousily’
 d. *’al hakrašim* ‘on the planks = lousily’
 e. *’al hadaka* ‘[exactly] on the minute’
 f. *’al hašniya* ‘[exactly] on the second’
 g. *’al hagova* ‘on the height = doing great’
 h. *’al hasus* ‘on the horse = doing great’

Prefixes are sometimes followed by other prepositions (underlined here), forming an adverbial that can seem contradictory. Some of them are used in formal language, such as these:

- (24) a. *mibaḥuts* ‘from [in] the outside’
 b. *milefanim* ‘from [to] front’
 c. *miberešit* ‘from [in] beginning’
 d. *milexaṯhila* ‘from [to as the] beginning’

Following this process, adverbials were also coined in slang, by attaching different prepositions:

- (25) a. *beke’ilu* ‘[in] as if = not really’
 b. *bamisaviv* ‘[in the from] around’
 c. *babetoxo* ‘[in the] inside it’
 d. *babifnoxo* ‘[in the] inside it’

In some cases, prepositions are added to existing nouns, adjectives or adverbs, and they create diverse forms of adverbials for the same meanings. For example:

- (26) a. *maher* 'quickly' > *bimhirut* 'in quickness'
 b. *klalit* 'generally' > *baklali* 'in the general'
 c. *beintayim* 'meanwhile' > *levein[a]tayim* '[to] meanwhile'
 d. *miyad* > *miyadit* 'immediately' > *bamiyadit* '[in the] immediately'
 e. *mizman* 'from time' > *mimizman* '[from from] time = long time ago'

The preposition 'ad 'till' is used in many adverbial expressions in Israeli slang, some of them loan translations, such as these:

- (27) a. 'ad kan 'till here = this is enough'
 b. 'ad ha'etsem 'to the bone'
 c. 'ad hatsavar 'up to the neck'
 d. 'ad ha'oznayim 'up to the ears'
 e. 'ad hagag 'up to the roof = extremely'

A special category of adverbials originated in Aramaic,¹⁸ and some of them are still used to this day. They are recognized by their suffix *-in* instead of the expected form *-im*, as shown below:

- (28) a. *bemeišarin* 'directly'
 b. *ba'akifin* 'indirectly'
 c. *begilufin* 'drunkenly'
 d. *leserugin* 'intermittently'
 e. *laħalutin*¹⁹ 'completely'
 f. *laħalufin* 'alternately'

In Israeli slang we find many foreign words combined with Hebrew prefixes and used as adverbials. For instance:

- (29) a. *be'izi* '[in] calmly' [English]
 b. *bedaun* 'in depression' [English]
 c. *bešvung* 'on the move' [Yiddish]
 d. *bestalbet* '[in] lazily' [Arabic]
 e. *besababi* '[in] coolly' [Arabic]

7. Summary and conclusions

The present article demonstrates that in Modern Hebrew, adverbs and adverbials constitute an open class, as there are many and very diverse types of them used to this day, mostly denoting manner. Some of the items that are known from Classical Hebrew continue in use, while new adverbs and adverbials were coined in recent years, mostly using the same ancient patterns.

Among old adverbs still used today, we find monosyllabic unmarked and underived words. Marked adverbs coined nowadays include discontinuous formation according to patterns

¹⁸ They are found in the Mishna and the Talmud, written about 2,000 years ago.

¹⁹ A very trendy word used instead of *laħalutin* is *legamre*, also taken from Aramaic.

known from Classical Hebrew, such as nominal forms, absolute infinitives, masculine adjectives, and present participles. In addition, foreign words borrowed from European languages and Arabic are used today as adverbs.

Most of the adverbials found in this study are formed by diverse prepositional prefixes, and some of them by the suffix *-a* (for the same meaning as the prefix *le-*) added to nouns, adjectives or adverbs, among them foreign words.

We have seen that, indeed, adverbs and adverbials are an open class of content words, which are very productive in Modern Hebrew, whether in the standard language or in popular language and slang. They generally follow existing ways of formation, and in most cases add connotative meanings.

References

- Amir Coffin, Edna & Shmuel Bolozky. 2005. *A reference grammar of Modern Hebrew*. Cambridge: Cambridge University Press.
- Berman, Ruth. 1987. Productivity in the lexicon: New-word formation in Modern Hebrew. *Folia Linguistica* 21(2). 225-254.
- Bolozky, Shmuel. 1999. *Measuring productivity in word formation: The case of Israeli Hebrew*. Leiden, New York & Köln: Brill.
- Bolozky, Shmuel & Ruth A. Berman. 2020. Parts of speech categories in the lexicon of Modern Hebrew. In Ruth A. Berman (ed.), *Usage-based studies in Modern Hebrew: Background, morpholexicon, and syntax*, 310-330. Amsterdam & Philadelphia: John Benjamins.
- Choueka, Yaakov. 2010. *Rav-Milim: Hamilon hašalem 'ivri 'ivri* [Many Words: A comprehensive dictionary of Modern Hebrew] (updated online).
- Cohen-Gross, Dalia. 1997. *Hamivne hamorfologi-havarati šel ha'ivrit hašadaša* [The morphological-syllabic structure of Modern Hebrew]. Ramat Gan: Bar-Ilan University dissertation.
- Diepeveen, Janneke & Freek van de Velde. 2010. Adverbial morphology: How Dutch and German are moving away from English. *Journal of Germanic Linguistics* 22(4). 381-402.
- Giegerich, Heinz J. 2012. The morphology of *-ly* and the categorial status of 'adverbs' in English. *English Language and Linguistics* 16(3). 341-359.
- Killie, Kristin. 2000. *Stative adverbs in English: A study of adverbial productivity and orientation*. Tromsø: University of Tromsø dissertation.
- Killie, Kristin. 2015. Secondary grammaticalization and the English adverbial *-ly* suffix. *Language Sciences* 47. 199-214.
- Killie, Kristin. 2022. Extravagance, productivity and the development of *-ingly* adverbs. In Matthias Eitelmann & Dagmar Haumann (eds.), *Extravagant morphology: Studies in rule-bending, pattern-extending and theory-challenging morphology*, 51-72. Amsterdam & Philadelphia: John Benjamins.
- Kogut, Simcha. 2002. Lema'amadam šel šmot hatō'ar veto'ore hapō'al bilšon hamikra: 'iunim morfologiyim vetahbiryim vetašlaxot paršaniyot [The status of adjectives and adverbs in Biblical Hebrew: Morphological and syntactic studies and exegetical implications]. *Šnaton leheker hamikra vehamizraħ hakadum* [Yearbook for Biblical and ancient Near Eastern studies] 13. 111-137.
- Maienborn, Claudia & Jena Martin Schäfer. 2011. Adverbs and adverbials. In von Heusinger, Klaus, Claudia Maienborn & Paul Portner (eds.), *Semantics: An international handbook of natural language meaning*, 1390-1420. Mouton de Gruyter.
- Marle, Jaap van. 1985. *On the paradigmatic dimension of morphological creativity* (Language Sciences 18). Walter De Gruyter.
- Marle, Jaap van. 1992. The relationship between morphological productivity and frequency: A comment on Baayen's performance-oriented conception of morphological productivity. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology*, 151-163. Dordrecht: Springer.
- Muchnik, Malka. 1994. Lo stam sleng [Not just slang]. *Leshonenu La'am* [Our Popular Language] 45(2). 65-74.
- Muchnik, Malka. 1996. Sofit ketana še higdila la'asot [A little suffix growing up]. *Helkat Lashon* 23. 13-22.
- Muchnik, Malka. 2000. Darxe tsuratam šel šmot hatō'ar bilšon hasleng [Adjective formation in Hebrew slang]. In Ora R. Schwarzwald, Shoshana Blum-Kulka & Elite Olshain (eds.), *Sefer Raphael Nir* [Raphael Nir Jubilee Book], 377-389. Jerusalem: Carmel.
- Muchnik, Malka. 2004. Is Modern Hebrew a synthetic or analytic language? Suffixed and independent pronouns in the writing of Haim Be'er. *Folia Linguistica* XXXVIII (3-4). 257-276.

- Nir, Bracha & Ruth Berman. 2010. Parts of speech as constructions: The case of Hebrew 'adverbs'. *Constructions and Frames* 2(2). 242–274.
- Nir, Raphael. 1993. *Darxe hayetsira hamilonit be'ivrit bat-zmanenu* [Word-formation in Modern Hebrew]. Tel Aviv: The Open University of Israel.
- Pittner, Karin, Daniela Elsner & Fabian Barteld. 2005. *Adverbs: Functional and diachronic aspects*. Amsterdam & Philadelphia: John Benjamins.
- Pounder, Amanda V. 2001. Adverb-marking in German and English: System and standardization. *Diachronica* 18(2). 301-358.
- Ravid, Dorit. 2020. Derivation. In Ruth Berman (ed.), *Usage-based studies in Modern Hebrew: Background, morpho-lexicon, and syntax*, 203-264. Amsterdam & Philadelphia: John Benjamins.
- Ravid, Dorit & Yitzhak Shlesinger. 2000. Modern Hebrew adverbials: Between syntactic class and lexical category. In Ellen Contini-Morava & Yishai Tobin (eds.), *Between grammar and lexicon*, 333-351. Amsterdam & Philadelphia: John Benjamins.
- Rosenthal, Ruvik. 2005. *Milon hasleng hamakif* [Dictionary of Israeli slang]. Jerusalem: Keter.
- Salazar García, Ventura. 2007. Flexibilidad categorial y adverbios de manera en español: Un enfoque funcional. In Pedro Barros García, Gonzalo Águila Escobar & Estéban Tomás Montoro del Arco (eds.), *Estudios lingüísticos, literarios e históricos*, 309-326. Granada: Editorial Universidad de Granada.
- Saydon, Vered. 2018. 'Od 'al ški'ei dikduk hamikra ba'ivrit haḥaya [More on biblical grammar components in the living Hebrew language]. *Lešonenu* [Our Language] 70. 305-323.
- Schwarzwald, Ora R. 1989. Lešimuš hamakor hamuḥlat ba'ivrit haḥadaša [On the use of the absolute infinitive in Modern Hebrew]. *Lešonenu* [Our Language] 53. 107-112.
- Schwarzwald, Ora R. 1996. Milim besiomet –ayim be'ivrit: Tsuratan, yetsiratan vešimušeihen [Words with -ayim endings in Hebrew: Their form, formation and use]. In Moshe Bar-Asher (ed.), *Mehkarim balašon ha'ivrit uvilšonot hayehudim* [Studies in Hebrew and Jewish Languages], 341-358. Jerusalem: Bialik Institute.
- Schwarzwald, Ora R. 1998. Word foreignness in Modern Hebrew. *Hebrew Studies* 39. 115-142.
- Schwarzwald, Ora R. 2001. *Modern Hebrew*. München: Lincom Europa.
- Schwarzwald, Ora R. 2019. *Mehkarim be'ivrit bat zmanenu* [Studies in contemporary Hebrew]. Jerusalem: The Academy of the Hebrew Language.
- Schwarzwald, Ora R. & Dalia Cohen-Gross. 2000. Hamiškalim hašxiḥim be'ivrit [Frequent patterns in Hebrew]. In Miri Horwitz (ed.), *Lešon ha'itonut beyamenu* [Journalistic language nowadays], 148-161. Or Yehuda: Reches.
- Tagliamonte, Sali & Rika Ito. 2002. Think really different: Continuity and specialization in the English dual form adverbs. *Journal of Sociolinguistics* 6(2). 236-266.
- Talmy, Leonard. 2000. *Towards a cognitive semantics*. Cambridge: Cambridge University Press.

