Proceedings of the LFG'21 Conference

On-Line

Miriam Butt, Jamie Y. Findlay, Ida Toivonen (Editors)

2021

CSLI Publications

http://csli-publications.stanford.edu/LFG/2021

Contents

Editor's Note

For this year's edition of the LFG Proceedings, we welcome Jamie Findlay to the team of editors.

The 2021 Conference on Lexical Functional Grammar was held on-line. The program committee for LFG21 were Tina Bögel and Agnieszka Patejuk. We would like to thank them for coordinating the review process and working with the conference organizers to put together this year's on-line program. The conference was originally scheduled to take place at the University of Oslo in Norway, with Helge Lødrup and Dag Haug as organizers. Due to the pandemic, the executive committee decided to move the conference on-line for the second year in a row. Stephen Jones, Joey Lovestrand, Kengatharaiyer Sarveswaran, Péter Szűcs and Fengrong Yang then took on the challenging task of organizing the on-line version of the conference. We would like to thank them and David Diem, who developed and maintained the LFG21 website, for their time and engagement!

Extended abstracts, handouts and videos of talks were able to be uploaded before the conference, along with an open commenting function. The synchronous part of the conference was held mainly in the form of QA sessions on the talks via Zoom. Poster sessions and social gatherings were facilitated via Discord. This format worked out very well and we would like to thank the ad-hoc committee for an outstandingly well organized conference that worked well and smoothly.

As usual, we would also like to thank the executive committee and the abstract and final paper reviewers, without whose prompt and thorough work the conference and the proceedings would not have been possible in this form.

The table of contents lists all the papers presented at the conference. Some papers were not submitted to the proceedings. For these papers, we suggest contacting the authors directly. We note that all of the abstracts were peer-reviewed anonymously (double-blind reviewing) and that all of the papers submitted to the proceedings underwent an additional round of reviewing. We would like to express our heartfelt thanks to all of the anonymous reviewers for the donation of their expertise and effort in what is often a very short turn-around time.

As we were putting together the proceedings, news reached us that our friend andd colleague Jürgen Wedekind had passed away suddenly at the end of December. On the behalf of the LFG community, we would like to here express our heartfelt grief at losing him too early.

Hard Copy: All of the papers submitted to the LFG21 proceedings are available in one large pdf file. The proceedings' file was created via pdflatex tools and with the help of scripts written originally by Tracy Holloway King and Stefan Müller. We thank Dikran Karagueuzian at CSLI Publications for his continuous support of our proceedings and our community.

Rethinking lexical integrity: Phrase-level and word-level case morphology

Oleg Belyaev
Lomonosov Moscow State University/Institute of Linguistics RAS/
Pushkin State Russian Language Institute

Proceedings of the LFG'21 Conference

On-Line

Miriam Butt, Jamie Y. Findlay, Ida Toivonen (Editors)

2021

CSLI Publications

pages 5–25

http://csli-publications.stanford.edu/LFG/2021

Keywords: L(R)FG, lexical integrity, morphological case

Belyaev, Oleg. 2021. Rethinking lexical integrity: Phrase-level and word-level case morphology. In Butt, Miriam, Findlay, Jamie Y., & Toivonen, Ida (Eds.), *Proceedings of the LFG'21 Conference, On-Line*, 5–25. Stanford, CA: CSLI Publications.

Abstract

In this paper, I provide a typological argument in favour of preserving lexical integrity in LFG, based on the behaviour of case markers in languages of the world. I demonstrate that case systems that conform to the definition of morphological case (m-case) as proposed in work by Otoguro and Spencer cannot have phrasal scope; conversely, only m-cases may trigger NP-internal concord. I interpret these findings as pointing to a principal distinction between morphology and syntax, with the domain of morphology limited compared to the traditional view: only features showing complex paradigmatic behaviour are truly morphological. I further evaluate three possible ways to account for this distinction in modern LFG (standard LFG, lexical sharing, L(R)FG), and conclude that, at present, none are fully acceptable.

1 Introduction

Lexical integrity has been a hallmark of LFG since its inception. The concept itself, however, is far from having a universally accepted definition, but there are two main formulations that are frequently used in the literature:

(1) Words are built out of different structural elements and by different principles of composition than syntactic phrases. (Bresnan and Mchombo 1995, 181)

(2) Morphologically complete words are leaves of the c[onstituent]-structure tree and each leaf corresponds to one and only one c[onstituent]-structure node. (Bresnan et al. 2016, 92)

The definition in (1) is rather broad and is compatible with a wide array of approaches, as long as *some* boundary between morphology and syntax is maintained. (2) is more specific in that it constrains possible analyses in a particular way: namely, it disallows empty nodes, terminal nodes occupied by affixes or features, and words mapping to more than one preterminal (category) node. However, the notion *morphologically complete word* is treated as a theoretical primitive; it is not clear which criteria can consistently distinguish between words and bound morphemes in a cross-linguistically uniform way. There has been surprisingly littled discussion of this problem in LFG. Bresnan and Mchombo (1995) provide a number of diagnostics for lexical integrity (LI), namely extraction, conjoinability, gapping, inbound anaphoric islands, and phrasal recursivity. However, all

[†]I am grateful to the audience of LFG2021, especially Ash Asudeh, Alessandro Jaker, Adam Przepiórkowski and Daniel Siddiqi, for the feedback, and to two anonymous referees for their comments. Special thanks are due to the editors of this volume, Miriam Butt, Jamie Findlay and Ida Toivonen. All errors are mine.
This research was performed according to the Development Program of the Interdisciplinary Scientific and Educational School of Moscow University "Preservation of the World Cultural and Historical Heritage".

these criteria are problematic because, as shown by Haspelmath (2011), there is no single criterion or set of criteria that can capture linguists' actual use of the term 'word'; the continuum between words and phrases does not seem to show any consistent clustering either. Even within one language, elements defined as words according to some criteria may fail to meet other criteria.

One example of how lexical integrity can be problematic is the phenomenon of so-called phrasal or suspended affixation, such as that found in Turkish:

(3) Turkish (Turkic < Altaic)

[*Almanya*	*ve*	*Amerika*]	***-dan***
Germany	and	America	-ABL

'from Germany and America' (Kabak 2007, 335)

Assuming lexical integrity, the existence of such phenomena leads to a contradiction. Case and number affixes certainly pass all criteria for affixhood in Turkish: they obey vowel harmony and receive stress, unlike clitics, some of which do follow harmony but which are all unstressable (Göksel and Kerslake 2005, 100). Therefore, they should be treated within the morphological domain according to the principle (1), i.e. that rules for assembling words are different from rules for assembling syntactic phrases. But examples like (3) show that case affixes may scope over coordinate phrases, attaching to their rightmost word. This suggests that their behaviour is more akin to that of clitics than affixes, i.e. (3) can be analysed as a Case head that has a coordinate NP as its complement. Conjoinability has been explicitly listed in Bresnan and Mchombo (1995) as a criterion for distinguishing syntactic phrases from word parts. Hence, either case markers should arbitrarily be considered to be clitics – thereby blurring the distinction between morphology and syntax – or the lexical integrity principle should be abandoned or at least relaxed, admitting such notions as "phrasal affixation" or "group inflection". Both conclusions severely weaken the notion of lexical integrity.

Bruening (2018) lists a number of other counterexamples to lexicalism involving phrasal syntax feeding word formation, i.e. words formed from syntactic phrases, such as *a ne'er-do-well* or *a shoot-'em-up* in English; and phrasal syntax having access to sub-word units, such as coordination of word parts, of which (3) is the most clear example, but which is also found in English, as in *pro-choice and -gun control* (Chaves 2008, 263).

Such contradictions may indicate that notions like "word" or "affix" are indeed theoretically problematic: if wordhood criteria do not serve as reliable predictors of any syntactic behaviour, a strict separation between morphology and syntax seems unnecessary and arbitrary. In the context of LFG, this is in fact perfectly possible: nothing in the framework hinges specifically on lexical integrity. And, indeed, at least two such attempts have been made. Lexical sharing (Wescoat 2002; Broadwell 2008; Lowe 2016) allows one violation of lexical integrity as understood in (2): a single morphological word may be associated with two adjacent heads. All other principles of lexicalism are preserved; importantly, the sharing pattern itself is defined in the lexicon, and so, the basic division between morphology and syntax is supposedly retained.[1] A more radical option is a new variant of LFG called L(R)FG, for "Lexical (Realizational) Functional Grammar" (Melchin, Asudeh, and Siddiqi 2020), essentially a hybrid of DM and LFG. In this approach, lexical entries represent morphemes that are mapped directly to terminal nodes of the c-structure tree. Like in DM, morphology is only in the mapping between f-descriptions (that are found in terminal nodes) and the lexicon; there is no lexical morphological component.

Therefore, lexical integrity for LFG is, primarily, an empirical question: if it can be demonstrated that some definition of morphology predicts an impenetrability to syntactic processes, lexical integrity can be preserved. It is now clear that the traditional assumptions of wordhood and affixhood do not translate to consistent syntactic predictions either cross-linguistically or language-internally. However, wordhood and bondedness do not have to play a central role in the morphology–syntax distinction. After all, modern morphology is not so much about morphemes (cf. Anderson 1992); neither is it much concerned with the definition of wordhood. Rather, morphological theory mainly works with paradigms and relations between their elements; the validity of its results is hardly dependent on our definitions of words and morphemes, or lack thereof. For instance, the results of such studies as Baerman, Brown, and Corbett (2005) on syncretism and Corbett (2007) on suppletion hold regardless of which diagnostics for wordhood are valid in the languages included in the sample.

The aim of this paper is to test whether morphological complexity – broadly understood as in Baerman, Brown, and Corbett (2017), i.e. as the existence of intra-paradigmatic relationships that go beyond concatenation – can serve as a better predictor for LI-consistent behaviour than bondedness in terms of word- or affixhood. In other words, if it can be shown that certain patterns of morphosyntactic expression (those that require reference to the notion *paradigm*) predict syntactic impenetrability (e.g. the diagnostics described in Bresnan and Mchombo 1995), LI can be maintained as a useful principle of grammar. However, its scope will have to be strongly restricted.

Of course, this hypothesis is difficult to test in its entirety because, as it stands, its formulation is too general; furthermore, its scope covers all kinds of morphology (inflectional and derivational) which are clearly outside the scope of a single study. Therefore, in this paper I focus on one particular morphological phenomenon that is relatively well-understood and well-represented in grammars: case systems. My point of departure is the notion of *morphological case* (m-case) as formulated in Spencer and Otoguro (2005), Otoguro (2006), and Spencer (2005). Spencer and Otoguro claim that the *morphological* feature CASE should only be defined for languages where "case" marking (i.e. any kind of nominal dependent

[1]How lexical sharing should be integrated in the morphological module is another question. One possibility is described in Belyaev (2021) for PFM (Stump 2001), based on the morphology–syntax interface model in Dalrymple (2015).

marking – *flagging* in terms of Haspelmath 2019) involves certain kinds of morphological complexities. For other languages, at best, only a syntactic feature CASE should be used.

In Belyaev (2018), I hypothesized that it is only those case systems that obey the definition of m-case as per Spencer and Otoguro which necessarily obey LI. Other "case systems", regardless of their description in grammars as affixes or clitics, may behave as separate syntactic heads scoping over noun phrases. This is, in effect, an implicational universal M-CASE $\longrightarrow$ $\neg$GROUP, where GROUP is the ability to mark the edge of a noun phrase. Conversely, my second hypothesis is that it is only m-case systems that can display NP-internal concord.[2] Based on a pilot sample of 107 languages, both hypotheses are confirmed, although the latter less strongly so because of low occurence of case concord in the sample in the first place. From this typological observation, I argue that any approach that involves a clear boundary between "lexical" morphology and syntax (such as traditional LFG or LFG with lexical sharing) is preferable to an approach that collapses the boundary between morphology and syntax in its entirety.

The paper is organized as follows. In section 2, I describe the approach of Spencer and Otoguro, the notion of m-case and how it can be used as the basis for a typological treatment of lexical integrity. In section 3, I present the result of a preliminary typological study that defines m-case as a comparative concept and confirms two putative universals that connect m-case status with the lack of group affixation and the possibility of NP-internal concord. In 4, I discuss the implications of these findings for LFG.

2 Case systems

Spencer and Otoguro (2005), Otoguro (2006), and Spencer (2005) based their analysis of case systems on Beard (1995), who proposed that case systems should only be stipulated for those languages where the morphology is complex enough to warrant a *morphological* feature CASE. In Spencer and Otoguro's interpretation, this criterion, which they call BEARD'S CRITERION, is that morphological case (M-CASE) should only be postulated if the connection between syntactic case features/functions and their formal exponents is more complex than just a one-to-one mapping. Examples from Otoguro (2006) are particularly illustrative.

[2]In fact, case concord is treated as one of the *criteria* for m-case status in Otoguro (2006). However, concord in a syntactic case feature is perfectly conceivable – for example, preposition concord, although mainly optional and restricted, was found in Old Russian (see Klenin 1989). Therefore, I treat m-case status and concord as independent variables.