Where did the Italian Verbal-Nexus N+N compounds come from?

Jan Radimský
University of South Bohemia in České Budějovice
radimsky@ff.jcu.cz

1. Introduction

This article aims to trace the origin and development of Italian Verbal-Nexus N+N compounds (henceforth VNX NNs), such as *trasporto merci* – ‘transport of goods’, based on the Google n-grams frequency lists (2020), which are the most extensive diachronic linguistic data currently available.

Italian VNX NNs represent a prominent – and probably the only productive – higher-order subordinate NN construction in Romance (Rainer 2016, Baroni, Guevara & Zamparelli 2009, Radimský 2018) and it is considered, along with others subordinate Italian NNs, a very recent innovation. Indeed, the existing literature does not report cases of subordinate N+N Italian compounds attested before 1950 (Tolemache 1945, Micheli 2020a, 2020b). The first examples are assumed to appear around the 1970s (Dardano 2009:226-229), presumably under a certain influence of American English, they tend to be associated with specific contexts of use, namely with telegraphic language of journalism (journal titles) and the style of bureaucratic documents, while in spoken Italian they are rather sparse (Baroni et al. 2009). However, to the best of my knowledge, the diachronic evolution of Italian subordinate NNs has not yet been empirically investigated on large corpora. In order to fill this gap, the present article provides a detailed analysis of a large sample of Italian VNX NNs in the most recent version of Google n-grams (2020) data, within the theoretical framework of Construction Morphology (CM, Booij 2010, 2016, Traugott & Trousdale 2013) and Relational Morphology (RM, Jackendoff & Audring 2020).

The paper is organized as follows: Section 2 will outline basic properties of Italian VNX NNs, Section 3 will sketch the theoretical background and hypotheses concerning the diachronic emergence of this pattern within the CM and RM framework; Section 4 will discuss the data-gathering process and Section 5 will present in turn results concerning the diachronic profile of the whole sample (relative type and token frequency, 5.1.), data about the first/last appearance of individual compounds (5.2.) and diachronic profiles concerning the most prominent N-1 and N-2 based families or ‘semi-schematic constructions’ (5.3.).

Although the technical processing as well as the interpretation of such a large data is very challenging, two important conclusions emerge from this analysis. First, we will show that Italian VNX NNs are older than previously assumed, their emergence is linked to the bureaucratic language of the newly established Italian kingdom in the mid-19th century and the pattern was popular especially during the Fascist period. Second, we will attempt to put forward hypotheses about the respective roles of N1(head)-based and N2(argument)-based families in the process of creation of the VNX NN compounding pattern, which may be of a more general interest.

2. Italian verbal-nexus NNs

Italian *Verbal-nexus NNs* (also referred to as *Argumental NNs*) represent a subtype of endocentric subordinate compounds consisting of a deverbal head and a non-head element which is interpreted as its argument. Over the past decade, a number of studies have been devoted to them, focusing on three questions in particular, namely:

- (i) Should they be analyzed rather as morphological constructions, i.e. compounds (cf. Gaeta & Ricca 2009, Masini & Scalise 2012, Radimský 2015, Lami & Weijer 2022), as compound-like syntactic phrases (Bisetto-Scalise 1999, Delfitto & Paradisi 2009) or as a heterogenous class (Baroni, Guevara & Zamparelli 2009)?
- (ii) How to delimit this category? First of all, should it only cover cases where the non-head element is the internal argument of the deverbal head – that is, in terms of Generative grammar, the direct object or the subject of the underlying unaccusative verb – (Baroni, Guevara & Zamparelli 2009, Baroni, Guevara & Pirrelli 2009), or should it include also other types of predicate-argument (or even predicate-adjunct) relationship (Scalise & Bisetto 2009, Radimský 2015)?
- (iii) What morphological (e.g. inflection) and syntactic (syntactic atomicity) properties do they have (Bisetto-Scalise 1999, Baroni, Guevara & Zamparelli 2009, Radimský 2015, Lami & Weijer 2022)?

In this paper, I will leave aside the question (iii) concerning morphological and syntactic properties of VNX NNs and as for the point (i), all VNX NNs will be treated as a homogeneous group of subordinate compounds that represents one morphological higher-level construction.¹ As for the delimitation of the VNX NNs (ii), I will adopt a permissive approach in line with Scalise and Bisetto (2009) that involve all different types of predicate-argument or predicate-adjunct relationship. However, the core group of ‘canonical’ VNX NNs in line with Baroni, Guevara and Zamparelli (2009) will be predominant in the data, as it is also in current use.

Indeed, the starting point of this research is a sample of 1,364 VNX NNs collected by Radimský (2015), where 80% of types (let us call them “canonical VNX NNs”) feature a deverbal event noun as head and the non-head (its argument) corresponds to the direct object of the underlying verb. The head may be either a zero-derived (1a-1b) or a suffixed (1c) noun.

- (1) a. *noleggio auto* (rental_car) – ‘car rental’
 b. *trasporto merci* (transport_goods.pl) – ‘goods transport’
 c. *trattamento rifiuti* (treatment_vaste.pl) – ‘waste treatment’

The remaining 20% of the sample represent various non-canonical VNX NNs, be it with respect to the properties of the non-head or the head element. That is, the non-head may have a different role than the direct object (2-6), the head may be a (deverbal) result noun (7) or a deadjectival noun (8).

- (2) *caduta massi* – “rockfall” (non-head = subject of an unaccusative verb)
- (3) *attacco hacker* – “hacker attack” (non-head = subject of a transitive verb)
- (4) *accusa maltrattamento* – “allegation of ill-treatment” (non-head = indirect object)
- (5) *applicazione laser* – “laser application” (non-head = adjunct)
- (6) *uscita autostrada* – “highway exit” (non-head = adjunct)

¹ Notice that within the Construction grammar framework adopted here, all constructions are of the same nature, be they morphological or syntactic, so the dilemma is irrelevant.

- (7) *deposito bagagli* – “luggage [storage room]” (the head is a result noun)
 (8) *pericolo terrorismo* – “terrorism danger” (the head is deadjectival)

Canonical VNX NNs, i.e., examples (1a-1c), may be described in terms of Construction and Relational morphology as a pair of sister constructions schematized in (8). As various scholars agree, this pattern represents a prominent higher-order subordinate NN construction available in Romance (Rainer 2016; Baroni, Guevara & Zamparelli 2009).

- (8) $[N_i N_j]_{Nk} \leftrightarrow [V_i > N_{i\text{-head}} \mathbf{REL}_X N_{j\text{-non-head}}]_k$
 $[V_i N_j]_{Nk} \leftrightarrow [V_i \mathbf{REL}_X N_{j\text{-direct_object}}]_k$

Single instances of VNX NNs are also attested in French (9), but by far lacking the regularity present in Italian data (Radimský 2018).

- (9) Fr. *exposition photos* – “photography exhibition”

3. Theoretical background

Construction Morphology as well as Relational Morphology are conceived of as usage-based models, which entails that schemas available in the Constructicon capture generalizations over a critical mass of already attested words. In other terms, when it comes to the emergence of new constructions in a diachronic perspective, “constructionalization” must be based on previous individual “innovation” (in the sense of Traugott & Trousdale 2013). It is not the aim of this study to find out where the various individual Italian “innovations” – i.e. first examples of VNX NNs – came from, but to date their origin and to trace the process of “constructionalization” that led to the emergence of the productive schema of VNX NNs described in (8) above.

The process of constructionalization is not a matter of just one schema, but of the whole hierarchical network of schemas in the constructicon. In our case, the subordinate VNX NN construction (10c) represents a specific case of the subordinate NN construction (10b) which is in turn an instance of the more general left-headed NN pattern (10a). A similar hierarchy can be observed in the reverse direction, because between the general VNX SUB NN schema (10c) and the individual instances of compounds (10e) we can assume the existence of semi-schematic VNX SUB NN constructions (10d) based either on the same head noun (10d1) or on the same non-head noun (10d2).

- (10) a. Left-headed NN construction
 $[N_i N_j]_{Nk} \leftrightarrow [N_{i\text{-head}} N_{j\text{-non-head}}]_k$
- b. SUB NN construction
 $[N_i N_j]_{Nk} \leftrightarrow [N_{i\text{-head}} \mathbf{REL} N_{j\text{-non-head}}]_k$
- c. VNX SUB NN construction
 $[N_i N_j]_{Nk} \leftrightarrow [V_i > N_{i\text{-head}} \mathbf{REL}_X N_{j\text{-non-head}}]_k$
 $[V_i N_j]_{Nk} \leftrightarrow [V_i \mathbf{REL}_X N_{j\text{-direct_object}}]_k$
- d. SUB NN semi-schematic constructions
 d.1 SUB NN semi-schematic constructions based on the same N1
 $[\mathbf{trasporto} N_j]_{Nk} \leftrightarrow [\mathbf{TRASPORTO}_{i\text{-head}} \mathbf{REL} N_{j\text{-non-head}}]_k$
 ...

- d.2 SUB NN semi-schematic constructions based on the same N2
 $[N_i \textit{merci}_j]_{Nk} \leftrightarrow [N_{i\text{-head}} \textit{REL merci}_{j\text{-non-head}}]_k$
 ...

e. Individual instances of NNs

- e.1 *noleggio auto* ('car rental')
 e.2 *trasporto merci* ('freight transport')
 e.3 *trattamento rifiuti* ('waste treatment')
 e.4 *scarico merci* ('goods unloading')
 e.5 *trasporto persone* ('passenger transport')
 ...

To the question of the interrelation of hierarchical constructions in the process of constructionalization, recent research in the framework of Construction Grammar gives a fairly unambiguous answer: it is a bottom-up process, where new schemas correspond to areas in which examples encountered so far cluster (cf. the notion of *coverage* by Goldberg 2019: 51-73), while increasing type frequencies of lower-order schemas do not automatically strengthen the mental representation of higher-order schemas (cf. Hilpert 2015 for compounds).

A similar view is also offered by Relational Morphology, which does not yet have a comprehensive model of diachronic language development, but whose premisses about constructionalization in language acquisition can be easily applied to the language change (Jackendoff & Audring 2020: 218-232). Constructionalization in Relational Morphology consists of two steps. First, relational links between the existing words must be built through the process of "Structural Intersection", and then it is necessary to determine whether these new relational schemas are productive. The key operation of Structural Intersection (Jackendoff & Audring 2020: 223-225) is quite straightforward in the case of derivation, where the shared phonological material corresponds to the affix, and the shared morphosyntactic as well as semantic properties must also be associated with it, at least in some way. When it comes to compounds, however, we encounter a serious difficulty, because between the individual instances of compounds (10e) and the closest schematic construction (10c) there is no shared phonological material, which entails that Structural Intersection would have to be entirely based on very abstract semantic and/or morphosyntactic categories and relations. It therefore seems reasonable to hypothesize that semi-schematic constructions, such as those in (10d), may play a prominent role in the process of compounds constructionalization. Such a view is not new: it is consistent with the assumption of Laurie Bauer (2017: 74) that "*it is not the N+N pattern of compounding which is productive, but patterns with individual lexemes within that*", as well as with the observation of Franz Rainer (2016:2714) that within Italian N+N compounds, "*neologisms tend to follow analogues or series of analogues with the same first or second constituent.*" Although it may seem counterintuitive from a functional point of view, a quantitative study on French N+N compounds has shown that such family-size effect is prominent with both N1 (=head noun) based and N2 (non-head noun) based families (Radimský 2020). One of the questions addressed in this investigation will therefore be: what is the role of semi-specified constructions (families) in diachrony?

4. Data gathering

The research is based on extensive diachronic data drawn from the *Google books* corpus that has been made available by Google in the form of raw frequency lists as the 3rd version of

Italian *Google n-grams*.² The size of the underlying Google books corpus is 120,410,089,963 tokens from 1,209,932 volumes,³ which – by a simple extrapolation of figures provided by Lin et al. (2012) – may represent roughly 16% of volumes ever published before 2010. Data for the extraction of N+N compounds come from bigrams and trigrams (in order to capture compounds with space-separated and hyphen-separated components, respectively) that were pre-treated and merged together into the *it2020_bi* dataset⁴ using the procedure described in detail by Radimský (2022). The whole *it2020_bi* dataset from which function words have been filtered out comprises 19,319,372 non-lemmatized types.

The starting point for subsequent data filtering was the sample of 1,364 contemporary VNX NN compounds (lemmas) identified in the ItWac corpus by Radimský (2015). On this basis, a sample of 1,185 VNX compounds (words) was retrieved in the *it2020_bi* dataset. In order to achieve a higher accuracy, most compounds have been checked back manually in Google books and many false positives have been eliminated. Word forms rather than lemmas have been used as basic units, because it turned out that by virtue of morphological ambiguity, some inflected forms are a frequent source of false positives in real texts, as exemplified in (11).

- (11) a. *valutazione*.sg *danno*.N.sg/V.pl. (NN) ‘damage assessment’ – true positive
 b. *valutazione*.sg *danni*.N.pl (NN) ‘damage assessment’ – true positive
 c. *valutazioni*.pl *danno*.N.sg/V.pl. (NV) ‘provide evaluation’ – false positive

False positives due to syntactic ambiguity, such as (12), have also been filtered out.

- (12) a. [*uscita merci*]? ?‘goods exit’
 b. [*se vengono dichiarate [per l'uscita]*] [*merci di cui non occorre che sia provata l'esportazione*] – ‘If goods whose export is not to be proved are declared for exit’

On the other hand, a number of new compounds were added to the sample due to the fact that additional types could be retrieved manually for prominent semi-specified constructions (families).

For the final sample of 1,185 VNX compounds, dated numbers of occurrences in Google books were available from 1850 to the present with a year-by-year precision.

5. Results

5.1. Diachronic profile of the whole sample

A comprehensive diachronic overview of the use of Italian VNX NNs is illustrated by Figures 1 and 2 that provide, respectively, the sum of relative token frequencies and the relative type frequency for all the compounds in our sample. To identify diachronic trends and draw regression lines, Theil-Sen estimator was used and supplemented, where necessary, with the Mann-Kendall test for significance testing (cf. Kovář & Herman 2013, Python implementation by Hussain & Ishtiaq 2019).⁵

² <https://storage.googleapis.com/books/ngrams/books/datasetsv3.html>

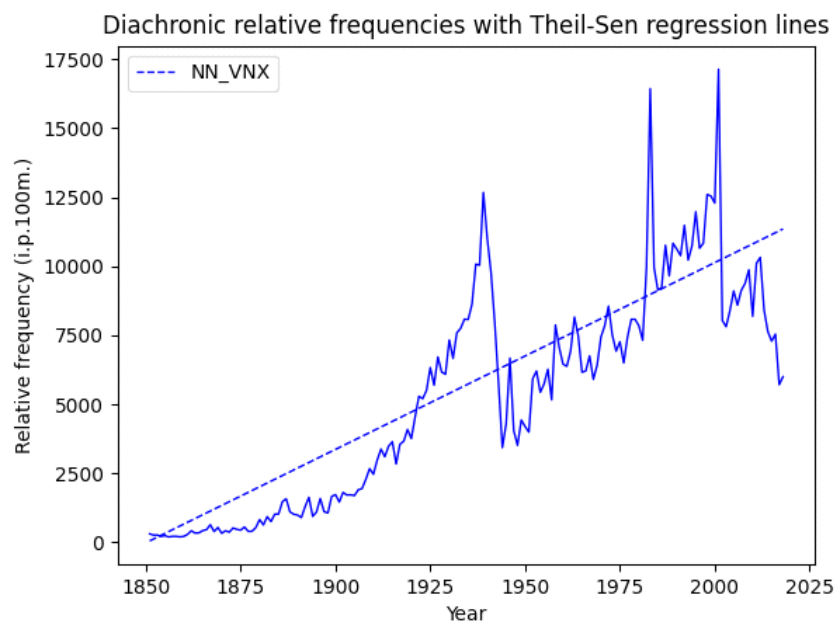
³ “Volumes” in Google books are intuitively associated with “books”, but a qualitative look at the data shows that nowadays, Google books contain also other types of printed and published materials.

⁴ The *it2020-bi* dataset is available for download at: <https://osf.io/46qcd/>

⁵ As Kovář & Herman (2013) point out, the Theil-Sen estimator is a rank-based non-parametric method suitable to test any form of dependence (not only linear). Since it does not assume a normal distribution of errors, it is not sensible to outliers and therefore it is particularly suitable for trend identification of word usage in diachronic corpora.

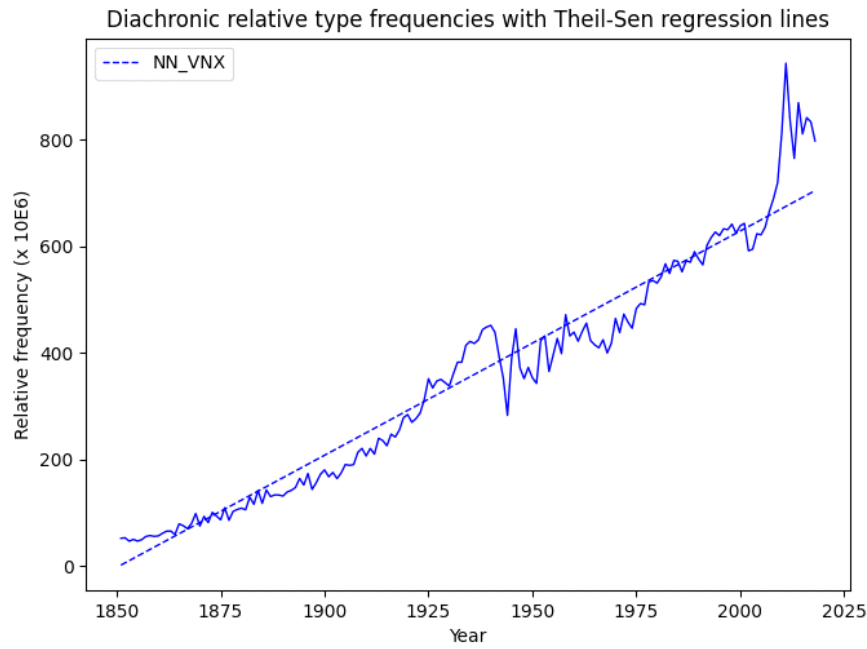
Figure 1 shows that the use of VNX NNs in Italian steadily increases between 1850 and 2000 with three major peaks in 1930's-1940's (the fascist period), 1980's and 2000's, respectively. We will not attempt to interpret the subsequent drop in frequency, since data for the period after 2010 might be strongly biased by a different composition of the underlying Google books corpus as a result of copyrights issues. The essential point in any case is that the history of Italian VNX NNs is roughly 100 years longer that assumed.

Figure 1: Relative token frequency of Italian VNX NNs



The relative type frequency curve in Figure 2 confirms the steady increase of the Italian VNX NN pattern since 1850's. Between 2000 and 2010, an interesting phenomenon occurs: the type frequency of the VNX NN construction exponentially increases, although its overall token frequency decreases. The question arises whether this could be considered a sign of “productivity upgrade” of the VNX NN pattern, which would mean that the solely relational schema (10c) is shifting towards a productive status in that period of time (Jackendoff & Audring 2020:228-231).