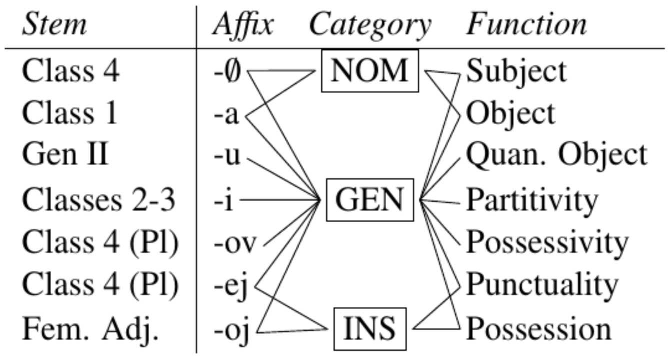


Figure 1: Russian case system according to Otoguro (2006)

A system like Russian (Figure 1) clearly requires reference to a morphological feature CASE. Indeed, no direct mapping between syntactic function and morphological exponence can be established: the latter is dependent on number (due to consistent case-number cumulation) and inflection class. For example, the suffix *-a* can be associated with two feature sets, which, in turn, are associated with different syntactic functions: genitive singular (in the *-a* inflection class) and nominative plural (in neuter nouns of the consonant-final inflection class). It is impossible to assign *-a* a single set of syntactic features or functions for all contexts; which of the two sets is used depends on the inflection class of the head noun.

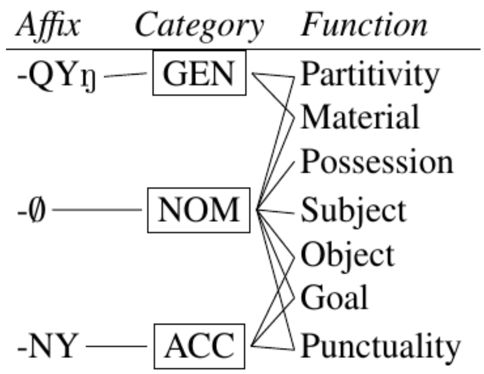


Figure 2: Bashkir case system according to Otoguro (2006), variant 1

Bashkir, like other Turkic languages, is different. In this language, the mapping between syntactic function and affix exponence is always one-to-one; what variation there is is explicable from morphonology. Hence, while it is possible to provide a "Russian-like" mapping, as in Figure 2, it seems more economical to assume that affixes are directly associated with specific syntactic functions, as in Figure 3. Thus, instead of "genitive" or "accusative", Bashkir "cases" can be referred to as the "-*QYŋ*-form", "-*NY*-form", etc. This makes such case markers not much different from adpositions – even though they are affixes from the point of view of Bashkir grammar. Bashkir may still require a *syntactic* case feature; the point that Otoguro makes is that a case feature is not required for a *morphological* description of Bashkir.

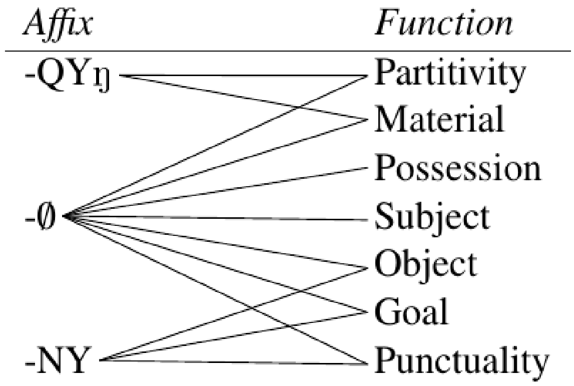


Figure 3: Bashkir case system according to Otoguro (2006), variant 2

Spencer and Otoguro's observations are very valuable, but they are mainly concerned with morphological theory; they do not claim that m-case should correlate with any syntactic behaviour. Moreover, they start from the assumption that all exponents involved are affixes (since adpositions or other kinds of syntactic case markers cannot, by definition, introduce m-case features); the differences are in the morphological features they realize. What I propose in this paper is to essentially reverse the argument, taking m-case status as a starting point and seeing whether it predicts syntactic behaviour consistent with LI. If this is true, pre-syntactic (lexical) morphology should be retained in the theory, but its domain, at least as far as case is concerned, should be limited: only m-case should be treated in the morphology. Other "case" markers can be dealt with in the syntax. This would follow the standard LFG division of labour between morphology and syntax, confirming its cross-linguistic validity; analyses of individual "case" markers, however, may have to be reconsidered in light of these findings.

3 Typology

This typology mainly repeats the finding earlier reported in Belyaev (2018), with certain minor additions and modifications.

3.1 Formulating the concept

One of the key components of a typological study is providing clear definitions of the parameters involved. Statements in descriptive grammars characterizing markers as "cases" or "adpositions" cannot be taken at face value: the set of criteria that the authors had in mind is often vague and is usually based on the traditional idea of wordhood or bondedness, as opposed to the paradigm-based notion of m-case. Directly applying Beard's Criterion is not an option either:[3] This requires a detailed morphosyntactic analysis of a language's case system, such as the ones in Otoguro (2006), which is not feasible for any sample of a substantial size. Therefore, Beard's Criterion should be reformulated as a comparative concept (in terms of Haspelmath 2010) that is applicable cross-linguistically and testable based on data that are easilly obtainable from published sources. To this end, I will rely on three criteria that, if observed in a case system, unambiguously classify it as an m-case system and are sufficiently well-defined in prior typological work:

syncretism (Sync) "a single inflected form [corresponding] to more than one morphosyntactic description" (Spencer 1991, 45);

cumulative exponence (Cumul) encoding of more than one grammatical feature, or a lexical meaning together with a grammatical feature, by a single exponent (Bickel and Nichols 2013);[4]

inflection classes (Infl) lexically conditioned variation in case exponence.

I assume that, if a case system demonstrates **at least one** of those, it is an m-case system. Thus:

(4) Beard $\equiv$ Sync $\vee$ Cumul $\vee$ Infl

Importantly, the definitions should be independent of affix/word status, because the goal here is to replace traditional notions of wordhood and affixhood, rather than augment them. Therefore, unlike Baerman, Brown, and Corbett (2005), I include any **system of basic NP flags** (i.e. markers that can attach to NPs lacking other dependent marking, see Haspelmath 2019) in the sample. Thus, in Russian, case+number affixes like *-om* in (5a) will be considered. In Japanese, I will consider "case" clitics such as genitive *no* and dative *ni* (5b), although they are not affixes according to most descriptions of Japanese.

[3] An anonymous reviewer wonders why Beard's Criterion cannot be applied directly if it simply means "complex morphology indicating case". But the notion of "complex morphology" depends on the analysis of the language in question. For example, seemingly cumulative exponence of inflectional features may be due to regular phonological processes erasing the boundary between two morphemes in particular environments. Similarly, lexical variation should be described in terms of inflection classes only if it does not follow from regular phonological rules.

[4] An anonymous reviewer suggests that cumulative exponence implicitly relies on a morpheme-based view of morphology. I am not sure that the notion is incompatible with all word-and-paradigm approaches, however. For example, in PFM (Stump 2001), realization rules targeting more than one feature may be viewed as involving cumulative exponence (although the "exponents" themselves do not exist as theoretical objects as such).

(5) a. Russian (Slavic < Indo-European)

[$_{\text{PP}}$ *nad* [$_{\text{NP}}$ *dom-**om***]]
above house

b. Japanese (Japonic < Altaic)

[[[[*ie* $_{\text{NP}}$] ***no*** $_{\text{KP}}$] *ue* $_{\text{NP}}$] ***ni*** $_{\text{KP}}$]
house GEN above DAT

'above the house'

3.2 Syntactic parameters

3.2.1 The hypothesis

As stated above, I test two hypotheses on the correlation between the morphological status of case and its syntactic expression. One is that m-case status is incompatible with group marking; that is, Group $\rightarrow \neg$ m-case. The other is that case concord is only compatible with m-case status: Concord $\longrightarrow$ m-case. In the former case, I assume that group marking is handled via locating the affix in a higher projection like KP, as in Broadwell (2008), or as an adjunct to NP, as in Spencer (2005) and Belyaev (2021), which scopes over both conjuncts. This, by definition, is incompatible with the notion of case as a lexically expressed, morphological feature, which m-case is supposed to represent. The latter hypothesis is less obvious; my assumption is that concord is only possible in grammatical *features*, not in *form*;[5] an adjective may agree with its head in a genitive case feature, but not in "*-Qyŋ*-form" or in the preposition *of*.

Both parameters represent facts that are usually reflected in descriptive grammars in one form or another. However, what exactly counts as group marking or concord is a non-trivial question. In the following section, I will provide empirical definitions of both that can be unambiguously identified in languages.

3.2.2 Group marking

I assume that group marking occurs whenever a case marker (flag) occurs at the edge of NP rather than at its head. Prenominal markers in head-final languages and postnominal markers in head-initial languages are thus uncontroversial. For example, English prepositions uncontroversially mark phrases rather than heads because they precede the NP regardless of what constituent begins it (6a). In contrast, Russian case and number suffixes always mark the head, even if it is followed by another modifier (6b).

[5] There has been discussion of "alliterative agreement", i.e. true agreement in form, in some Bantu languages; see Corbett (2006, 87–90). Even if such genuine systems exist, they are expected to be rare. Note that all known claims are for agreement in gender/noun class, not case.

(6) a. English (Germanic > Indo-European)
to [*John's friend*]

b. Russian (Slavic > Indo-European)

[*drug-**u***	*Vas'-i*]
friend-DAT.SG	Basil-GEN.SG

'**to** Basil's friend'

But prefixes/proclitics in head-initial languages and suffixes/enclitics in head-final languages are less trivial, because in this case the head coincides with the edge of the phrase. Therefore, a more reliable criterion is the ability to mark the edge conjunct of *coordinate phrases*, such as in the following example from Nivkh, a head-final language:

(7) Nivkh (isolate)

mañḍu+əs	[*sək*	*p'-umgu-gu*	*p'-ōla-gu*	] ***-kir***
Chinese+owner	all	REFL-woman-PL	REFL-child-PL	-INST

lumr+uski-ɣət-ţ
sable+pay-DISTR/INTS/COMPL-IND

'The owner of the Chinese **with** all his wives and his children paid for the sables.' (Nedjalkov and Otaina 2013, 56)

However, sometimes data on coordination is unavailable. In these cases, I relied on any evidence that shows flags marking an edge constituent that is not a head, such as in the following example from Sanuma, where the instrumental marker *-nö* marks the postnominal adjective rather than the head:

(8) Sanuma (Yanomam)

[*kamakali*	*te*	*wasu*]	***-nö***	*ipa*	*ulu*	*a*	*noma*	*-so*	*-ma*
high:fever	3:SG	deadly	-INST	my	son	3:SG	die	-FOC	-CMPL

'My son died from a deadly high fever.' (Borgman 1990, 123)

3.2.3 Case concord

Because it is difficult to distinguish concord from the use of two separate NPs, I only consider instances of **obligatory** case concord within a **continuous** sequence; thus, phenomena like the abovementioned Old Russian preposition repetition (Klenin 1989) are excluded, since they are not obligatory. Unlike group marking, case concord is relatively rare. It is mostly found in Eurasia (Indo-European, East Caucasian, South Caucasian) and Australia, but also in other areas:

(9) Southern Sierra Miwok (Utian)

pakal-te-m	*ʔansi-nţi-j*	[*oţi·ko-**j***	*pe·so-**j***]
pay-VERB-1SG	son-my-OBJ	two-OBJ	dollar-OBJ

'I'm paying my son **two dollars**.' (Callaghan 1987, 22)

3.3 Sample

The sample I used for the pilot study in Belyaev (2018) is largely based on the intersection of the syncretism sample in Baerman, Brown, and Corbett (2005) (and the corresponding WALS feature Baerman and Brown 2013) and the WALS sample "Exponence of Selected Inflectional Formatives" (Bickel and Nichols 2013). I only exclude languages for which there is not enough data or no access to primary sources; in many cases I have included closely related languages instead. A few well-attested and well-described languages have also been added. In sum, the sample includes 107 languages with a fairly high level of genetic and areal diversity.[6] It is illustrated in Figure 4 (where orange dots mark languages with m-case according to my criteria, and blue dots mark languages with no m-case). The map has been drawn using the `lingtypology` R package (Moroz 2017).

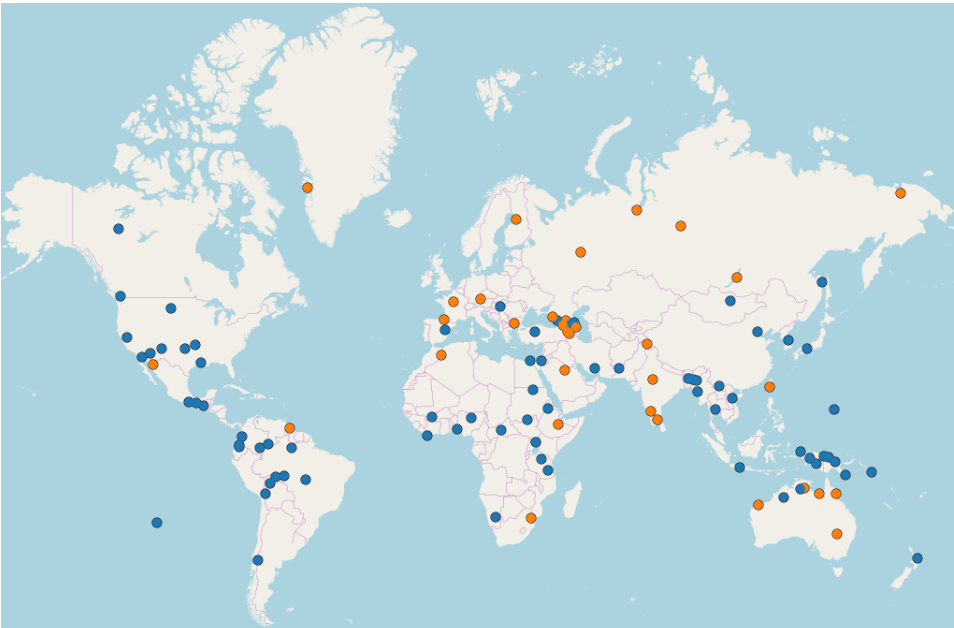


Figure 4: Languages in the sample

3.4 Results

3.4.1 Universal 1

The first hypothesis concerns the relationship between group marking and m-case status: group marking should be impossible in m-case systems.