Figure 2: Relative type frequency of Italian VNX NNs



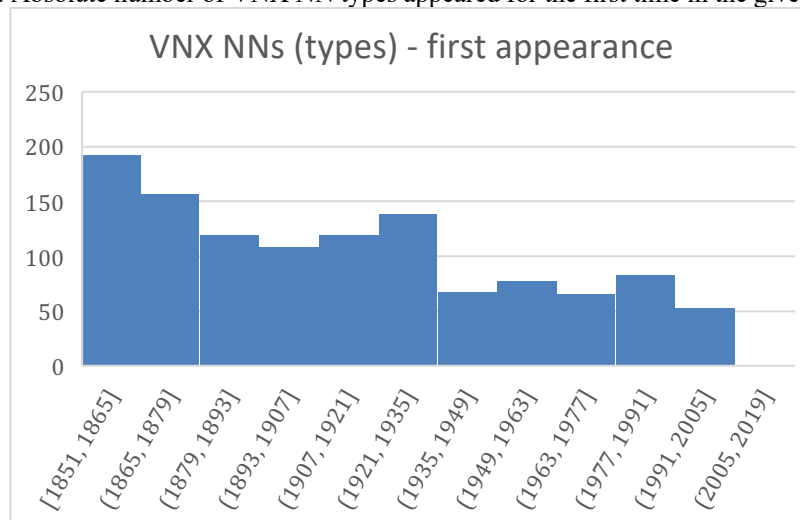
5.2. First and last appearance of VNX NN compounds

The diachronic evolution of VNX NN compounds can also be observed from a different perspective by examining years of the first and the last appearance of the types from the sample. The mean and the median of these figures are given in Table 1, while the graph in Figure 3 displays absolute numbers of types that appeared for the first time in the different 14-year periods that evenly cover the entire time span under investigation.

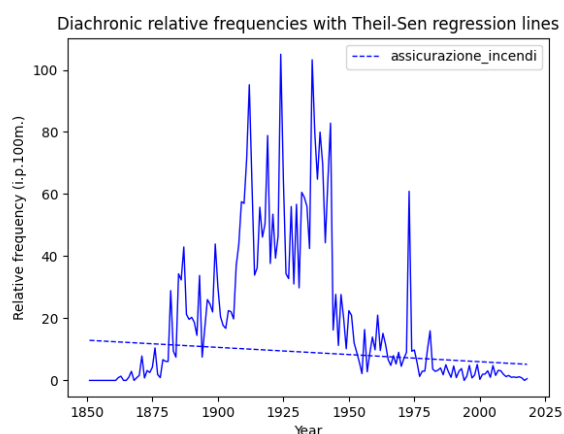
As for the years of first attested occurrences, the mean and the median are surprisingly low, considering that the first examples of VNX NNs were assumed to appear around the 1970s (Dardano 2009:226-229). The graph in Figure 3 confirms this observation and shows that most of the types (approximately 2/3) were attested for the first time already before the year 1935. Since 1935 there has been a slower but steady inflow of new types.

Table 1: Mean and median of years of the first and the last appearance of VNX NN compounds

	First appearance	Last appearance
Mean	1912,853	2017,55
Median	1909	2019

Figure 3: Absolute number of VNX NN types appeared for the first time in the given time span

Conversely, the figures concerning years of the last attested occurrences given in Table 1 are very high, which entails that almost all types persist in usage until the present time. This naturally does not mean that they have the same or increasing token frequency – many of them had their “period of glory” in the past and their token frequency decreases, as illustrated by the example in Figure 4.

Figure 4: Diachronic token fq. of *assicurazione incendi* (“fire insurance“)

5.3. Role of semi-schematic constructions

In this section we will attempt to empirically examine the role of semi-schematic constructions (also referred to as *N-1* and *N-2 based families*) in the diachronic evolution of VNX NN compounds. Diachronic type frequency curve of the 7 most prominent *N-1* and *N-2* based families is given in Figures 5 and 6, respectively.

Figure 5: Diachronic type frequency of the 7 most prominent N1-based families ('risk', 'management', 'organization', 'transport', 'request', 'management', 'change')

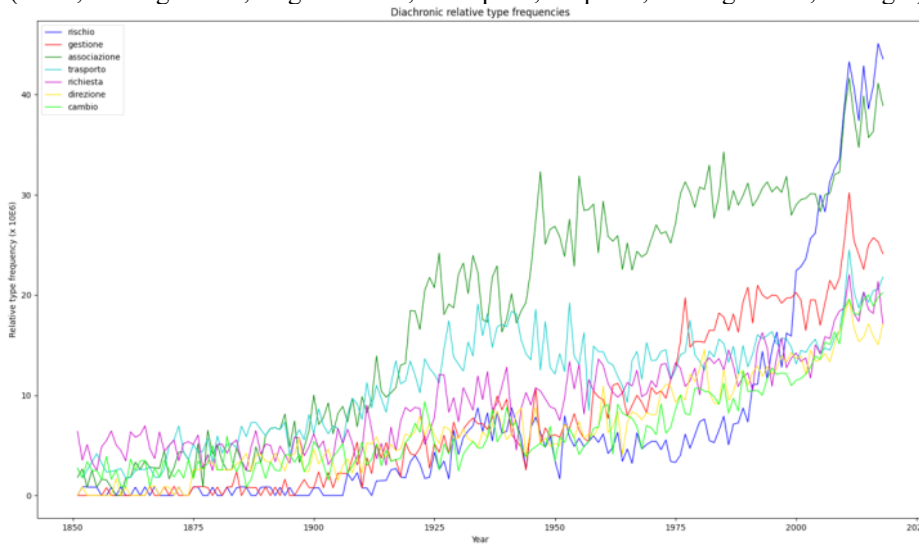
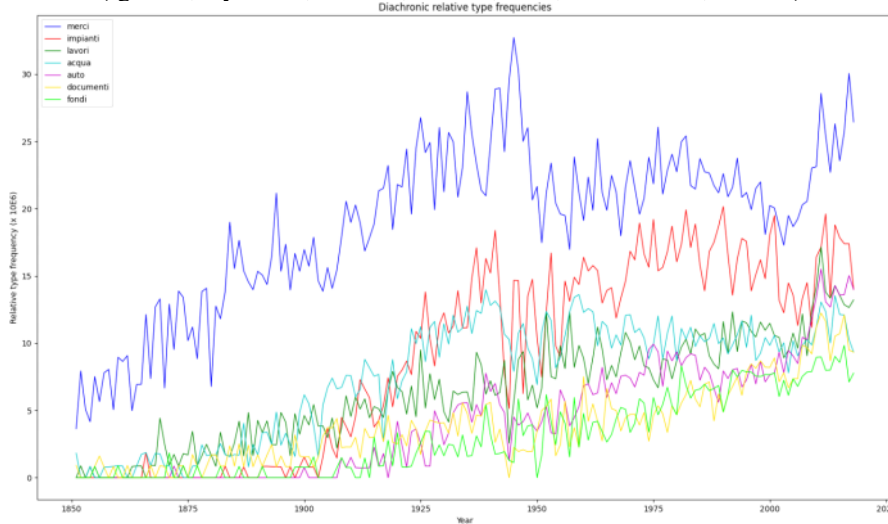


Figure 6: Diachronic type frequency of the 7 most prominent N2-based families ('goods', 'systems', 'works', 'water', 'car', 'documents', 'funds')



One might intuitively expect that the leading role in the process would be played by head-based families, i.e. by semi-schematic constructions with a specified N1. However, the picture given by Figures 5-6 is more complicated. The difference between N-1 and N-2 based families is not a quantitative one, because N2-based families display similar type frequencies as N-1 based families do, but the respective curves are differently distributed in time.

The type frequency of some N1-based families begins to increase slowly after 1910 (*associazione* – ‘organization’, *trasporto* – ‘transport’), but a clear and rapid growth of all seven N1-based families takes place only after 1975 (trend = increasing with $p < 2 \times 10^{-6}$ and slope between 0,136 – 1,0). Conversely, the growth of N2-based families took place earlier and slowed down considerably after 1950. Notice that until 1925, the type frequency of the leading construction $[N_i \text{ *merci*}]_{Nk}$ (‘goods’) outperforms all the others, including the N1 families, and it is already fully saturated around 1950. The frequency of the other N-2 based constructions also increases until 1950. But after 1975, when a rapid growth of N1-based families takes place, the type frequency of the seven N-2 based families display either no significant trend (*merci* – ‘goods’, *impianti* – ‘systems’, *acqua* – ‘water’) or only a slow increase (*lavori* – ‘works’, *auto*

– ‘car’, *documenti* – ‘documents’, *fondi* – ‘funds’), with respective slopes between -0,05 and 0,14.

The aggregate type frequency data for all families from the sample also lead to the same conclusion. Table 2 provides the mean and the median of years in which each family of the sample reached the highest type frequency.

Table 2: Mean and median of years of the highest type frequency of all N-1 and N-2 based families

	N1	N2
Mean	1978,9	1973,3
Median	2010	1951

Although the means are quite similar, the medians are very different, which entails that many N2-based families displayed the highest type frequency already in the early 20th century, while N1-based families contributing to today’s growth of the VNX NN pattern became more saturated in the second half of the 20th century.

Such a difference between the diachronic role of N-1 and N-2 based families might be explained in functional terms. Since argument nouns (N2s), such as *merci* (‘goods’), *impianti* (‘systems’) or *acqua* (‘water’), are closely related to concrete topics and therefore to concrete genres and texts, it is likely that they will be easier to replicate in these areas within similar structures – i.e. that they will more easily begin to form semi-schematic constructions. Conversely, head nouns (N1s) are less linked to concrete topics, so it can be expected that N-1 based semi-schematic constructions will need more time and more source examples before they emerge. And since argument nouns have, for obvious semantic reasons, a more restricted combinability than common deverbal heads in purely quantitative terms, N2-based constructions will reach full saturation quite early, so that their type frequency can no longer continue to grow.

6. Conclusions

As this first large scale diachronic investigation on the topic suggests, the history of Italian VNX NNs is more intriguing than assumed in previous literature. First instances (‘innovations’) of this type did not appear around 1950’s – 1970’s (Micheli 2020a, 2020b, Dardano 2009), but at least already since 1850’s. A qualitative look into the data reveals that they were emerging especially in the context of the bureaucratic and economic language of the newly established Italian kingdom in the mid-19th century. Besides that, the particular popularity of the VNX NN pattern during the fascist period in 1930’s – 1940’s might also be accounted for by the fascist regime’s affinity for Marinettian futuristic aesthetics that glorified speed, directness and simplicity in language, so that no useless function words, such as prepositions, were particularly welcome. It was only in the second half of the 20th century that VNX NNs fully penetrate into journalistic language – which is entirely consistent with the widely shared assumption that the bureaucratic language of the newly formed Italian state was an important source of innovations that were later conventionalized in the journalistic language and in other registers of Italian (cf. Viale 2008:91-94).

Analyses of type and token frequency curves suggest that the pattern has steadily grown during the whole period since 1850’s to the present, with two periods of particularly rapid type frequency increase (1930’s-40’s and since 2000). The latter, correlated even with a token frequency decrease, might perhaps be considered as a progressive shift of the relational VNX NN construction towards a productive state in terms of Relational morphology (Jackendoff & Audring 2020: 228-231). We have also analyzed the different role of N-1 and N-2 based semi-schematic constructions in the complex process of constructionalization, showing that the type

frequency of N-2 based constructions grew earlier than the type frequency of N1-based families. Therefore, it might be hypothesized that only after 2000's the whole VNX NN pattern reached the necessary coverage (in terms of Goldberg 2019: 51-73) by various individual instances and semi-schematic constructions in order to be ready for a 'productivity shift' observed on the global type frequency curve.

Finally, it has to be emphasized that the investigation presented in this paper certainly does not tell the whole story about the emergence of Italian VNX NNs. Besides obvious methodological issues (such as subsequent reduction of false positives, qualitative identification of contexts and genres, not to speak about the still problematic composition of the underlying Google books corpus), the research will need to be complemented in the future by at least two aspects. First, other prominent Italian NN compounds (i.e. coordinate, attributive and grounding) have to be studied in diachrony, including their interaction with the VNX NN pattern. Second, diachronic competition between the VNX NNs and the respective prepositional NPN structures have to be examined thoroughly, since any Italian VNX NN, such as *trasporto merci*, has a licit NPN equivalent (*trasporto di merci*). To put it differently, the emergence of the Italian VNX NN pattern is a prominent illustration of the fact that "grammatical constructions tend to emerge in domains that are already relatively well represented by other constructions" (Hilpert 2021:149). It would be tempting to understand why this occurs.

References

- Baroni, Marco, Emiliano Guevara & Roberto Zamparelli. 2009. The dual nature of Deverbal Nominal Constructions: Evidence from acceptability ratings and corpus analysis. *Corpus Linguistics and Linguistic Theory* 5(1). 27–60. doi:[10.1515/CLLT.2009.002](https://doi.org/10.1515/CLLT.2009.002).
- Baroni, Marco, Emiliano Guevara & Vito Pirrelli 2009. Sulla tipologia dei composti N+N in italiano : principi categoriali ed evidenza distribuzionale a confronto. *Linguistica e modelli tecnologici di ricerca : atti del XL Congresso internazionale di studi della Società di linguistica italiana (SLI) : Vercelli, 21-23 settembre 2006*. 73-96. Bulzoni Editore.
- Bauer, Laurie 2017. *Compounds and Compounding*. Cambridge University Press. doi:[10.1017/9781108235679](https://doi.org/10.1017/9781108235679).
- Bisetto, Antonietta, & Scalise, Sergio 1999. Compounding: Morphology and/or syntax? In Lunella Mereu (ed.), *Boundaries of Morphology and Syntax*, 31-48. Amsterdam: Benjamins.
- Booij, Geert E. 2010. *Construction morphology* (Oxford linguistics). Oxford & New York: 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. doi:[10.1017/9781139814720.016](https://doi.org/10.1017/9781139814720.016).
- Dardano, Maurizio 2009. *Costruire parole: la morfologia derivativa dell'italiano* (Itinerari). Bologna: il Mulino.
- Gaeta, Livio & Davide Ricca 2009. Composita solvantur: Compounds as lexical units or morphological objects? *Italian Journal of Linguistics* 21(1). 35–70.
- Goldberg, Adele E. 2019. *Explain me this: creativity, competition, and the partial productivity of constructions*. Princeton, New Jersey: Princeton University Press.
- Hilpert, Martin 2015. From hand-carved to computer-based: Noun-participle compounding and the upward strengthening hypothesis. *Cognitive Linguistics* 26(1). 113–147. doi:[10.1515/cog-2014-0001](https://doi.org/10.1515/cog-2014-0001).
- Hilpert, Martin 2021. *Ten lectures on diachronic construction grammar* (Distinguished lectures in cognitive linguistics vol. 26). Leiden & Boston: Brill.
- Hussain, Md. & Ishtiaq Mahmud 2019. pyMannKendall: a python package for non parametric Mann Kendall family of trend tests. *Journal of Open Source Software* 4(39). 1556. doi:[10.21105/joss.01556](https://doi.org/10.21105/joss.01556).
- Jackendoff, Ray & Jenny Audring 2020. *The texture of the lexicon: relational morphology and the parallel architecture*. Oxford & New York: Oxford University Press.
- Kovář, Vojtěch & Ondřej Herman 2013. *Methods for Detection of Word Usage over Time*. Tribun EU.
- Lami, Irene & Joost van de Weijer 2022. Compound-internal anaphora: evidence from acceptability judgements on Italian argumental compounds. *Morphology* 32. 359–388. doi:[10.1007/s11525-022-09398-w](https://doi.org/10.1007/s11525-022-09398-w).

- Lin, Yuri, Jean-Baptiste Michel, Erez Lieberman Aiden, Jon Orwant, Will Brockman & Slav Petrov 2012. Syntactic annotations for the Google Books Ngram Corpus. *Proceedings of the ACL 2012 System Demonstrations* (ACL '12), 169–174. USA: Association for Computational Linguistics.
- Masini, Francesca & Sergio Scalise 2012. Italian compounds. *Probus* 24(1). 61–91. doi:[10.1515/probus-2012-0004](https://doi.org/10.1515/probus-2012-0004).
- Micheli, M. Silvia 2020a. *Composizione italiana in diacronia: Le parole composte dell'italiano nel quadro della Morfologia delle Costruzioni*. Berlin / Boston: De Gruyter. doi:[10.1515/9783110652161](https://doi.org/10.1515/9783110652161).
- Micheli, M. Silvia 2020b. *La formazione delle parole: italiano e altre lingue* (Studi superiori ; Linguistica 1237). Roma: Carocci editore.
- Radimský, Jan 2015. *Noun+Noun compounds in Italian: a corpus-based study* (Episteme. Theoria). Vydání 1. České Budějovice: Jihočeská univerzita v Českých Budějovicích.
- Radimský, Jan 2018. Does French have verbal-nexus Noun+Noun compounds? *Linguisticae Investigationes* 41(2). 213–223. doi:[10.1075/li.00020.rad](https://doi.org/10.1075/li.00020.rad).
- Radimský, Jan 2020. A paradigmatic account of lexical innovation: the role of repeated components in French N+N compounds. In Jenny Audring, Nikos Kousoukos & Christina Manouilidou (eds.) *Mediterranean Morphology Meetings; Vol 12 (2019): Rules, patterns, schemas and analogy*. Patras. 77-91. doi:[10.26220/mmm.3251](https://doi.org/10.26220/mmm.3251).
- Radimský, Jan 2022. Towards a diachronic analysis of Romance morphology through Google n-grams. *Corpus* (23) doi:[10.4000/corpus.7205](https://doi.org/10.4000/corpus.7205).
- Rainer, Franz 2016. Italian. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-Formation. An International Handbook of the Languages of Europe*, vol. 4, 2712–2731. De Gruyter Mouton. doi:[10.1515/9783110379082-017](https://doi.org/10.1515/9783110379082-017).
- Scalise, Sergio & Antonietta Bisetto 2009. The Classification of Compounds. In Rochelle Lieber & Pavol Štekauer (Eds.), *The Oxford Handbook of Compounding*. Oxford: Oxford University Press. doi:[10.1093/oxfordhb/9780199695720.013.0003](https://doi.org/10.1093/oxfordhb/9780199695720.013.0003).
- Traugott, Elizabeth Closs & Graeme Trousdale. 2013. *Constructionalization and constructional changes* (Oxford linguistics 6). Oxford: Oxford University Press.
- Viale, Matteo. 2008. *Studi e ricerche sul linguaggio amministrativo*. Padova: CLEUP.

A typological comparison of infixes and circumfixes

Tim Zingler
University of Innsbruck
tim.zingler@uibk.ac.at

1. Introduction

While the existence of infixes and circumfixes is acknowledged in many relevant textbooks and handbooks on morphology and/or typology (e.g., Whaley 1997: 117; Hall 2000: 540; Moravcsik 2000: 548), very little is known about their behavior and distribution (cf. also Harris 2010: 100, n. 1). If anything, both types are usually claimed to be cross-linguistically rare, even though no robust evidence in support of this is typically presented. One of the few language families known to have both affix types is Austronesian, and to the extent that the relevant literature provides examples of either type, they tend to come from this family. This situation suggests that a considerable amount of empirical groundwork is required in order to properly understand these two phenomena and to adequately account for them in theories of morphology. The present contribution is intended as a first step toward that goal. It will aggregate the results of two separate studies on the cross-linguistic distribution of infixes and circumfixes, Zingler (2022a, b). Both works are of an inductive nature and as such mostly point to areas that call for further research or that would appear to be of particular theoretical interest. These domains will be highlighted throughout this work.

In Section 2, I will outline some basic facts about infixation and circumfixation and some of the issues and motivations that guided the research underlying this paper. In Section 3, I will juxtapose the main results of the two studies, while Section 4 will be an attempt to find plausible explanations for those patterns. Section 5 is the conclusion, which will sum up the major insights and sketch some crucial desiderata for future work.

2. Infixation and circumfixation: Some basic facts and issues

It should be emphasized at the outset that our relatively limited understanding of infixation only concerns its morphological aspects. That is, the phonological properties of infixes have been studied in cross-linguistic detail and have subsequently been integrated into linguistic theory, particularly Optimality Theory (cf. Yu 2007; Inkelas 2014). Some of the major questions such phonologically oriented works are interested in are which positions of a root an infix can appear in (cf. also Wilson 2014) and under which structural conditions certain infixes might occur outside the root (i.e., as prefixes or suffixes). However, such approaches do not address many other questions that any typologically informed theory of morphology or linguistics will eventually need to answer. First and foremost, phonological accounts obviously pay little attention to the functions that infixes express, and hence the range of meanings that infixes encode has hardly been explored. Put differently, there is much to be gained from a “semasiological” approach to infixation, which essentially takes infixes as the “independent variable” and then explores what kind of functions (i.e., “dependent variables”) this formal type is associated with.