(10) GROUP ⟶ ¬ M-CASE
M-CASE ⟶ ¬ GROUP

[6] An interactive map of the sample, where one can click to see language names, is available at: http://ossetic-studies.org/obelyaev/case-sample-map.html.

The hypothesis is strongly confirmed, with only three real exceptions, as seen in the contingency table in Table 1.

	¬ M-CASE		M-CASE	
GROUP	**56**	95%	3	5%
	76%		9%	
¬ GROUP	**18**	37%	**30**	63%
	24%		91%	
$\chi^2(1, N = 107) = 40.9059, p < 0.00001$				

Exceptions: Basque (isolate), French (Romance), Burushaski (isolate).

Table 1: Contingency table for Universal 1

Note that many languages in the sample, such as Ossetic (Iranian > Indo-European, Erschler 2012), or Kryz (Lezgic > East Caucasian, Authier 2009, 34), or Oromo (Kushitic > Afro-Asiatic, Owens 1985, 8ff.) do have both group affixation and m-case features. But they are not exceptions because these languages actually possess two case systems: an m-case system, more tightly integrated, often covert, that does not scope over coordination, and an agglutinating, non-m-case system that does scope over coordination. For possible analyses of such mixed systems in LFG, see Belyaev (2014) and Belyaev (2021).

The remaning exceptions may be due to limitations in the typological methodology. For example, French counts as an exception due to cumulation of prepositions with definiteness, number, and gender: *au* [o] (TO.DEF.M.SG) is not synchronically derivable from *à* 'to' + *le* (DEF.M.SG). Furthermore, [o] is syncretic with definite plural (orthographic *aux*). However, this depends on the morphophonological analysis. Furthermore, cumulation in French is "accidental": it does not occur all across the paradigm, and non-cumulative exponents of both case and definiteness are easy to isolate (*à* can be used without an article, or with the feminine singular article – *à la*, etc.). This contrasts with systematic cumulation, such as between case, number and gender in Indo-European case systems (e.g. in Russian or German). Perhaps a distinction should be made between this "real" cumulation and portmanteaux like in French; however, such a distinction is difficult to formalize typologically, and since the exceptions are few anyway, this does not seem to be a serious problem.

Remarkably, there also seems to be a tendency in the opposite direction for non-m-case systems to possess group marking, although it is weaker than Universal 1. Furthermore, individual diagnostics for m-case status are different in their predictive power: INFL, taken alone, is exceptionless. This is in line with Spencer and Otoguro's (2005) observation that inflection classes are the most reliable criterion for m-case status.

3.4.2 Universal 2

The second hypothesis is that case concord is only possible in m-case systems:

(11) Concord ⟶ m-case
¬ m-case ⟶ ¬ Concord

This hypothesis is also confirmed, as seen in Table 2.

	m-case		¬m-case	
Concord	**17**	89%	2	11%
	52%		3%	
¬ Concord	**16**	18%	**72**	82%
	48%		97%	
$\chi^2(1, N = 107) = 37.2353, p < 0.00001$				

Exceptions: Wardaman (Yangmanic), Southern Sierra Miwok (Utian).

Table 2: Contingency table for Universal 2

The statistical significance is high. However, the universal still looks less reliable than Universal 1, because the number of systems with case concord is low in the first place: only 19 in the 107-language sample. The sample should be extended in future work to cover more families and geographic areas.

A possible critique of this universal is that its consequent, m-case, is a disjunction between Sync, Cumul and Infl. This is not a problem for Universal 1, because a disjunction in the antecedent is actually more restrictive than a simple statement. But in Universal 2, it means that a violation of one of the three may be "saved" by the lack of violation of one of the others, thus weakening the universal. It should therefore be noted that, even when individual diagnostics are taken in isolation, the universal is still statistically significant, although the number of exceptions is higher.

3.4.3 Universal 3

A curious corollary of Universals 1 and 2 is a generalization which may be called a third universal:

(12) A case feature in which there is concord cannot have group exponence.

That is, the following implication holds:

(13) Concord ⟶ ¬Group
Group ⟶ ¬Concord

This generalization seems obvious for the conventional view of group/phrasal affixation, where the affix literally attaches to the edge of a noun phrase (14); if affixes attach to adjectives, affixation should occur at the lexical level.

(14) [[ADJ N] CONJ [ADJ N]]-**CASE**

But there are other approaches to suspended affixation, treating it as ellipsis (Erschler 2012) or feature deletion (Kharytonava 2012). In this case, other options may be possible, such as (15), where the case marker occurs on the head and modifiers of the last conjunct but is absent (deleted) from all other conjuncts. Universal 3 predicts that such examples are impossible, and indeed, to the best of my knowledge, none are attested in the literature.

(15) [[ADJ N] CONJ [ADJ-**CASE** N-**CASE**]]

Thus, these findings support the conventional approach to group affixation. In the context of LFG, they also support the syntactic analyses of Broadwell (2008), Belyaev (2014), and Belyaev (2021) rather than a hypothetical edge feature passing approach along the lines of (16). The latter approach does not predict that case features are realized on the edge conjunct that coincides with the direction of attachment of the affix (prefixes attach to the first conjunct, suffixes attach to the last conjunct). It also does not explain why case features are always realized on edge conjuncts, and no systems marking, for example, penultimate conjuncts exist.

(16) $\text{NP} \longrightarrow \underset{\downarrow\in\uparrow}{\text{NP}^{*}} \quad \underset{\uparrow=\downarrow}{\text{Conj}} \quad \underset{\substack{\downarrow\in\uparrow \\ (\uparrow\,\text{CASE})=(\downarrow\,\text{CASE})}}{\text{NP}}$

4 Discussion

4.1 Implications for LFG

In my view, in the context of LFG, these typological findings support preserving lexical integrity in some form; that is, a distinction between lexical morphology and syntactic exponence of grammatical features. However, the latter is to be understood in a wider sense than in the conventional view that relies on language-specific wordhood diagnostics. Syntactic exponence should be treated as the "default"; lexical (morphological) exponence should only be assumed if there is evidence for effects that require resorting to morphology-specific mechanisms. In the domain of case, morphological systems are an obvious minority (33 languages in my sample); only they should be treated as introducing the feature CASE in the lexicon. All other "case" exponents, regardless of their status with respect to wordhood diagnostics or their descriptions in grammars, should be described as corresponding to separate heads in the syntax, as in the analyses of Broadwell (2008) and Belyaev (2021). This agrees with much of current LFG practice of dividing labour between morphology and syntax, but gives it a solid cross-linguistic justification. Another implication is that distinguishing between syntactic and morphological treatment of case markers should be based on Beard's Criterion rather than diagnostics based on bondedness.

At the same time, the morphology–syntax distinction may still be viewed as redundant because it is not formally impossible to analyze *all* case marking phenomena syntactically. What I argue is that such an approach fails to explain the proposed typological generalizations, whereas they follow naturally from the distinction between (lexical) morphology and syntax. In the following, I would like to illustrate this point by analyzing one example from three different approaches that could be used within LFG.

4.2 The Kryz example

I shall consider the following example from Kryz (Lezgic > East Caucasian):

(17)

[*kasib-a*	*sun-**ci***	*fur-**a***	*na*	*xinib-**ci***	] *-**ğar***
poor-a	one-OBL	man-GEN	and	woman-GEN-SUPEREL	

‘About a poor man and his wife.’ (Authier 2009, 199)

Within the framework proposed in this paper, Kryz has both m-case and non-m-case markers, which are both treated as “cases” in Authier (2009). The only m-case marker in Kryz is the genitive (*-a* and the second *-ci*[7] in 17), which fits most of the m-case criteria: it has different forms in different inflection classes and it is sometimes syncretic with the nominative (i.e. zero-expressed). Other “case” markers, such as the superelative *-ğar* in (17), attach to the genitive stem and have a consistent form across all lexemes, singular and plural.

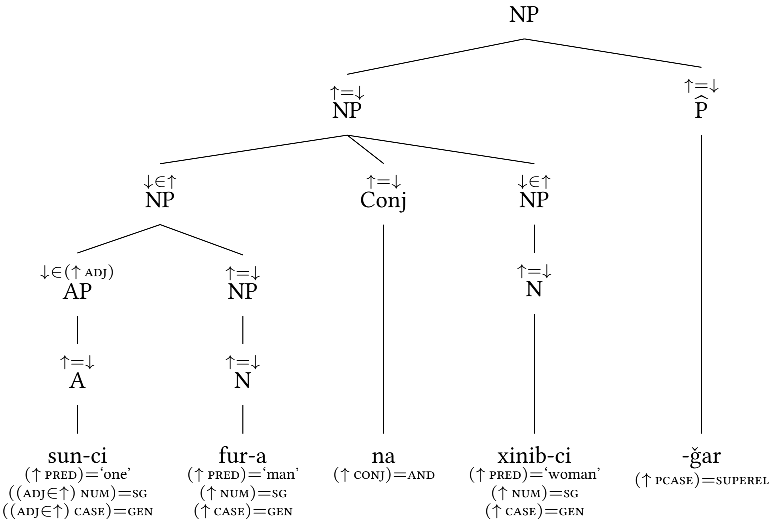
As expected, the genitive marker does not show group exponence; in (17), it appears on both conjuncts (as *-a* on ‘man’ and *-ci* on ‘woman’). In contrast, the superelative *-ğar* scopes over both genitive-marked conjuncts. In accordance with Universal 2, case concord is only found in the genitive; that is, adjectives distinguish between nominative and oblique (recall that oblique cases are based on the genitive). For example, in (17), the numeral ‘one’ has the oblique concord suffix *-ci*, which is equivalent to the genitive affix on the noun ‘woman’. Therefore, Kryz is a paradigm example of all the typological generalizations and distinctions made in this paper.

The most straightforward approach would be to take the term “case” used in the grammar at face value and assume that all case marking is morphological, i.e. lexical. This will not work, because secondary cases like the superelative scope over coordinate phrases. It is technically possible to analyze this via edge feature passing as in (16), but I have stated above why this approach is problematic from a typological point of view; furthermore, this requires treating secondary case as nondistributive, which will create additional problems, such as preventing proper case assignment to coordinate phrases (the set will be assigned a case feature that can be distinct from the features of its elements).

Secondary cases could be treated as clitics, such as in (18). On the analysis of “case” markers as $\widehat{\text{P}}$, see Spencer (2005) on Hindi.

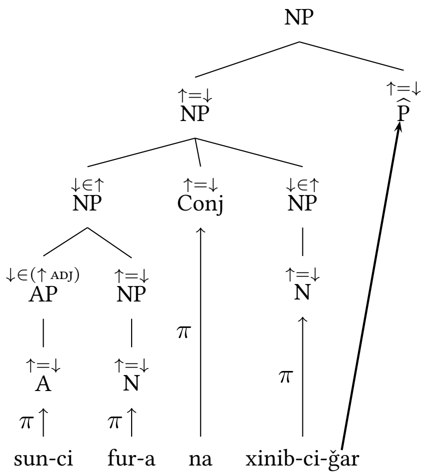
[7] The first *-ci*, on ‘one’, is glossed as OBL because it is treated as an oblique concord marker rather than a genitive case marker by Authier. The two are, of course, related.

(18)



This works as a technical solution, but it misses the fact that Authier (2009) treats elements like *-ǧar* as cases for a reason: they morphonologically pattern with affixes rather than clitics.[8] If this evidence is to be taken seriously, a compromise would be to use lexical sharing (Wescoat 2002; Lowe 2016), as in (19), where the second conjunct co-instantiates the non-projecting $\widehat{\mathrm{P}}$ (case) node and the N node.

(19)



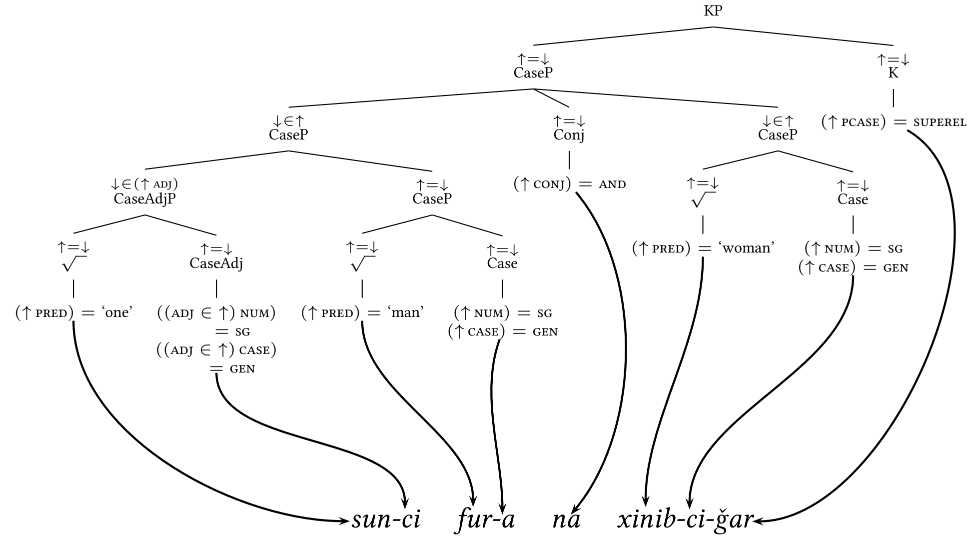
Lexical sharing is not without its problems, however. The most serious problem is that it fails to capture the typological generalizations provided above, namely, that m-case defies syntactic exponence, even through lexical sharing. Syntactic exponence is always affixal and agglutinating. Lexical sharing only specifies which f-description is assigned to which head, but does not capture the contribu-

[8] Authier (2009) does not explicitly discuss the criteria for treating these secondary cases as affixes rather than clitics, but these may be deduced from the data. For example, vowel hiatus is resolved with case markers: *k'ul-ci* (house-GEN) + inessive *-a* ⟶ *k'ul-c-a* ‘in the house’ (Authier 2009, 36), but *riki* (door.GEN) + *ara-c-a°ar* ‘through’ ⟶ *riki ara.c-a°ar* ‘through the door’ (Authier 2009, 96).

tion of individual morphemes; the internal structure of the word form is handled by the morphological component. Thus any features with any formal expression can be handled as lexically shared; generalizations like the ones presented in this paper either cannot be captured or must be captured through additional stipulations in the morphology itself.

Finally, a third alternative is to abandon LI (in its traditional form) altogether and treat all morphology as syntactically expressed. Such is the approach taken in L(R)FG (Melchin, Asudeh, and Siddiqi 2020). A sketch of an L(R)FG analysis of (17) is provided below:

(20)



The problem with this approach is that it completely collapses the morphological difference between the primary case markers and "secondary cases". The fact that only the latter can undergo "phrasal affixation" cannot be explained by a morphology vs. syntax distinction. Rather, it has to be described as a constraint on coordination: KPs and Case(Adj)Ps can be coordinated, but not bare roots (there is no rule that coordinates bare roots). However, this is not realistically translatable to a cross-linguistic constraint, unlike the analyses above. It is not clear why the possibility of coordination would correlate with m-case status of the affixes: why are stems that host m-case markers non-conjoinable, while stems that host other case markers are?

However, this is not so much a feature of L(R)FG itself as a framework, but of its theoretical assumptions. Much like LFG does not have to be lexicalist, arguably, L(R)FG does not have to follow DM assumptions that every morpheme corresponds to a functional head. It is fully compatible with a lexical component, where some morphological features are realized together with the root; indeed, even now this solution must be taken for certain suppletive forms, such as English *my* and other possessive pronouns, to prevent forms like **me's* or **you's*. In this spirit, L(R)FG can be used similarly to lexical sharing, assuming functional heads only where this is syntactically motivated by facts such as group affixation. One advantage over lexical sharing is an explicit mapping between exponents and their corresponding syntactic tree nodes.

5 Conclusions

In this paper, I have presented a typological argument, earlier presented in a more brief form in Belyaev (2018), in favour of lexical integrity based on the notion of (M-)CASE as formulated in Spencer and Otoguro (2005). This approach is based on the properties of the case paradigm and leads to more robust generalizations than prior definitions based on *words* and *affixes*. Specifically, two typological generalizations are shown to be statistically significant: first, m-case status predicts lack of group marking; second, case concord is only possible in m-case systems.

Therefore, in contrast to work such as Haspelmath (2011), the morphology-syntax distinction can be seen as cross-linguistically adequate. However, the scope of morphology is more narrow than traditionally assumed. Most kinds of nominal flag systems fall into the same class as adpositions, regardless of "bondedness".

For LFG, this conclusion suggests that lexical integrity is a reasonable assumption. A natural explanation for the typological data is that flags adhering to Beard's Criterion (m-cases) are always co-expressed at N heads and can never have syntactic expression. A theory that has no strict boundary between morphological and syntactic material fails to account for this.

But a conventional LFG approach that follows a strict definition of lexical integrity is also problematic, as some case affixes that correspond to syntactic heads nevertheless display properties of word-internal elements, and should not be treated in the same way as clitics or independent words. Two possible alternatives, which relax lexical integrity somewhat, are lexical sharing (Wescoat 2002; Lowe 2016) and L(R)FG (Melchin, Asudeh, and Siddiqi 2020). Both, in my view, are problematic: lexical sharing, because it does not model the association between specific affixes and syntactic heads, relegating all work to the morphology and thus allowing shared heads to have any kind of morphological expression; L(R)FG, because it completely removes the boundary between morphology and syntax and fails to provide a satisfactory explanation of the typological generalizations presented herein. A hybrid approach that has a place for both "syntactic" and "lexical" morphology, while providing clear criteria for separation between the two, would be preferable.

An interesting observation that emerges from these typological generalizations is that languages seem to prefer syntactic expression by default. Nothing prevents non-m-cases from being expressed in the lexicon, but they seem to predominantly favour expression in separate syntactic heads. One may speculate that lexical morphology is a "last resort" for language learners: the formation of linguistic expressions is relegated to the lexicon only if the paradigm structure cannot be accounted for in the syntax.

This typological study, and its results, remain preliminary. The sample is not fully balanced, especially with respect to case concord: more data from other linguistic areas and language families should be included in order to make Universal 2 more reliable. The comparative concept is also too crude as it fails to distinguish between different kinds of cumulation (cf. the French example above), syncretism (phonologically motivated vs. systematic), and inflection classes (purely idiosyncratic vs. semantically motivated variation). This, however, is an inherent feature of the typological method, which has to rely on relatively coarse-grained concepts in order to achieve a large coverage of languages; it is the goal of the theory to provide the initial hypotheses and explain any exceptions.

Other typological parameters of case systems, such as case compounding, Suffixaufnahme, and affix order, may be considered as well, in addition to group marking and concord. However, it is not clear whether these phenomena are frequent enough in languages of the world to provide raw data for a robust typological study.

Finally, if my explanation of the observed universals is on the right track, similar observations should hold for other nominal features, such as number, and other word classes, such as verbs. Notions like m-case should be devised for these domains as well. Case, however, seems to be an appropriate initial testing ground, being a purely syntactic feature whose set of values is determined solely on the basis of its marking patterns.

References

Anderson, Stephen R. 1992. *A-morphous morphology.* Cambridge Studies in Linguistics 62. Cambridge: Cambridge University Press.

Authier, Gilles. 2009. *Grammaire kryz, langue caucasique d'Azerbaïdjan, dialecte d'Alik.* Leuven: Peeters.

Baerman, Matthew, and Dunstan Brown. 2013. "Case syncretism." In *The World Atlas of Language Structures Online.* Leipzig: Max Planck Institute for Evolutionary Anthropology.

Baerman, Matthew, Dunstan Brown, and Greville G. Corbett. 2005. *The syntax-morphology interface: a study of syncretism.* Cambridge Studies in Linguistics 109. Cambridge: Cambridge University Press.

———. 2017. *Morphological complexity.* Cambridge: Cambridge University Press.

Beard, Robert. 1995. *Lexeme-Morpheme Base Morphology.* Stony Brook, NY: SUNY Press.

Belyaev, Oleg. 2021. "Paradigm structure influences syntactic behaviour." In *Modular design of grammar,* ed. by I Wayan Arka, Ash Asudeh, and Tracy Holloway King, 251–281. Oxford: Oxford University Press.

Belyaev, Oleg I. 2014. “Osetinskij kak jazyk s dvuxpadežnoj sistemoj: gruppovaja fleksija i drugie paradoksy padežnogo markirovanija” [Ossetic as a two-case language: Suspended affixation and other case marking paradoxes]. *Voprosy jazykoznanija* 6:31–65.

———. 2018. “Morfosintaksis padeža i struktura imennoj paradigmy” [Morphosyntax of case and the structure of the noun paradigm]. *Typology of Morphosyntactic Parameters* 1 (1): 51–72.

Bickel, Balthasar, and Johanna Nichols. 2013. “Exponence of selected inflectional formatives.” In *The world atlas of language structures online,* ed. by Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Borgman, Donald M. 1990. “Sanuma.” In *Handbook of Amazonian languages,* ed. by Desmond C. Derbyshire and Geoffrey K. Pullum, 2:15–248. Berlin: De Gruyter Mouton.

Bresnan, Joan, Ash Asudeh, Ida Toivonen, and Stephen Wechsler. 2016. *Lexical-functional syntax.* Second edition. Blackwell Textbooks in Linguistics 16. Oxford: Wiley Blackwell.

Bresnan, Joan, and Samuel A. Mchombo. 1995. “The lexical integrity principle: evidence from Bantu.” *Natural Language & Linguistic Theory* 13:181–254.

Broadwell, George Aaron. 2008. “Turkish suspended affixation is lexical sharing.” In *Proceedings of the LFG ’08 Conference,* ed. by Miriam Butt and Tracy Holloway King. S: CSLI Publications.

Bruening, Benjamin. 2018. “The lexicalist hypothesis.” *Language* 94 (1): 1–42.

Callaghan, Catherine A. 1987. *Northern Sierra Miwok Dictionary.* University of California Publications in Linguistics 110. Berkeley, CA: University of California Press.

Chaves, Rui P. 2008. “Linearization-Based Word-Part Ellipsis.” *Linguistics and Philosophy* 31:261–307.

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

———. 2007. “Canonical typology, suppletion, and possible words.” *Language* 83 (1): 8–42.

Dalrymple, Mary. 2015. “Morphology in the LFG architecture.” In *Proceedings of the LFG15 Conference,* ed. by Miriam Butt and Tracy Holloway King, 64–83. Stanford, CA: CSLI Publications.

Erschler, David. 2012. “Suspended affixation in Ossetic and the structure of the syntax-morphology interface.” *Acta Linguistica Hungarica* 59 (1–2): 153–175.

Göksel, Aslı, and Celia Kerslake. 2005. *Turkish: A comprehensive grammar.* London: Routledge.

Haspelmath, Martin. 2010. “Comparative concepts and descriptive concepts in crosslinguistic studies.” *Language* 86 (3): 663–687.

———. 2011. “The indeterminacy of word segmentation and the nature of morphology and syntax.” *Folia Linguistica* 45 (1): 31–80.

Haspelmath, Martin. 2019. "Indexing and flagging, and head and dependent marking." *Te Reo* 62, no. 1 (Special issue in honour of Frantisek Lichtenberk): 93–115.

Kabak, Bariş. 2007. "Turkish suspended affixation." *Linguistics* 45 (2): 311–347.

Kharytonava, Volha (Olga). 2012. "Taming affixes in Turkish: with or without residue?" In *Irregularity in morphology (and beyond),* ed. by Thomas Stolz, Hitomi Otsuka, Aina Urdze, and Johan van der Auwera, 167–185. Studia Typologica 11. Berlin: De Gruyter.

Klenin, Emily. 1989. "On preposition repetition: A study in the history of syntactic government in Old Russian." *Russian Linguistics* 13:185–206.

Lowe, John J. 2016. "English possessive 's: clitic AND affix." *Natural Language & Linguistic Theory* 34:157–195.

Melchin, Paul B., Ash Asudeh, and Dan Siddiqi. 2020. "Ojibwe agreement in Lexical-Realizational Functional Grammar." In *Proceedings of the LFG'20 Conference,* ed. by Miriam Butt and Ida Toivonen, 268–288. S: CSLI Publications.

Moroz, George. 2017. *lingtypology: easy mapping for linguistic typology.*

Nedjalkov, Vladimir P., and Galina A. Otaina. 2013. *A syntax of the Nivkh language.* Studies in Language Companion Series 139. Amsterdam: John Benjamins.

Otoguro, Ryo. 2006. "Morphosyntax of case: a theoretical investigation of the concept." PhD diss., University of Essex.

Owens, Jonathan. 1985. *A grammar of Harar Oromo (Northeastern Ethiopia).* Kuschitische Sprachstudien 4. Hamburg: Helmut Buske Verlag.

Spencer, Andrew. 1991. *Morphological theory.* Oxford: Blackwell.

———. 2005. "Case in Hindi." In *Proceedings of the LFG05 Conference,* ed. by Miriam Butt and Tracy Holloway King. Stanford, CA: CSLI Publications.

Spencer, Andrew, and Ryo Otoguro. 2005. "Limits to case: a critical survey of the notion." In *Competition and variation in natural languages: the case for case,* ed. by Mengistu Amberber and Helen de Hoop, 119–145. Amsterdam: Elsevier.

Stump, Gregory T. 2001. *Inflectional morphology: a theory of paradigm structure.* Cambridge Studies in Linguistics 93. Cambridge: Cambridge.

Wescoat, Michael T. 2002. "On lexical sharing." PhD diss., Stanford University.

A hybrid model of auxiliary contraction: evidence in children's speech

Joan Bresnan
Stanford University

Proceedings of the LFG'21 Conference

On-Line

Miriam Butt, Jamie Y. Findlay, Ida Toivonen (Editors)

2021

CSLI Publications

pages 26–50

http://csli-publications.stanford.edu/LFG/2021

Keywords: auxiliary contraction, exemplars, hybrid model, probabilistic, near-categorical, acquisition

Bresnan, Joan. 2021. A hybrid model of auxiliary contraction: evidence in children's speech. In Butt, Miriam, Findlay, Jamie Y., & Toivonen, Ida (Eds.), *Proceedings of the LFG'21 Conference, On-Line*, 26–50. Stanford, CA: CSLI Publications.