A second major concern is *where* infixes are found, which refers to both their genealogical and their geographical distribution. While Austronesian and, to a lesser extent, Austroasiatic are standardly claimed to be centers of infixation (e.g., Ultan 1975: 172; Naumann & Vogel 2000: 941), there is considerably less agreement concerning their occurrence in the Americas. Specifically, Moravcsik (2000: 548), Akmajian et al. (2001: 19), and Crystal (2008: 243) argue that infixes are relatively widespread in that part of the world, whereas Mithun (1999: 39) explicitly states that they are not common in North America. The overall picture is further complicated by the fact that comparable works on South American languages (e.g., Campbell & Grondona 2012; O'Connor & Muysken 2014) do not make any general claims about the status of infixation. This state of affairs was the impetus for Zingler (2022a), which provides an overview of infixes in American languages on the basis of a convenience sample of about 170 languages. That study primarily relied on reference grammars, with handbooks, textbooks, and journal articles as supplementary sources. The Americas were chosen because the above-mentioned references suggested that the phenomenon is sufficiently common in those two macro-areas to support an in-depth study. Put differently, no other areas with potentially sizable quantities of unknown infixes emerged during the literature review. As a consequence, any claims about the nature and distribution of infixes in this work will be based on that particular study of American infixes, unless explicitly indicated otherwise.

The situation regarding circumfixes is rather different. This structural type has not only failed to attract attention from morphology but has also held little interest for phonology. Hence, circumfixes are even less well explored than infixes, and this crucially impacted the empirical focus of Zingler (2022b). Specifically, the latter work used a worldwide convenience sample of roughly 450 languages to arrive at a first impression of the distribution and behavior of circumfixes. Here, too, I primarily relied on reference grammars and made use of handbooks, textbooks, and journal articles as supplementary sources. Beyond that, the goals of Zingler (2022b) were essentially the same as the ones in Zingler (2022a).

That said, Zingler (2022b) did impose *some* restrictions on the database. One major caveat concerning its sample is that no attempt was made to include a representative number of Austronesian circumfixes. This was due to the fact that this family is extraordinarily large and would thus have had an outsized impact on the database if no such safeguard had been in place. Also, since the goal of typology is to explore cross-linguistic patterns that emerge irrespective of common descent, such an extensive coverage of Austronesian would have undermined the overall purpose of the investigation. Therefore, Austronesian circumfixes were only considered when they expressed a function that was also encoded by non-Austronesian circumfixes, in which case they are part of a larger and independent pattern.

The second major exception to the otherwise inclusive sampling in Zingler (2022b) concerns the Indo-European family. As the discussions in the volumes edited by Müller et al. (2015, 2016) show, many of the Indo-European items that have been or could be described as circumfixes are rather controversial as far as their structural make-up is concerned. As such, an inclusion of (potential) Indo-European circumfixes would often have required a detailed discussion of formal and semantic nuances, which would have gone beyond the goals of Zingler (2022b). The solution to this problem was to entirely exclude Indo-European languages from the sample. This presumably makes Zingler (2022b) one of the few studies with a worldwide sample in which Indo-European languages are underdocumented. Of course, the Indo-European circumfixes are controversial because the relevant languages are well-known, including the analytical challenges they present. In other words, it is likely that many of the non-Indo-European circumfixes in Zingler (2022b) would be seen as equally controversial if the relevant languages were sufficiently understood. As such, that study points toward certain descriptive issues that still require clarification, either on a language-specific or on a general basis.

A perusal of the relevant literature reveals that there are no universally applied definitions of infixation or circumfixation. It is therefore necessary to describe the structural parameters on which the collection of the two databases relied. I will briefly summarize the main criteria here. Details, references, and illustrations can be found in Zingler (2022a: 172-178) and Zingler (2022b: 58-65). With regard to infixes, the major distributional requirement was that they occur *intramorphemically* rather than *intermorphemically*. There is a terminological tradition in which non-peripheral affixes are described as infixes such that the *en-* in English *disenfranchise* would count as an infix. Yet, this is of course very different from the way in which infixation is usually understood. On that more common definition, which also underlies Zingler (2022a), infixes have to split another morph in two. Furthermore, infixes – like all morphs – are linguistic signs, and every item claimed to be an infix therefore has to express a discernible function. This criterion primarily serves to distinguish infixation from epenthesis.

Meanwhile, circumfixes must not be the compositional combination of a prefix and a suffix such that both the preposed and the postposed element independently exist *and* the putative circumfix expresses the combined function of these preposed and postposed items. Like most other criteria mentioned here, this one is often difficult to verify due to the limited information in the sources, but it has to be a general principle. In addition, there are also criteria that apply to both infixes and circumfixes. For instance, in order to be included in the relevant databases, both types have to be instantiated by clearly additive material. That is, phenomena such as root change (including ablaut) do not count as infixation because the allegedly infixed word form is not segmentally longer than the corresponding form without the alleged infix. Similarly, processes in which, for instance, consonant gradation of the word onset expresses a specific function in conjunction with suffixation are not instances of circumfixation because there are no segmental additions at the beginning of the word form in such cases.

Finally, and perhaps most importantly, infixes and circumfixes are affixes and therefore have to be bound to a single word class. This criterion divides into two different issues. On the one hand, it is often impossible to determine on the basis of the evidence available in the sources whether an element really is limited to a single word class (provided that there is even a clear idea of how to define word classes). Put differently, it is somewhat probable that there are more endoclitics and even circumclitics in the languages of the world than currently recognized. On a more general level, however, it can be stated with confidence that both infixes and circumfixes almost exclusively occur with verbs and, to a much lesser extent, with nouns (or with lexemes expressing verbal and nominal concepts). On the other hand, it is unclear how to define “bound” (cf. Haspelmath 2021). In Zingler (2022a, b), this concept was primarily invoked to exclude cases in which full words may be inserted into other words, as in the famous case of English “expletive insertion” (cf. McCarthy 1982). However, this English process seems to be virtually unique, and there were no analogous instances relevant to the classification of circumfixes either. Therefore, the criterion of “boundness” was presumably satisfied to a greater extent across the two studies than any of the other parameters sketched here.

Examples (1), from Mískito, and (2), from Georgian, illustrate items that meet all the above-mentioned criteria for infixes and circumfixes, respectively, to the extent that the relevant sources permit such a conclusion. Both will be discussed below.

- (1) *na<m>pa*
tooth<2SG>
'your tooth'
(Lehmann & Moravcsik 2000: 750)

- (2) *me-cxvar-e*
 AGT₁-sheep-AGT₂
 ‘shepherd’
 (Hewitt 1995: 103)

With respect to (1), the concept of tooth is highly likely to be monomorphemic, which in turn renders it likely that the possessive marker is inserted intramorphemically. Since this marker is assigned the straightforward function of second person possessive, it also qualifies as a full-fledged sign. Furthermore, its sub-syllabic structure suggests that it is indeed a bound unit rather than a potentially free word. Finally, there is no indication in the source that the *m* might be understood as nasalization. With regard to (2), it is unclear how the meaning of agent(ive) could be compositional, and this makes it likely that the element is indeed a single affix, as described in the source. Given this meaning, the item is also a linguistic sign, and there is again nothing to suggest that the preposed and postposed parts are anything but segmental additions. While possessive marking on nouns and argument indexation on verbs often use formally identical items (Siewierska 1998), the two functions are of course distinct, which is why the element in (1) does not violate the principle of attaching to only a single word class. Similarly, the formal structure that instantiates the agentive circumfix in (2) also derives ordinal from cardinal numerals. But since these functions are unrelated, the two structures are not the same sign. Therefore, (2) does not illustrate a violation of word class loyalty either.

3. Data

Zingler (2022a) found 61 infixes from 48 American languages, which belong to 32 families. Meanwhile, Zingler (2022b) found 83 circumfixes from 59 languages, which belong to 33 families. (The genealogical classification is based on *Glottolog*; Hammarström et al. 2022. Each isolate counts as a separate family.) Tables 1 and 2 show which functions the infixes and the circumfixes express, respectively. The classification of the functions is further described below, whereas the potential explanatory value that the different functions have for each of the affix types will be a major focus of the next section.

Table 1: The functions of infixes.

<i>Function</i>	<i>Number of languages</i>
Number (including collective, distributive, pluractional)	18
Tense/aspect/mood (modality)/evidentiality (TAME)	14
Person	9
Voice/valency	8
Miscellaneous	12

Table 2: The functions of circumfixes.

<i>Function</i>	<i>Number of languages</i>
Negation (including privative, irrealis, prohibitive)	20
Nominalization	8
Mood/modality	7
Miscellaneous	48

The two tables reveal several obvious asymmetries. First and foremost, the infixes are a much more semantically coherent set than the circumfixes. The former largely divide into four semantic domains, with the “miscellaneous” category accounting for less than 20% of the data. Yet, even this is deceptive because several of the infixes subsumed under the “miscellaneous”

label express some kind of intensification. Hence, intensification could be regarded as a fifth major function expressed by the infixes, which would further reduce the proportion of infixes with “miscellaneous” functions.

By contrast, the circumfixes are a very heterogeneous class in which the “miscellaneous” group is by far the largest and accounts for more than half the data. In fact, it would be futile to list which functions are included among the “miscellaneous” circumfixes because they subsume most major inflectional and derivational meanings. The main exceptions are from the nominal domain: demonstratives and definiteness markers, as well as noun class and gender markers. The absence of these categories from the database is also essentially the only semantic property that the circumfixes share with the infixes. That is, among the categories that they do express, there is essentially no overlap. The only candidate might be mood/modality, but only a few of the TAME infixes actually express mood and/or modality. Furthermore, the concept of mood/modality is so wide (not to say “ill-defined”) that this degree of convergence appears unremarkable on the surface.

In addition to their semantic patterns, the two affix types also show interesting geographical distributions. While these will not be discussed any further here, they should be mentioned in the interest of completeness. In discussing areal patterns, I will again follow the classification in *Glottolog*. My circumfix sample only contains two items from the macro-area of Australia, which is a conspicuously low number compared to all the other macro-areas. By contrast, the macro-area of Papunesia accounts for 25 of the circumfixes, and circumfixation is relatively common in that part of the world even if Austronesian is factored out. On the other hand, there are no infixes in most of the northernmost families of North America (specifically, Eskimo-Aleut, Iroquoian, and Athabaskan-Eyak-Tlingit) but noticeably many across Mesoamerica.