Abstract

The present study finds parallel patterns of auxiliary contraction in a corpus study of children's speech and an earlier corpus study of adults' speech (Bresnan 2021). The combination of probabilistic and near-categorical patterns is accounted for by the hybrid model of auxiliary contraction of Bresnan (2021). These findings show that children's language, like that of adults, depends on both the usage probabilities of multiword sequences and their prosodic and rhythmic patterns reflecting the syntactic context.

The hybrid formal and usage-based model of auxiliary contraction of Bresnan (2021) combines the formal grammar of LFG including lexical sharing, and a dynamic, exemplar-based lexicon.[†] It accounts for contraction phenomena unexplained by either of the component theories alone: (1) the usage-based lexicalization of contractions, (2) the probability of cooccurrence of word and auxiliary predicting the probability of their contraction, (3) the prosodic wordhood of contractions, and (4) the rightward metrical dependence of unstressed auxiliaries in weak positions. The present study finds similar patterns of auxiliary contraction in a corpus study of children's speech, showing that children's language, like that of adults, depends on both the usage probabilities of multiword sequences and their prosodic and rhythmic patterns reflecting the syntactic context.

Section 1 briefly sketches the hybrid model and Section 2 exemplifies the probabilistic and near-categorical patterns in adult speech that follow from the model, summarizing highlights of Bresnan (2021). The new contribution of the present study, Section 3, presents evidence from a corpus study showing that similar patterns occur in children's speech.

1 The hybrid model

To grasp the hybrid model of auxiliary contraction quickly from the point of view of an LFG grammarian, take a standard LFG grammar, add lexical sharing and connect it to prosodic and metrical structures; then swap out the LFG lexicon for an exemplar-based lexicon, and visualize the resulting LFG lexical schemata as labeling clouds of lexical exemplars. These steps, described in more detail in Bresnan (2021), are briefly illustrated here as background.

In the lexical sharing theory of auxiliary contraction illustrated in Figure 1, adjacent terminal category nodes D, I are mapped to the same lexical exponent *you're*.[1] Unlike the dominance relations in c-structure trees, which are indicated by straight lines connecting nodes, the mapping from terminal syntactic

[†]I thank Ida Toivonen and an anonymous reviewer for suggested revisions. For the analyses and graphics I used R (R Core Team 2020) with contributed packages lme4 (Bates et al. 2015), lattice (Sarkar 2008), and rms (Harrell Jr 2021). This open-source software can be downloaded at no cost from https://cran.r-project.org/.

[1]Here the category labels D, I and C are used for convenience and explicitness, in order

categories to lexical exponents is many-to-one and is indicated by the arrows pointing from the terminal c-structure categories (D, I, V, and the like) to the lexical exponents *you're* and *going* in Figure 1. The adjacent nodes D and I in Figure 1 are pointing to the same exponent *you're*, and are said to "share" it. For formal details of the instantiation of lexical schemata of the atomic units *you* and *'re* as D and I and of the joint constraints on the entire contraction *you're*, see Wescoat (2005).

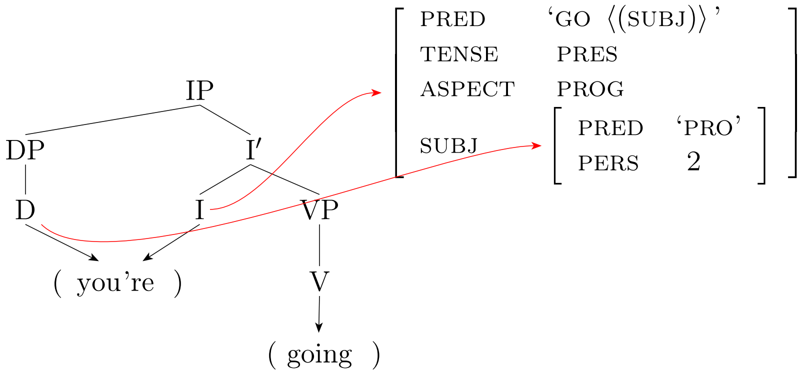


Figure 1: Lexical sharing

To Wescoat's 2005 formal theory of English auxiliary contraction, Bresnan (2021) adds prosodic and metrical connections. First, lexical sharing implies prosodic wordhood of the lexical exponent as illustrated in Figure 2, because all lexical words are prosodic words (Inkelas 1991, Inkelas and Zec 1993, Selkirk 1996).[2]

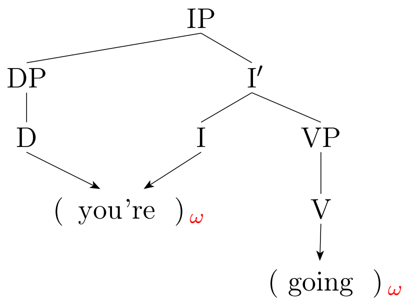


Figure 2: Lexical sharing implies prosodic wordhood

Second, in the lexical sharing analysis the contracted and uncontracted forms of the auxiliary *are, 're* have the same c-structure position. See Figure 3.

to represent respectively pronouns and the uninverted and inverted positions of the tensed auxiliaries. The specific choices and granularity of the c-structure category labels are not crucial to the model (Bresnan 2021, n. 25, p. 123).

[2]See Bresnan (2021, n. 27, p. 125) on apparent exceptions.

For unstressed auxiliaries the I in Figure 3 (and likewise C) is a metrically weak position requiring a strong—that is, stressed—complement (Bresnan 2021, pp. 117–119, 125).

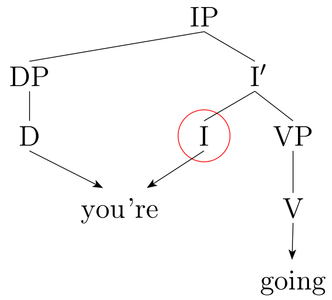


Figure 3: Contracted *'re* has the same c-structure position as uncontracted *are*, a metrically weak position for unstressed auxiliaries.

Arguing against affixed-word analyses of contraction and in favor of simple cliticization, Wescoat (2005) motivates the syntactic position of the auxiliary in Figure 3 with evidence from coordination, where parallel I′ nodes can be conjoined despite the head of the first being lexically shared with the subject (1a,b), and where subject-auxiliary contractions cannot be conjoined, because they are nonconstituent D I sequences (1c).[3]

(1) a. ***I***[***'m*** *looking forward to seeing you*] *and* [*will be there on Sunday*]

b. ***You***[***'ll*** *do what I say*] *or* [*will suffer the consequences*]

c. *[***They're*** *and* ***you're***] *going.*

An alternative analysis of (1)a,b that does not involve I′ coordination is left peripheral ellipsis of the rightward subject of conjoined IPs (Bresnan and Thráinsson 1990):

(2) [***You'll*** *do what I say*] *or* [*(**you**) will suffer the consequences*]

However, this alternative is inapplicable to cases like (3a), where the operator *who* has scope over coordinated complement I′ (or C′). Here a left-peripheral source is not semantically equivalent to (3a). The question in (3a) is about the ones that will both forget and suffer the consequences, while in (3b), the ones that will forget are not necessarily the same as the ones that will suffer the consequences. Thus despite the availability of left peripheral ellipsis, I′/C′ coordination still provides evidence for Wescoat's theory of lexical sharing.

[3] As a reviewer notes, non-constituent D I sequences can appear in the conjoined clause residue of right node raising (RNR), as in *They might and you will do it*, where a VP is extracted from both sentential conjuncts. RNR is prosodically marked by accents on the right edges of the residue conjuncts and does not bear on the absence of ordinary NP conjunction for examples like (1c) expected under affixed-word analyses (Wescoat 2005).

(3) a. ***Who***[***’ll*** *forget*] *and* [*will suffer the consequences*] *?*

b. ≠ [***Who****’ll forget*] *and* [***who*** *will suffer the consequences*] *?*

In the hybrid model, these LFG components are linked to a dynamic exemplar-based lexicon (Bybee 2001, 2006, Bybee and Hopper 2001) as mathematically modelled by Pierrehumbert (2001, 2002, 2006) at the level of word phonetics. Figure 4 provides a simplified visualization of tensed auxiliary contractions in this model. The labels *you, you’re,* and *are* with their varying pronunciations stand for (partial) ‘lexical entries’ in traditional linguistic terminology and correspond to structural descriptions at several levels. Each entry maps onto a matching set of remembered instances of its utterance—the memory traces, or exemplars, structured into ‘clouds’ by similarity.[4] The visualization is simplified to show only varying pronunciations of remembered instances; it omits links to further grammatical, pragmatic, semantic, and social information. Fresh experiences and memory decay lead to continual updating of the entries in the mental lexicon, so that frequent, recent instances are more highly activated than infrequent, temporally remote ones.

labels: *you* [juː/jə] *you’re* [juːɹ/jʊɹ/jɔɹ] *are* [ɑɹ/əɹ]

memory traces:
[jə] [jʊɹ] [əɹ]
[jə] [jə] [jʊɹ] [əɹ] [əɹ]
[jə] [jɔɹ] [jɔɹ] [juːɹ]
[juː] [jʊɹ] [jʊɹ] [jʊɹ] [ɑɹ] [əɹ][ɑɹ]
[juːɹ] [jɔɹ] [jʊɹ]
[juː] [jʊɹ] [jʊɹ]
[juː] [jə] [jʊɹ] [ɑɹ]
[juːɹ] [juːɹ] [əɹ]
[jɔɹ] [ɑɹ]

Figure 4: Exemplar-based lexicon

The hybrid lexicon replaces the ‘lexical entries’ in Figure 4 with LFG lexical schemata within the lexical sharing theory, so that LFG structures serve to label or index the clouds of memory traces. The result is visualized in Figure 5 with extensional depictions of the lexical schemata for contractions (Bresnan 2021).

[4]The similarity within exemplar clouds is symbolised here by their matching color, viewable in the online version of this paper.

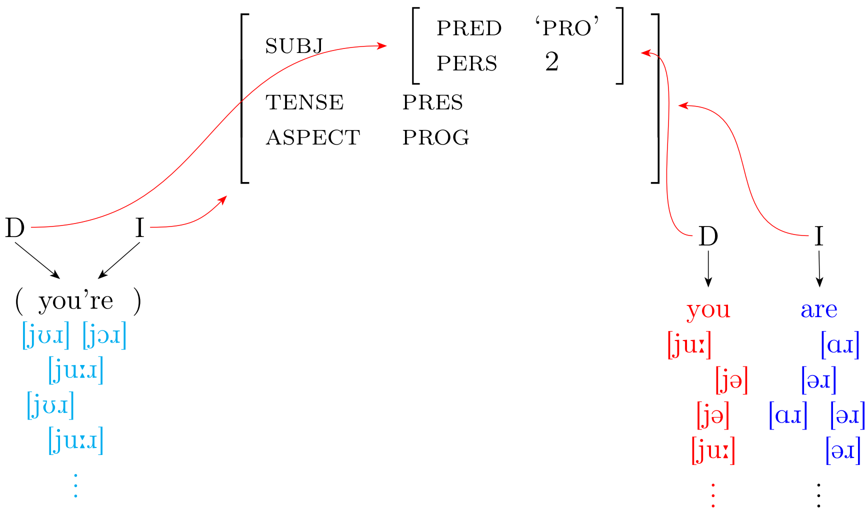


Figure 5: LFG functional schemata label lexical exemplar clouds

2 Consequences of the model

The hybrid model has broader explanatory scope than either of its usage-based or formal-grammar-based components alone. The main consequences are briefly reviewed here; see Bresnan (2021) for detailed discussion of evidence and analyses of data.

Lexicalized contractions On the usage-based theory of the lexicon, more frequently used words and multiword expressions are phonetically more reduced and become lexically stored (e.g. Bybee 2001, 2006, Bybee and Hopper 2001, Pierrehumbert 2001, 2002, 2006, Seyfarth 2014, Sóskuthy and Hay 2017). Table 1 shows some examples of this phenomenon in auxiliary contractions collected by Wescoat (2005, 471–2). Arguing for a lexical source for these and other nonsyllabic auxiliary contractions, he observes that the laxed vowels occur *even in slow or emphatic speech*, unlike on-line contextual adjustments in the phonology of rapid connected speech.

Table 1: Wescoat's (2005, 471) "morphological idiosyncracies" in auxiliary contractions cited as evidence for their lexical source. Unlike fast-speech phenomena, "*I'll* [ɑl] and *you're* [jɔɹ] may be heavily stressed and elongated".

I'll	[aɪl/ɑl]	*I'm*	[aɪm/*ɑm]	*I've*	[aɪv/*ɑv]
you'll	[juːl/*jɔl]	*you're*	[juːɹ/jɔɹ]	*you've*	[juːv/*jɔv]

Probability of contraction Recent work on English auxiliary contraction has found that probabilistic measures derived from frequencies of use of hosts and auxiliaries correlate with the likelihood of contraction (Frank and Jaeger 2008, Spencer 2014, Barth and Kapatsinski 2017, Barth 2019, Bresnan 2021). These results are expected in the exemplar-based lexicon: given production biases toward lenition and shortening, contractions of hosts and auxiliaries tend to increase with their production.

For example, in Bresnan's (2021) study of auxiliary contraction in New Zealand English the nouns having the highest share of cooccurrences with *is/'s* are *one, mum, dad*, and *thing*: 83.7% are contracted with the auxiliary, compared to the average of 56.5% for all nouns.

Prosodic wordhood of contractions The prosodic wordhood of tensed auxiliary contractions is supported by the word-level phonological processes in contractions and by the absence of pausing and interruptions between the host and the contracted auxiliary (Bresnan 2021), as shown in (4a,b).

(4) a. **(we... um... 've)$_\omega$ all done it*
(cf. *we've all done it*)

b. **(he ... uh... 's)$_\omega$ odd*
(cf. *he's odd*)

Rightward metrical dependence Unstressed auxiliaries occurring in I/C are in a metrically weak position which must be followed by a strong sister phrase (bearing stress), as Figure 6 illustrates.

Figure 6: Rightward metrical dependence of weak I/C

When unstressed auxiliaries in the weak I position lack a following strong sister phrase, the result is ungrammatical, as exemplified in (5a,b).

(5) a. *You're* [ɹ] *going.*
**You're* [ɹ].

b. *You are* [əɹ] *going.*
**You are* [əɹ].

And the same holds for weak C position. For example, in (6a,b), the IP or S complement to the inverted auxiliary in C must contain a strong (stressed) element.[5] Stressless *it* is followed by stressed *doing* in the (6a), but not in (6b), where contraction is ill-formed.[6]

(6) a. *That bird, what's it doing?*

b. **That bird, what's it?*

This shared metrical behavior of contracted and uncontracted auxiliaries in metrically weak positions is called the "central generalization of contraction" by Selkirk (1984, 405): "only auxiliaries that would be realized as stressless in their surface context may appear in contracted form" and it is also the core generalization of Labov's (1969) analysis, which phonologically derives the contracted forms from the uncontracted in the same phrase structure position. In contrast to previous morphological analyses of some contractions (e.g. Sadler 1998; Spencer 1991; Bender and Sag 2001; Börjars et al. 2019, pp. 87–88), it follows from the present lexical sharing analysis in which the contracted and the unstressed uncontracted auxiliary forms occupy the same syntactic position where the shared metrical requirement of a strong sister constituent in c-structure applies.

Natural speech is full of dysfluencies and incomplete utterances, including what appear to be violations of rightward metrical dependence (e.g. *I'm . . . Oh never mind.*) However, these are usually marked either by conventional discourse fillers signalling a planned delay (Clark and Tree 2002) (e.g. *I'm um . . .*) or by the absence of utterance-final intonation. Hence they do not undermine rightward metrical dependence as a property of fluent, complete utterances.

3 In children's speech

Similar patterns in auxiliary contractions appear in children's speech: lexical contractions, the probability of cooccurrence of host and auxiliary predicting probability of contraction, and the prosodic wordhood and rightward metrical dependence of contractions.

3.1 Lexical contractions in the previous literature

Early work on auxiliary contractions in children's speech investigates the order of acquisition of the alternative forms, with conflicting conclusions.[7] Much of

[5]See Wescoat (2005) and Bresnan (2021) on the analysis of inverted auxiliary contractions within the lexical sharing theory.

[6]See Bresnan (2021, p. 118 and n. 17) for further discussion concerning stressed and stressless *it* in relation to auxiliary contraction.

[7]Brown (1973) and de Villiers & de Villiers (1973) each give evidence for the opposite order of acquisition of uncontracted and contracted *be* forms. Kuczaj (1979) attributes the conflicting conclusions of the earlier work to sampling differences.

the subsequent literature on the development of auxiliaries is concerned with the debate between constructivist and generativist approaches to the development of the tense/aspect and agreement/inflection systems, and generally disregards the topic of auxiliary contraction itself (e.g. Wexler 1994, 1998, Rice et al. 1998, Wilson 2003, Theakston and Lieven 2005, 2008, Theakston et al. 2005, Rowland and Theakston 2009, Theakston and Rowland 2009, Rispoli et al. 2009, Rissman et al. 2013, Rispoli 2016).

The constructivist line of research on auxiliaries, although not focused on contraction *per se*, has shown that contractions of auxiliaries with their hosts are acquired as lexically stored units in children's speech. For example, in a longitudinal corpus study of the acquisition of three exponents of the category of 'inflection' in English—the copula *be*, the auxiliary *be*, and 3sg present agreement—Wilson (2003, 75) shows that ***children learn lexically specific host-auxiliary chunks***—sequences such as *he's, that's*—independently of learning general subject-auxiliary combinations.[8] Pine et al. (2008) replicate Wilson's findings in a different longitudinal dataset of children's speech with additional controls.[9] Wilson (2003, 84) further observes that because they constitute ***prosodic words***, lexically specific chunks like *he's, that's* may be "more readily extractable units than other recurring sequences such as *is* V-*ing*, which does not constitute a single prosodic word."

Regarding children's acquisition of such contractions as units, (Wilson 2003, 85) makes an important point:

> The position that items like *he's* and *I'm* may be unanalysed in child grammar has been held by many researchers. However, an important point needs to made. Although we will argue that *he's* and *I'm* are often unsegmented in child grammars, this does not imply that they are simply equivalent to *he* and *I*, as some researchers have seemed to suggest (e.g. Pinker 1996, 261). Empirically it is clear that they are not, because it is very rare that children say things like *I'm want it*, which would be expected if they did not distinguish between *I'm* and *I*. In terms of the present account, *I* and *I'm* are claimed to be represented very

[8]Wilson (2003, 88) counts both contracted and uncontracted copulas and auxiliaries, excluding contexts in which *be* cannot be contracted, such as before VP ellipsis. He does not report separate counts for contracted vs. uncontracted forms, and notes that where contraction could be orthographically indicated, the children "almost always" used it. His transcripts, selected from five longitudinal corpora in CHILDES, spann the ages 1;6–2;3, 1;11–2;5, 1;8–2;7, 2;8–3;5, and 2;3–3;5. The transcripts of the present study (Section 3.2) are a superset of Wilson's (2003, 87), drawn from the same five corpora together with three additional longitudinal corpora, and including a wider range of children's ages.

[9]Note that the term "auxiliary" in the present study includes the copula, following Bresnan (2021, n. 1, p. 109). In contrast, both Wilson (2003) and Pine et al. (2008) refer to the same verb forms as "auxiliary" or "copula" depending on the construction they occur in. In the present framework, copular and auxiliary constructions are otherwise distinguished (cf. Bresnan 2021, pp. 134–135).

> differently in the child's grammar: the unit *I'm* exists only as part of the construction in (6c) [*I'm* V-*ing*], and other construction(s) for copula sentences. It has no independent existence as a lexical item which would allow it to be used to construct a sentence like *I'm want it.*

In the present framework, *I'm* is represented as a shared lexical exponent of adjacent pronoun and auxiliary categories, which affects its meaning and syntactic distribution. Even if children's very early usage of such contractions may treat them as a single fused word rather than a composite of morphemes, their contexts of distribution indicate that they generally carry some version of the functional information expected under lexical sharing (cf. Figure 5). Recent corpus and experimental work has argued that the inventory of words and chunks gradually developed with statistical learning during language acquisition is used during children's comprehension and production and persists into adulthood (e.g. Arnon et al. 2017, McCauley and Christiansen 2019, Isbilen et al. 2020).

It is thus reasonable to infer from the previous literature that contractions of auxiliaries with their hosts are acquired as lexically stored chunks in children's speech.

3.2 Data of the present study

The questions the present study of children's speech addresses are parallel to those in Bresnan's (2021) study of adult speech: Does the probability of cooccurrence of host-auxiliary sequences predict their probability of contraction? Do contractions behave like prosodic words? Are weak (unstressed) I/C auxiliaries rightward metrically dependent?

Data to answer these questions comes from a joint project with Arto Anttila and Research Assistant Gwynn Lyons at Stanford in the Summer of 2015. The project selected eight longitudinal corpora consisting of 386,155 utterances from conversational interactions children between $1\frac{1}{2}$ and over 5 years of age and their caretakers, contributed to the CHILDES database of North American English (MacWhinney 2000a) by Brown (1973), Clark (1978), Demetras (1986), Kuczaj (1979), Sachs (1983), and Suppes (1974). From these corpora the project team extracted 87,318 utterances of both child and child-directed speech by means of Python scripts using the morphological parsing tier provided with these corpora in CHILDES (MacWhinney 2000b). The extracted utterances contained any of the six tensed auxiliary verbs *is, are, am, will, have, has*, orthographically transcribed as full or contracted (*'s, 're, 'm, 'll, 've, 's*). Python scripts also collected **ngrams** from a broader set of North American English child corpora with longitudinal samples, consisting of 584,941 utterances, including child-directed speech, from both the eight selected corpora and ten additional corpora.

After manual inspection and exclusion of misparses and dysfluencies, unintelligible or incomplete hosts of the auxiliaries, main verb uses of *have* and *has*, infinitive forms, and possessives and plurals mistaken for the auxiliary *'s*, the "cleaned" dataset consists of 79,683 utterances, or 0.913 of the original data. From this dataset the target children's utterances were extracted and further inspected, removing 79 instances of main verb *have*, unsegmented expressions containing contractions (e.g. *suh* for *it's a*), unintelligible contexts, and possessive *'s* mistagged in the morphological tier as contractions of *is*. This children's dataset contains 25,270 utterances and is the source of the statistics in Sections 3.4 and 3.5.

To examine whether usage probabilities affect contraction in children's speech, it is necessary to focus on the portion of data where contractions are not already ruled out by the grammar itself. Therefore cases where contraction is prohibited for reasons of grammar (cf. MacKenzie 2012) were all excluded: where the auxiliary occurs in utterance final position, is directly preceded by a pause, lacks a leftward host altogether, is stressed by a preceding or following intensifying adverb (*too, really, probably*, is preceded by a non-noun (*hey, yeah, okay, uhhuh, away, hi, either, maybe, hurry, together*) or a host having a final sibilant when the auxiliary verb is *is* or *has*. The resulting subset of data contains 21,385 utterances, and is the source of the statistics in Section 3.3.

How reliably do the transcriptions indicate contraction? For Bresnan's (2021) corpus studies of adult speech, the researchers verified that samples of the transcribed contractions matched the acoustic files or phonetic transcriptions. For the data collection used in the present study the researchers did not have recordings for most of the CHILDES corpora used, so in principle the adult transcriptions of children's speech might reflect the adult transcribers' knowledge of grammar.[10] However, the manual for the CHAT transcription format used in these corpora (MacWhinney 2000b) provides cautions and training for the issues and problems that arise when transcribing children's speech, including the many divergences between speech and writing and many ways of transcribing and coding divergences between child and adult speech, and for marking unclarities. Transcribers were instructed to adhere as closely as possible to the child's actual output utterances regardless of deviations from the adult language. The transcriptions include many child pronunciations of words (e.g. "gween" for "green" and "dat" for "that") and there are multiple instances of transcriptions of utterances which would be ungrammatical in the adult child-directed speech, such as omitted, doubled, and superfluous auxiliaries: *it horsie, what is he's doing?, it's makes loud noise*. These show that the transcribers focused on distinctive properties of children's speech and did not generally assimilate it to adult knowledge of language.

[10] This possible objection was provided to the author by Chit-Fung Lam in personal communication dated July 17, 2021.

3.3 Probability of contraction

For Wilson (2003, 86), "The constructivist account predicts that on the whole, copula and auxiliary *be* should occur more frequently with closed-class (or highly frequent) subjects with which *be* can be learned as a chunk." His study and that of Pine et al. (2008) find that in children's speech the cooccurrence frequencies of subjects with 3rd singular inflections on main verbs, copulas, and auxiliaries (both contracted and uncontracted) are generally higher with pronoun subjects.

In what follows the **conditional probability** of a word in the context before an adjacent auxiliary in contracted or uncontracted form is used (cf. Bresnan 2021):

(7) $$\boldsymbol{P(host|aux)}$$

The probability in (7) is estimated from corpora by the ngram calculation shown in (8):

(8) $$\frac{\boldsymbol{count(host\ aux)}}{\boldsymbol{count(aux)}}$$

The natural logarithm is used to compress extreme values. For example, in the ngram collection (Section 3.2) there were 7 bigrams of *Agra is* or *Agra 's* and 103,457 unigrams of *is* or *'s*. So log P(*Agra*|*is/'s*) is calculated as log(7/103,457) = −9.601001. And log P(*Mommy*|*is/'s*) = log(533/103,457) = −5.26839, while log P(*Mommy*|*will/'ll*) = log(70/10,139) =−4.975649.

On the choice of "backward" rather than forward conditional probability—measuring the probability of the potential host given the following auxiliary, rather than the probability of the auxiliary given the potential host—see Bresnan (2021, 113–114) and references. McCauley and Christiansen (2019) argue for the same "backward" condition in their model of chunking in child language learning.

In our dataset, 686 different pre-auxiliary nouns (from the letter *a* to *Zorro*) were identified, along with 43 different types of pre-auxiliary pronouns and pro-forms.[11] (9) shows these pronouns as transcribed in the corpora:

(9) Pre-auxiliary pronouns:
anybody, dat, de, everybody, everyone, everything, he, her, here, him, how, I, it, me, mine, nobody, none, nothing, now, she, so, some, somebody, someone, something, that, them, there, these, they, this, those, we, wha, what, when, where, who, why, you, yours, em, then

The pronouns cooccur with following auxiliaries far more often in our dataset than lexical nouns do:

[11] The term 'pronoun' is used henceforth to include pro-forms such as pro-adjectives and pro-adverbs.