Of course, the results obtained here may simply be artifacts of the convenience samples on which they rest. However, both studies appear sufficiently comprehensive to suggest that there are indeed principled differences between the distributions of the two types. Crucially, there is nothing to suggest that the two databases are incommensurate because one is based on a worldwide sample and the other one is only based on American languages. (But of course, the baseline expectation is that Austronesian infixes would diversify the range of infixal functions.) The next section will propose certain explanations for the most prominent associations between form and function described above.

4. Analysis

The basic explanations that I will pursue for the patterns in Tables 1 and 2 are of the diachronic kind. Of course, first-hand historical documentation is hardly available for any of the languages in either of the samples, and I will therefore rely on the diachronic pathways established for infixes and circumfixes more generally. In addition, I will supplement these accounts with insights gained from cross-linguistic studies of grammaticalization (or “constructionalization”). Put differently, I will assume that the synchronic functions of infixes and circumfixes are an indicator of their diachronic development and that, more broadly, diachrony can help to make sense of synchronic phenomena (Bybee 2015).

Since the infixes constitute a fairly homogeneous group, explanations for their functional associations are presumably more straightforward than they are in the case of the highly diverse circumfixes. In fact, Yu (2007: ch. 5) provides a detailed account of the historical developments by which infixes arise. He describes three major paths: reduplication mutation, metathesis, and entrapment. The concept of reduplication mutation roughly refers to a process during which an erstwhile peripheral (i.e., “adfixal”) reduplicant is reanalyzed as part of the root, and a part of the base is simultaneously reanalyzed as the reduplicant. As such, the reduplicant “accidentally” lands inside the root. Schematically, this can be thought of as $ab-ABCD \rightarrow AB<ab>CD$, where

lower-case letters mark affixal segments and upper-case letters mark root segments. While Yu (2007: ch. 5) adds many more complexities to this account, which cannot be verified for the sample of Zingler (2022a), what matters for the present semasiological approach is simply that reduplication as such is often the origin of infixes. This is crucial in that the meanings expressed by reduplication are also widely found among the American infixes. In particular, this is true of the iconic functions of reduplication, where a repetition of form expresses a repetition of meaning (cf. Moravcsik 1978: 316-325). Ultimately, the vast majority of number-marking infixes fall into this category, as do most of the TAME infixes since these largely express imperfective aspects such as repetitive, iterative, etc. In addition, the intensifiers that account for several of the “miscellaneous” infixes are also iconic and are thus also plausibly derived from reduplicative constructions. As I discuss in Zingler (2022a: 209-212), many of the infixes expressing these functions are indeed (partly) reduplicative and, equally crucially, none of the other infixal functions are expressed by reduplicative infixes. As such, a major portion of the infix database can straightforwardly be attributed to an iconic origin that leads to infixation via reduplication (mutation).

The process of metathesis is simpler and better-known than reduplication mutation, but also harder to detect in synchronic data. In Zingler (2022a), I argued that the infixes most likely to have emerged by this process are the ones that express voice/valency. This claim is based on Bybee (1985), who shows that functions that are most relevant to the meaning of the verb stem also occur close to the verb stem (which is another instance of iconicity). She argues explicitly that voice and valency are the most relevant categories in the verbal domain (Bybee 1985: 4-5), from which it should follow that voice/valency affixes tend to be immediately adjacent to the stem. While she does not investigate this particular prediction in detail, her claims are borne out for all the categories that she does explore. Other amenable generalizations in this context are that metathesis typically affects adjacent and short units and that affixes are typically short. Hence, voice/valency affixes should meet both the structural and the distributional prerequisites for undergoing metathesis and for thus ending up inside the neighboring root. Of course, the factors that give rise to metathesis are primarily phonological, but these cannot be covered here.

The last process that Yu (2007) discusses is “entrapment.” This refers to another kind of reanalysis in which an erstwhile adfix occurs between two elements that ultimately cumulate to become a single morph such that the former adfix develops into an infix due to the semantic changes taking place around it. Yu (2007: 148) limits this process to cases in which one surrounding morph is an affix and the other one is a root, but there do not seem to be compelling reasons to exclude scenarios in which both surrounding items are stems. For a prefixal origin, this could thus be schematically represented as *Stem 1* – [*Prefix* – *Stem 2*] → [*Stem 1* – *Infix* – *Stem 1*]. It follows that entrapment is a process that may take place when affixes are not “externalized” (Haspelmath 1993). As before, the sparse synchronic data do not permit specific entrapment analyses for any of the infixes. However, Bybee’s (1985) account is yet again suggestive at a more general level. That is, she argues that person markers are least relevant to the meaning of a verb and are thus also furthest removed from the verb stem and generally less tightly fused to it. It seems to be a defensible assumption that a lower degree of formal bonding with the stem is a necessary condition for dissociating from that stem and for assuming a more neutral position between the two parts of the new stem. If so, person markers would be the most likely category of the ones in Table 1 to develop infixal exponents via entrapment. In that case, all the major functional domains in Table 1 would be accounted for by one of Yu’s (2007) diachronic scenarios.

While this account of the functional patterns shown by infixes is perforce speculative, it should also be mentioned that there are a few items in the database that are not captured by any of the diachronic trajectories (for instance, a stem formation marker in Central Sierra Miwok).

Since the circumfixes show a much wider range of functions than the infixes, they would also seem to require a much more expansive explanation. Yet, such an explanation is complicated by the lack of previous work on the topic. Therefore, the following remarks are based on more general principles of language change and will have to be supplemented or replaced by more detailed diachronic work. Two of the very few works that make substantive claims about (the emergence of) circumfixes are Greenberg (1980) and Harris (2010). Both are essentially in agreement that circumfixes arise from the reanalysis of an originally independent prefix and an originally independent suffix as a single affix. There is nothing in my data that would argue against this scenario, and this account has the additional merit that it would explain why circumfixes have a much wider range of functions than infixes to begin with. Simply put, there are essentially no semantic constraints on prefixes or suffixes, and hence there also should not be any semantic constraints on circumfixes that arise from prefixes and suffixes.

That said, there is an obvious link between circumfixation and negation. Of the 20 negation markers in the database, 13 express “standard negation” in the sense of Miestamo (2005), whereas the remaining seven items express privative, irrealis, or prohibitive functions. Hence, the latter primarily differ from standard negation in that they do not scope over a declarative predicate, but the decision to subsume irrealis under negation is more controversial (cf. Zingler 2022b: 71 for discussion). While I do not know of any concrete evidence for how negation circumfixes arise in a language, certain diachronic patterns nevertheless permit a relatively plausible explanation. The crucial ingredient of such an account is arguably the negative cycle, during which a negator loses its pragmatic force and comes to be supported by a second negator, which eventually becomes the main marker of negation. This process is well-documented for French but is also observable in many other languages (van Gelderen 2008). Presumably, then, if the two parts of the negative construction are found on opposite sides of the clausal head, and the stage at which they co-occur is sufficiently long for both to morphologize, a negation circumfix may emerge. The fact that negation circumfixes are rare overall therefore seems to confirm the intuition that these diachronic developments, or at least their combination, are also rare. In this context, it is also important to highlight that negators show a tendency to be preposed. As such, they meet a necessary requirement for the emergence of circumfixes that the exponents of most other categories satisfy to a much lesser degree. Bybee (1985: 177) even finds a slight prefixation preference among the negation affixes in her worldwide sample. It is this prefixing bias that gives negators a head start in the “circumfix race” compared to the markers of virtually any other category. (For a functional explanation as to why negators tend to occur in preposed position, cf. Dryer 1988: 102; Berg 2020: 386.)

The second major set among the circumfixes consists of nominalizers. This group is arguably in even greater need of explanation given the cross-linguistic fact that nouns show considerably fewer prefixes than verbs (Cutler et al. 1985: 730; Seifart et al. 2018: 5723). Hence, nouns violate a crucial requirement for the development of circumfixes. Yet, while any diachronic explanation of this pattern will also have to be speculative, one angle seems somewhat promising. In my sample, most of the stems that form the input for nominalization via circumfixes are verbal. The unexpected clustering of circumfixes in the domain of nominalization might thus find an explanation if it could be shown that the nominalization circumfixes originally encoded verbal categories and only became nominalizers via semantic change. There is some tentative evidence for that scenario. While Cristofaro (2019) shows that nominalizers mostly derive from nouns with general meanings such as ‘person,’ ‘thing,’ or ‘place,’ the grammaticalization literature also reveals that nominalizers seem to emerge in the context of participles (cf. Kuteva et al. 2019: 334, 434-4). Hence, if a single verbal stem is separately marked for both nominalization and non-finite status, there are two items that could eventually be reanalyzed as a single nominalizer. That these categories are likely to co-occur also follows from their semantic compatibility, specifically, from the fact that participles and

nominalizers assume intermediate positions on the noun-verb cline (cf. Hopper & Thompson 1984: 704; Comrie & Thompson 2007: 346). Crucially, this account leaves open the possibility that the prefixal part of the eventual circumfixes started out as a verbal marker (i.e., non-finiteness), which would be the unmarked case. As before, the fact that this scenario appears highly specific and thus unlikely is not a problem because the outcome of that scenario is indeed very infrequent.

There is also another conceivable diachronic path for nominalization circumfixes, though support for this claim comes from only one family in my data. The Tacanan branch of the Pano-Tacanan family has several different nominalizing circumfixes with *e-* as the first part and a variety of items as the second part (cf. Guillaume 2008: 435 for Cavineña; Vuillermet 2012: 313 for Ese Ejja). Crucially, the initial *e-* is homophonous with a former noun classifier that has since lost its function. It therefore stands to reason that this classifier was only “accidentally” present in the nouns originally formed by nominalizing suffixes and that it was absorbed by these suffixes because it was no longer associated with a function. On that reading, then, these circumfixes arose from the combination of a suffix and a meaningless element.