(10) Instances of pre-auxiliary use (tokens):

pronouns	nouns
19,549	1,836

Figure 7 shows how the mean log conditional probability of potential hosts given the target auxiliaries differs by host type in each of the 8 selected corpora of children's speech. An ANOVA test comparing two linear mixed-effect models of $logP(host|aux)$, both including a random effect of child and differing only in the presence of a fixed effect of host type (pronoun vs. noun), yielded a significant effect of host type: $\chi^2(1) = 20{,}630$, $p = 2.2 \times 10^{-16}$.

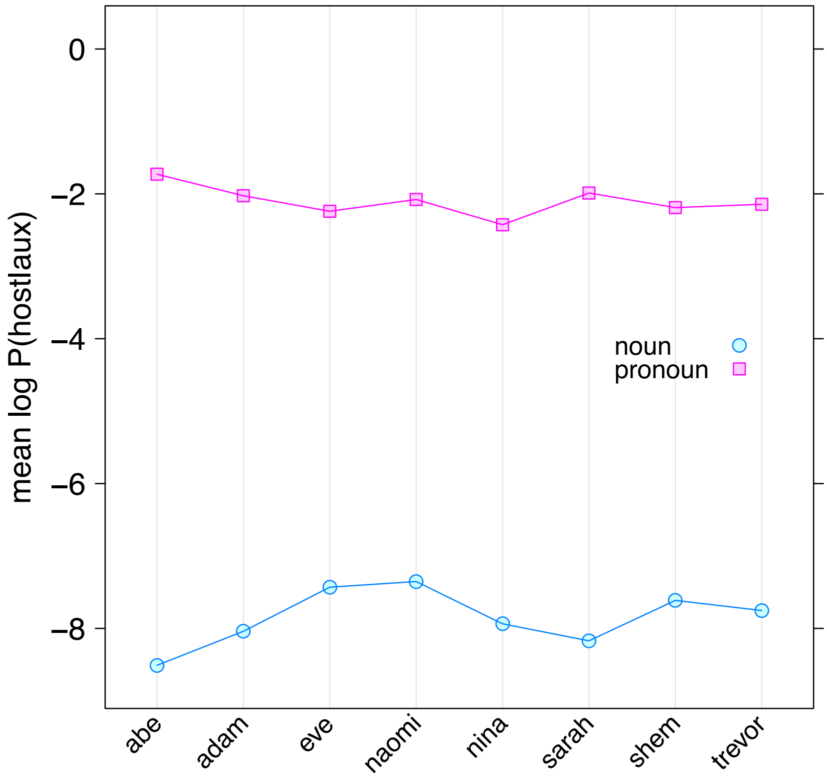


Figure 7: Mean log conditional probability of pre-auxiliary nouns and pronouns produced by children in 8 selected corpora

Given this substantial difference in cooccurrence probabilities, we would expect from the hybrid auxiliary model to find more contractions with the proform and pronoun subjects than with lexical noun subjects. Figure 8 bears this prediction out for each target child, showing again that the proportion contracted differs by host type. An ANOVA test comparing two logistic mixed-effect models of proportion contracted weighted by the numbers of total observations, both including a random effect of child and differing only in the presence of a fixed effect of host type (pronoun vs. noun), yielded a significant effect of host type on contraction: $\chi^2(1) = 2{,}338.1$, $p = 2.2 \times 10^{-16}$.

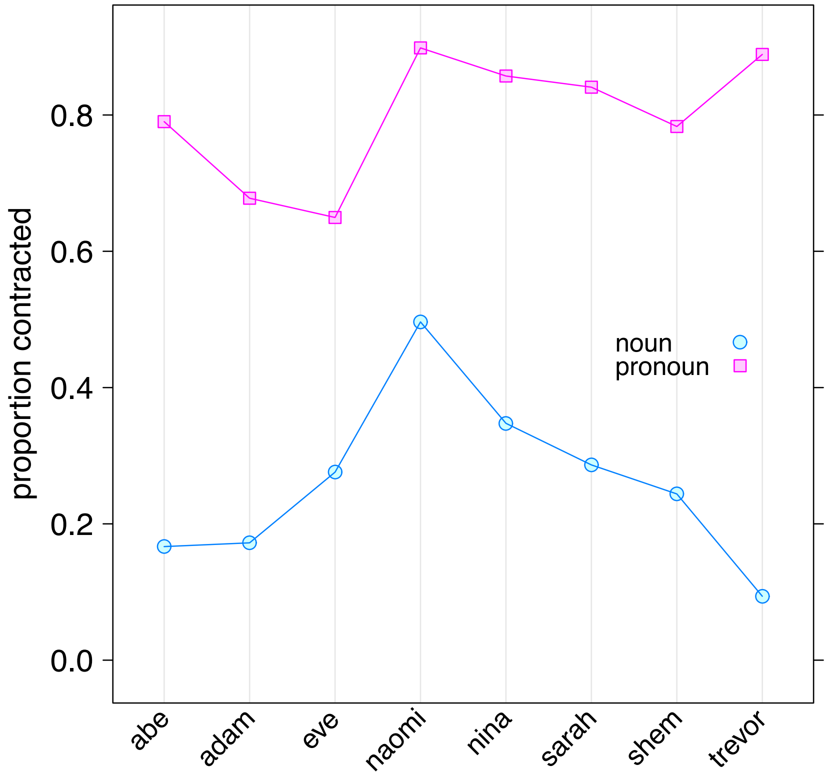


Figure 8: Proportions of contractions with pre-auxiliary nouns and pronouns produced by children in 8 selected corpora

Noun hosts, words for mother and father and *one* Among non-pronoun hosts, words for mother and father (*Mommy, Daddy, mommy, daddy, Papa, Mama, Mom, Dad*) and *one* have the highest conditional probabilities of occurring before *is/'s*. Of these, 43.1% are contracted, compared to the average of 21.1% of all other nouns in the dataset.

These findings support Wilson's (2003, 84) remarks on "chunking with *be*":

> Any particular open-class subject, such as *the pony*, presumably occurs much less frequently than any closed-class subject, so it is proposed that it is much less feasible for the child to abstract constructions such as *the pony's* V-*ing*. However it is plausible that some high-frequency lexical subjects such as *Mommy* and

Daddy might also be learned as units along with *be*. Therefore, to be precise, the claim is not that there is an inherent difference between open- and closed-class subjects in terms of whether they can be chunked with *be*, but rather, chunking should occur much more often with closed-class subjects than it does with open-class subjects.

What is the evidence from other nouns in our dataset? We should not expect a direct mapping from the conditional probabilities of individual pre-auxiliary nouns to their proportions contracted in our data. The reason is that there are so few instances of pre-auxiliary nouns in our dataset; recall (10). While their cooccurrence statistics—the log P(host|aux) values—were derived from the much larger collection of ngrams (Section 3.2), over 57% of the nouns preceding third person singular present tense forms of *be* (*is/'s*) in our 8 selected corpora have a frequency of 1.

Any low-frequency noun host in this dataset might occur once or a few times with contraction, resulting in a higher proportion of contractions than words for *Mommy, Daddy* and *one*. For example, the proper name *Agra* occurs only once, in the utterance *Agra's tired*, making *Agra* 100% contracted before *is/'s* in the data, more than *Mommy* at 51.3%. Yet their cooccurrence probabilities are the reverse: log P(*Agra*|*is/'s*) is less than log P(*Mommy*|*is/'s*), as we saw in the discussion of (8.)

To see the effects of conditional probability of cooccurrence of host and auxiliary on contraction, we must step back from individual data points and look at larger trends in the data. *Agra* falls in the second lowest 25% of the nouns in the dataset in log P(noun|*is/'s*) value. Many of the other nouns in this quartile occur uncontracted. *Mommy*, meanwhile, is in the top 25%. If all the nouns had an equal chance of contracting with *is*, the proportion of contractions would be expected (all else being equal) to be constant across the quartiles of conditional usage probabilities.[12] But if contraction is a function of usage probabilities, we would expect the rate of contraction to rise as the quartiles of log P(noun|*is/'s*) rise.

Therefore if we simply divide the set of unique nouns into quartiles by their log P(noun|*is/'s*) values and examine the overall proportion of contractions in each quartile, we can get a rough picture of the data trend, as shown in Figure 9. The figure shows that as the log P(noun|*is/'s*) values increase, the overall proportion contracted of the nouns within each quartile also increases. Table 2 gives the numbers from which Figure 9 is constructed.

[12] All else is never equal. Bresnan (2021, 132–137) shows by means of a multiple regression model of *is* contraction in adult speech that there is an effect of conditional probability of cooccurrence on contraction after adjusting for multiple other effects. A similar regression analysis of child speech is beyond the space and data limitations of the present study, however.

Table 2: log P(noun|*is*/'*s*) quartiles

quartile ranges:	[−11.6,-10.2)	[−10.2,-9.60)	[−9.60,-8.71)	[−8.71,-4.67)
total types:	191	62	106	110
total instances:	222	116	225	743
total contractions	46	25	64	291
proportion contracted:	0.207	0.216	0.284	0.392

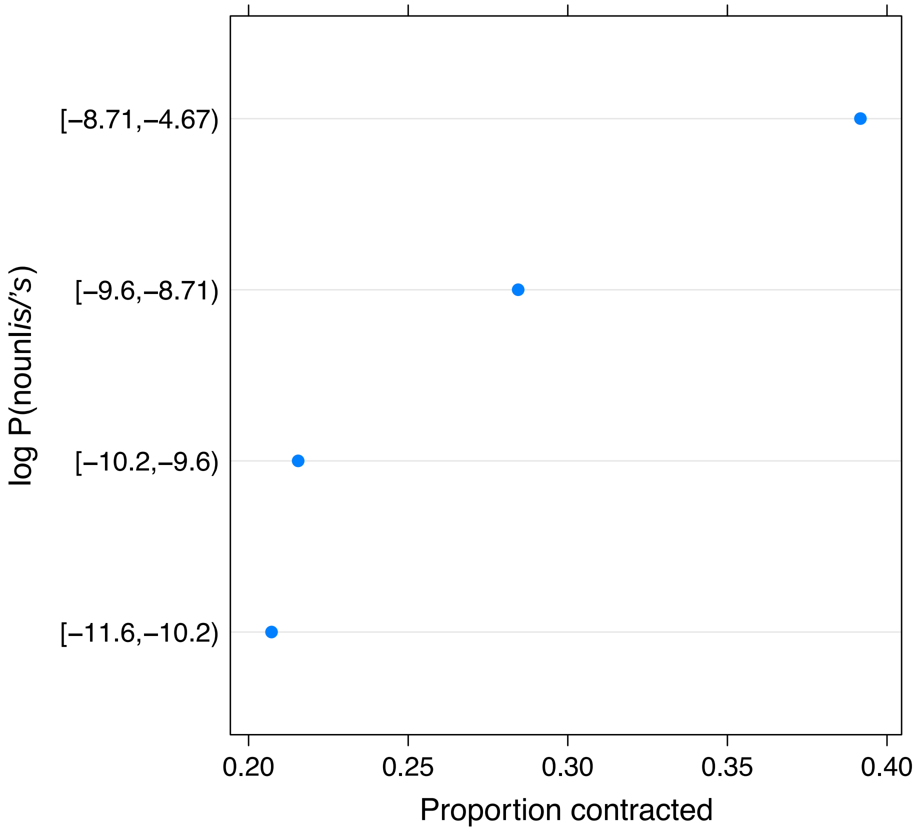


Figure 9: Overall proportions of nouns contracted with *is* by quartiles of log P(noun|*is*/'*s*) in 8 selected corpora

An ANOVA test comparing two logistic mixed-effect models of proportion contracted weighted by the numbers of observations, both including a random effect of child and differing only in the presence of the fixed effects of the quartiles of conditional probability of cooccurrence with *is*/'*s* shown in Table 2, yielded a significant effect of the quartiles on proportion contracted, compared to the hypothesized equality of proportions as the grand mean: $\chi^2(3) = 26.946$,

$p = 6.04 \times 10^{-06}$.

The visually rising trend in proportions contracted shown in Figure 9 was verified by the quartile model itself, the fixed effects of which are given in Table 3. Here the intercept is the mean proportion contracted of the lowest quartile, and for each higher quartile the model contrasts its mean proportion contracted to the mean proportion contracted of all of the previous quartiles. As in Figure 9, the proportion contracted of the second quartile did not reliably differ from that of the first quartile, but each of the higher quartiles differed reliably from those lower than it. Thus, there is a significant overall rise in proportions contracted with the rise in quartiles.

Table 3: Model estimates showing a significant effect of rising quartiles of log P(noun|*is/'s*) on proportion contracted.

	estimate	standard error	Z value	Pr(>|Z|)
intercept	−0.914	0.259	−3.532	0.000
quartile(−10.2, −9.6]	−0.007	0.144	−0.048	0.962
quartile(−9.6, −8.71]	0.161	0.070	2.316	0.021
quartile(−8.71, −4.67]	0.161	0.034	4.733	2.21×10^{-06}

From these results it is reasonable to conclude that in children's speech, as in the adult speech studied by Bresnan (2021), the conditional probability of cooccurrence of sequences of host and auxiliary in usage affects their contraction. This conclusion holds true both for pronouns compared with nouns and within the lexical nouns themselves.