Any explanation of the remaining categories will have to be even less comprehensive. For instance, Bybee et al. (1990: 29-34) show that aspect markers are much more frequently preposed than tense markers. That arrangement might pave the way for a process in which a perfective prefix and a past tense suffix coalesce into a single circumfix, given their functional near-equivalence. However, the TAME category most widely found among the circumfixes is mood/modality rather than tense/aspect, and no analogous developments suggest themselves in the former domain. Among the other functional clusters in the data is a small set of adjectival degree markers. Greenberg (1966 [2005]: 40) suggests that the Hungarian superlative circumfix arose via the further affixation of the comparative form. Such a development might be further facilitated if the comparative and superlative function are only weakly distinguished (cf. Hewitt 1995: 49; Cuzzolin & Lehmann 2004: 1215 for Kartvelian), in which case the second affix might primarily provide clarification.

Overall, then, the emergence of circumfixation is still largely a mystery, but some advances toward an understanding of this phenomenon can be made once independently established cross-linguistic phenomena such as the negative cycle or the suffixing preference are taken into account. Further progress on this topic will mostly depend on the work of specialists on the relevant languages.

5. Conclusion

Among the ideas espoused here is the claim that the functional patterns of the two affixation strategies can usually be explained with reference to their diachrony. Another conclusion is that infixation seems to be more widely distributed than circumfixation. This claim is based on the fact that I have data from roughly the same number of language families for both types, even though the infixes are only from the Americas. Hence, infixes seem to arise more easily than circumfixes, which would be plausible in that the emergence of infixes almost exclusively depends on formal mechanisms, whereas both formal and functional criteria have to be met in order for circumfixes to arise. On the other hand, there are several language families that make rather extensive use of circumfixes (Chukotko-Kamchatkan, Kartvelian, Austronesian), while infixes seem to be highly sporadic in every American language (family) they occur in. Hence, if infixes are indeed a “wider” property, circumfixes in turn might be a “deeper” feature and might even have to be reconstructed for some proto-languages.

Above all else, this work has shown that infixes and circumfixes are not “rare” in any absolute sense. There are dozens of language families that have at least one of the two types and many that have both. Instead, the relative infrequency of these types compared to prefixes

and suffixes is straightforwardly tied to their diachronic emergence (cf. Harris 2008, 2010). In conjunction, then, both the synchronic and the diachronic properties of infixes and circumfixes call for much further work. Put more bluntly, our current state of knowledge regarding these two types is somewhat of an embarrassment for morphological theory.

Desiderata for future research are as obvious as they are variegated. Each of the diachronic explanations for the circumfixes suggested here would ideally be supported by concrete evidence from at least a few languages. With regard to both affix types, synchronic explanations would also have to be pursued, especially in light of psycholinguistic arguments of why infixes *should* be rare (Cutler et al. 1985: 751-752). In addition to the obvious need for more data (especially from Austronesian languages), there are also many analytical issues that will require meticulous research. For instance, several of the circumfixes in my database appear to be at least partly compositional, whereas several of the infixes might actually be glottalization features rather than segmental additions. As described in Section 2, these issues are essential for the identification of the relevant types, and my databases and those similar to mine might have to be considerably revised once the relevant semantic and phonetic issues are properly understood. In addition, more comprehensive morphosyntactic analyses will have to determine to what extent the items considered infixes and circumfixes here can occur with different word classes and might thus be better conceived of as endoclitics and circumclitics, respectively.

References

- Akmajian, Adrian, Richard Demers, Ann Farmer & Robert Harnish. 2001. *Linguistics*. 5th edn. Cambridge, MA: MIT Press.
- Berg, Thomas. 2020. Ordering biases in cross-linguistic perspective: The interaction of serial order and structural level. *Linguistic Typology* 24(2). 353-397.
- Booij, Geert, Christian Lehmann & Joachim Mugdan (eds.). 2000. *Morphology*, vol. 1. Berlin: De Gruyter.
- Bybee, Joan. 1985. *Morphology*. Amsterdam: Benjamins.
- Bybee, Joan. 2015. *Language change*. Cambridge: Cambridge University Press.
- Bybee, Joan, William Pagliuca & Revere Perkins. 1990. On the asymmetries in the affixation of grammatical material. In William Croft, Keith Denning & Suzanne Kemmer (eds.), *Studies in diachronic typology for Joseph H. Greenberg*, 1-42. Amsterdam: Benjamins.
- Campbell, Lyle & Verónica Grondona. 2012. *The indigenous languages of South America*. Berlin: De Gruyter.
- Comrie, Bernard & Sandra Thompson. 2007. Lexical nominalization. In Timothy Shopen (ed.), *Language typology and syntactic description*, vol. 3, 334-381. Cambridge: Cambridge University Press.
- Cristofaro, Sonia. 2019. Nominalization in cross-linguistic diachronic perspective. In Robert Zariquiey, Masayoshi Shibatani & David Fleck (eds.), *Nominalization in languages of the Americas*, 169-194. Amsterdam: Benjamins.
- Crystal, David. 2008. *A dictionary of linguistics and phonetics*. 6th edn. Malden: Blackwell.
- Cutler, Anne, John Hawkins & Gary Gilligan. 1985. The suffixing preference: A processing explanation. *Linguistics* 23(5). 723-758.
- Cuzzolin, Pierluigi & Lehmann, Christian. 2004. Comparison and gradation. In Geert Booij, Christian Lehmann, Joachim Mugdan & Stavros Skopeteas (eds.), *Morphology*, vol. 2, 1212-1220. Berlin: De Gruyter.
- Dryer, Matthew. 1988. Universals of negative position. In Michael Hammond, Edith Moravcsik & Jessica Wirth (eds.), *Studies in syntactic typology*, 93-124. Amsterdam: Benjamins.
- Gelderen, Elly van. 2008. Negative cycles. *Linguistic Typology* 12(2). 195-243.
- Greenberg, Joseph. 1966 [2005]. *Language universals*. Berlin: De Gruyter.
- Greenberg, Joseph. 1980. Circumfixes and typological change. In Elizabeth Traugott, Rebecca Labrum & Susan Shepherd (eds.), *Papers from the fourth international conference on historical linguistics*, 233-241. Amsterdam: Benjamins.
- Guillaume, Antoine. 2008. *A grammar of Cavineña*. Berlin: De Gruyter.
- Hall, Christopher. 2000. Prefixation, suffixation and circumfixation. In Booij et al. (eds.), 535-545. Berlin: De Gruyter.
- Hammarström, Harald, Robert Forkel, Martin Haspelmath & Sebastian Bank. 2022. *Glottolog 4.6*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://doi.org/10.5281/zenodo.6578297>. Available online at <http://glottolog.org> (last access 10-25-2022).

- Harris, Alice. 2008. Explaining exuberant agreement. In Þórhallur Eypórssón (ed.), *Grammatical change and linguistic theory*, 265-283. Amsterdam: Benjamins.
- Harris, Alice. 2010. Explaining typologically unusual structures: The role of probability. In Jan Wohlgemuth & Michael Cysouw (eds.), *Rethinking universals*, 91-103. Berlin: De Gruyter.
- Haspelmath, Martin. 1993. The diachronic externalization of inflection. *Linguistics* 31(2). 279-309.
- Haspelmath, Martin. 2021. Bound forms, welded forms, and affixes: Basic concepts for morphological comparison. *Voprosy Jazykoznanija* 1(1). 7-28.
- Hewitt, B. G. 1995. *Georgian*. Amsterdam: Benjamins.
- Hopper, Paul & Sandra Thompson. 1984. The discourse basis for lexical categories in Universal Grammar. *Language* 60(4). 703-752.
- Inkelas, Sharon. 2014. *The interplay of morphology and phonology*. Oxford: Oxford University Press.
- Kuteva, Tania, Bernd Heine, Bo Hong, Haiping Long, Heiko Narrog & Seongha Rhee. 2019. *World lexicon of grammaticalization*. 2nd edn. Cambridge: Cambridge University Press.
- Lehmann, Christian & Edith Moravcsik. 2000. Noun. In Booij et al. (eds.), 732-57.
- McCarthy, John. 1982. Prosodic structure and expletive infixation. *Language* 58(3). 574-590.
- Miestamo, Matti. 2005. *Standard negation*. Berlin: De Gruyter.
- Mithun, Marianne. 1999. *The languages of Native North America*. Cambridge: Cambridge University Press.
- Moravcsik, Edith. 1978. Reduplicative constructions. In Joseph Greenberg, Charles Ferguson, Edith Moravcsik (eds.), *Universals of human language, vol. 3: Word structure*, 297-334. Stanford: Stanford University Press.
- Moravcsik, Edith. 2000. Infixation. In Booij et al. (eds.), 545-552.
- Müller, Peter, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.). 2015. *Word-Formation*. Vols. 1-3. Berlin: De Gruyter.
- Müller, Peter, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.). 2016. *Word-Formation*. Vols. 4-5. Berlin: De Gruyter.
- Naumann, Bernd & Petra Vogel. 2000. Derivation. In Booij et al. (eds.), 929-943.
- O'Connor, Loretta & Pieter Muysken (eds.). 2014. *The native languages of South America*. Cambridge: Cambridge University Press.
- Seifart, Frank, Jan Strunk, Swintha Danielsen, Iren Hartmann, Brigitte Pakendorf, Søren Wichmann, Alena Witzlack-Makarevich, Nivja de Jong & Balthasar Bickel. 2018. Nouns slow down speech across structurally and culturally diverse languages. *National Academy of the Sciences of the United States of America (PNAS)* 115(22), 5720-5726.
- Siewierska, Anna. 1998. On nominal and verbal person marking. *Linguistic Typology* 2(1). 1-55.
- Ullman, Russell. 1975. Infixes and their origins. In Hansjakob Seiler (ed.), *Linguistic Workshop III*, 157-205. Munich: Fink.
- van Gelderen, Elly (see Gelderen).
- Vuillermet, Marine. 2012. *A grammar of Ese Ejja, a Takanan language of the Bolivian Amazon*. Lyon: Université Lumière Lyon 2 dissertation.
- Whaley, Lindsay. 1997. *An introduction to typology*. Thousand Oaks: SAGE.
- Wilson, Jennifer. 2014. Evidence for infixation after the first syllable: Data from a Papuan language. *Phonology* 31(3). 511-23.
- Yu, Alan. 2007. *A natural history of infixation*. Oxford: Oxford University Press.
- Zingler, Tim. 2022a. Infixation in the Americas: A cross-linguistic survey. *International Journal of American Linguistics* 88(2). 171-228.
- Zingler, Tim. 2022b. Circumfixation: A semasiological approach. *Word Structure* 15(1). 55-113.