3.4 Contractions as prosodic words

The preceding section showed that where contraction is grammatically possible in the children's data, the proportion contracted is affected by the conditional probability of cooccurrence of the host and auxiliary. In contrast, this and the following section examine where contraction should not be grammatically possible because of the constraints imposed by prosodic words and rightward metrical dependence.

In our dataset all contracted auxiliaries have a leftward host. There are numerous instances like (10a) and none like (10b):

(11) a. *am I a lady ?*
am I going tell Daddy where dis [: this] ball came from ?
has pooped her diaper .
are eating grass .

b. **'m I a lady ?*
**'m I going tell Daddy where dis [: this] ball came from ?*
**'s pooped her diaper .*
**'re eating grass .*

The table in (12) shows the counts of each type:

(12)

	unContracted	Contracted
Host:	7,681	16,089
noHost:	1,500	0

A two-sided exact Fisher test to determine whether the odds of contraction with no host differ from chance yielded a p-value $< 2.2 \times 10^{-16}$ (95% confidence interval = 0.00, 0.00; odds ratio = 0).

Furthermore, unfilled pauses, transcribed as "(.)", appear before and after contractions (13a), but they never break up contractions (13b):

(13) a. *I'm (.) no one .*
it's (.) a house .
they're (.) they're at the beach .

b. **Adam (.) 'll fix the clothesline.*
**the pie (.) 's in the oven .*
**what number (.) 's the hands on ?*

All pre-auxiliary pauses occur with a full auxiliary (14):

(14) *Adam (.) will fix de [: the] clothes+line .*
the pie (.) is in the oven .
what number (.) is the hands on ?

The table in (15) shows the counts of each type:

(15)

	unContracted	Contracted
no preAuxPause:	9,012	16,095
preAuxPause	169	0

A two-sided exact Fisher test to determine whether the odds of contraction with a pre-auxiliary pause differ from chance yielded a p-value $< 2.2 \times 10^{-16}$ (95% confidence interval = 0.00, 0.01; odds ratio = 0).

The required presence of a host of the contracted auxiliary and the absence of pauses or interruptions between them are properties of prosodic wordhood. As these data indicate, the same patterns appear in the children's speech dataset of the present study as in adult speech Bresnan (2021).

3.5 Rightward metrical dependence

Rightward metrical dependence implies that a contracted auxiliary should never occur in the final position of an utterance. Overall, about 93% of utterance-final auxiliaries in the dataset are uncontracted. Counts are shown in (16).

(16)

	unContracted	Contracted
not utteranceFinal:	8,462	16,035
utteranceFinal:	719	54

A two-sided exact Fisher test to determine whether the odds of contraction in utterance-final position differ from chance yielded a p-value $< 2.2 \times 10^{-16}$ (95% confidence interval = 0.03, 0.05; odds ratio = 0.04). Some examples of the expected uncontracted final instances are given in (18).

(17) *dere [: there] it is .*
I don't know where caboose is .
here you is .
so we can know where de [: the] mailman is ?
Dad (.) see how strong I am ?
I will .
I am .
I finded where the swing is .
can you tell what these are ?
this baby is gonna go to the beach like this girl is .

An examination of the relatively small number of exceptional contractions in final position suggests that they may arise from younger speakers who have not fully learned the metrical properties of complete utterances and from incomplete utterances transcribed as complete, arising from the inherent difficulties in defining where a child's utterance ends. See (18a,b) for two examples that violate the rightward metrical dependence of contracted auxiliaries at younger ages.

(18) a. Nina at 1;11.6
MOT: *do you want to find the cow ?*
*CHI: ***here's*** *.*
*MOT: *where's the cow ?*
*CHI: ***here's cow*** *.*
*MOT: *no (.) that's a horse .*
*CHI: *horse .*

b. Nina at 2;5.26
act: nina starts hugging her rubber doll .
*CHI: *he's hugging me .*
*MOT: *who's hugging you ?*
*CHI: ***he's*** *.*
*MOT: *that funny doll ?*
act: nina twists the rubber doll in many shapes .
*CHI: *he [/] he bend .*

Nevertheless, the data sample of exceptions is too small to yield a reliable inferential test of an age effect.

Exceptions to rightward metrical dependence could also arise from incomplete utterances transcribed as complete. The ***utterance*** is the basic syntactic unit in the CHILDES corpora, but the CHAT transcription manual states that it is not always clear where the child's utterance ends. MacWhinney (2000b) observes that whether words the children utter are transcribed as a complete utterance depends on the transcriber's knowledge of their possible constraints on utterance length, their difficulties in saying a word, and the level of syntactic integration they have achieved, among other factors.

For example, in (19) the first line, ending in *I'm*, is transcribed as a complete utterance with the utterance terminator '.'; yet the sentence appears to continue on the next line with the verb *gonna*, which provides a rightward stressed context that allows the contraction.

(19) Adam at 4;5.11
*CHI: *if I finish dese [: these] cutting dese [: these] noodles* ***I'm*** *.*
*CHI: ***gonna*** *have_to +...*

Likewise, in (20) and (21) the final contraction is repeated in the next line, which completes the preceding line marked as a complete utterance:

(20) Sarah at 3;5.07
*CHI: *yeah because* ***I'll*** *.*
*CHI: ***I'll show*** *you how to do it now (.) okay ?*

(21) Trevor at 3;10.2
*CHI: ***or I'll*** *.*
*FAT: *what ?*
*CHI: ***or I'll shoot*** *.*

An extreme example of repetition of a part until completion is (22), where the first four consecutive occurrences of *where's?* are transcribed as complete utterances, violating rightward metrical dependence, although the fifth occurrence of *where's* provides a rightward stressed context that allows the contraction.

(22) Naomi at 3;8.19
*CHI: ***where's ?***
*CHI: ***where's ?***
*CHI: ***where's ?***
*CHI: ***where's ?***
*CHI: ***where's the other truck ?***

In sum, exceptions to rightward metrical dependence of auxiliary contractions might reflect either immature or incomplete utterances, the latter arising from unclarities in determining where a child's utterance ends, but the sample is too small to provide reliable quantitative estimates.

What is that?* vs. *What is it? Apart from the occurrence of a relatively few utterance-final contractions there is further support for the rightward metrical dependence of contraction. Consider children's utterances of two common questions in the dataset: *what is that?* and *what is it?* In the former, contractions are optional, but in the latter, contractions do not occur, as shown in (23).

(23)

	unContracted	Contracted
what is that ?	113	372
what is it ?	189	0

A two-sided exact Fisher test to determine whether the odds of contraction with *what is that?* vs. *what is it?* differ from chance yielded a p-value $< 2.2 \times 10^{-16}$ (95% confidence interval = 0.00, 0.01; odds ratio = 0).

Why is contraction disallowed before *it* but allowed before *that*, when neither is utterance final? The words *that, doing* in *What's that? And what's it doing?* provide rightward stressed elements in a metrically strong complement for *what's* contractions; the subject *it* alone does not, because it is unstressed. In other words, contraction does require a metrically strong complement in these cases of inverted auxiliaries. Bresnan (2021) discusses similar cases in adult speech.

4 Conclusion

In sum, the hybrid model of auxiliary contraction combining LFG and a dynamic exemplar-based lexicon (Bresnan 2021) accounts for four patterns in children's speech—both probabilistic and near-categorical—that closely match those of adults. Pattern 1 is the **usage-based lexicalization** of contractions: the evidence that contractions of auxiliaries with their hosts are acquired as lexically stored chunks in children's speech (Wilson 2003, Pine et al. 2008). Pattern 2 is the positive correlation between host-auxiliary **contractions and their conditional probability of cooccurrence in usage**. This pattern is manifest in the dataset in two ways: first in the contrasts between a large set of 43 closed-class pronouns/pro-forms and lexical nouns; and second, within the lexical nouns themselves, where their quartiles of conditional probabilities before an auxiliary—including words for *Mommy, Daddy* and *one*—correspond positively to the proportions of contractions. Pattern 3 is the requirement that contraction have a host to the left of the auxiliary and no pauses or interruptions between them—properties of **prosodic wordhood** which characterize lexical words. And Pattern 4 is the maturing pattern of host-auxiliary **contractions requiring a metrically strong complement in complete ut-**

terances. The last two patterns follow from connecting the theory of lexical sharing (Wescoat 2005) to prosodic and metrical properties (Bresnan 2021), as outlined in Section 2.

The evidence of the present study shows that children's language, like that of adults, depends on both the usage probabilities of multiword sequences and their prosodic and rhythmic patterns reflecting the syntactic context.

In terms of the developmental debate between constructivists and generativists referenced in Section 3.1, the present framework does not require one to choose sides between the acquisition of lexically specific multiword items and early abstract knowledge of the tense/agreement system. It is a design feature of LFG as a theory of lexical syntax to encode abstract functional information (f-structure) in lexically specific fragments. This design accounts for both the range of syntactic variation across languages and for the ease of breaking linguistic streams into syntactic chunks, referred to as the "fragmentability of language" by Bresnan (2001), Bresnan et al. (2015). What is new in the present hybrid model of LFG is lexical sharing, which allows a single lexical exponent of multiple adjacent syntactic terminal categories, and the usage-based model of the lexicon, which explains the formation and storage of these shared lexical exponents as a function of their conditional probabilities of cooccurrence.

References

Arnon, Inbal, McCauley, Stewart M. and Christiansen, Morten H. 2017. Digging up the building blocks of language: Age-of-acquisition effects for multiword phrases. *Journal of Memory and Language* 92, 265–280.

Barth, Danielle. 2019. Effects of average and specific context probability on reduction of function words *BE* and *HAVE*. *Linguistics Vanguard* 5(1), published online by DeGruyter, 2019-05-08 DOI: https://doi.org/10.1515.

Barth, Danielle and Kapatsinski, Vsevolod. 2017. A multimodel inference approach to categorical variant choice: Construction, priming and frequency effects on the choice between full and contracted forms of *am, are* and *is*. *Corpus Linguistics and Linguistic Theory* 13(2), 203–260.

Bates, Douglas, Mächler, Martin, Bolker, Ben and Walker, Steve. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1), 1–48.

Bender, Emily M. and Sag, Ivan A. 2001. Incorporating contracted auxiliaries in English. In Ronnie Cann, Claire Grover and Philip Miller (eds.), *Grammatical Interfaces in HPSG*, pages 1–15, CSLI Publications.

Börjars, Kersti, Nordlinger, Rachel and Sadler, Louisa. 2019. *Lexical-Functional Grammar. An Introduction*. Cambridge, UK: Cambridge University Press.

Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Oxford: Blackwell.

Bresnan, Joan. 2021. Formal grammar, usage probabilities, and auxiliary contraction. *Language* 97(1), 108–150.

Bresnan, Joan, Asudeh, Ash, Toivonen, Ida and Wechsler, Stephen. 2015. *Lexical-Functional Syntax,* 2nd *Edition*. Blackwell Texbooks in Linguistics, John Wiley & Sons.

Bresnan, Joan and Thráinsson, Höskuldur. 1990. A note on Icelandic coordination. In *Modern Icelandic Syntax*, pages 355–365, Brill.

Brown, Roger. 1973. *A First Language: The Early Stages*. Harvard University Press.

Bybee, Joan. 2001. *Phonology and Language Use*. Cambridge: Cambridge University Press.

Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82(4), 711–733.

Bybee, Joan and Hopper, Paul J. (eds.). 2001. *Frequency and the Emergence of Linguistic Structure*. John Benjamins Publishing.

Clark, Eve V. 1978. Awareness of language: Some evidence from what children say and do. In R. J. A. Sinclair and W. Levelt (eds.), *The child's conception of language*, pages 17–43, Springer.

Clark, Herbert H. and Tree, Jean E. Fox. 2002. Using *uh* and *um* in spontaneous speaking. *Cognition* 84(1), 73–111.

Demetras, Martha Jo-Ann. 1986. *Working Parents' Conversational Responses to their Two-Year-Old Sons*. University of Arizona: PhD dissertation.

Frank, Austin and Jaeger, T. Florian. 2008. Speaking rationally: Uniform information density as an optimal strategy for language production. In *The 30th Annual Meeting of the Cognitive Science Society (CogSci08)*, pages 939–944, Washington, D.C.

Harrell Jr, Frank E. 2021. *rms: Regression Modeling Strategies*. R package version 6.2-0.

Inkelas, Sharon. 1991. *Prosodic Constituency in the Lexicon*. Garland Publications.

Inkelas, Sharon and Zec, Draga. 1993. Auxiliary reduction without empty categories: A prosodic account. *Working Papers of the Cornell Phonetics Laboratory* 8, 205–253.

Isbilen, Erin S., McCauley, Stewart M., Kidd, Evan and Christiansen, Morten H. 2020. Statistically induced chunking recall: A memory-based approach to statistical learning. *Cognitive Science* 44(7), e12848.

Kuczaj, Stan A. 1979. Influence of contractibility on the acquisition of *Be*: Substantial, meager, or unknown? *Journal of Psycholinguistic Research* 8(1), 1–11.

MacKenzie, Laurel E. 2012. *Locating Variation above the Phonology*. University of Pennsylvania: PhD. dissertation.

MacWhinney, Brian. 2000a. *The CHILDES Project: The Database, 3rd Edition*, volume 2. Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, Brian. 2000b. *The CHILDES Project: Tools for Analyzing Talk, 3rd Edition*, volume 1. Mahwah, NJ: Lawrence Erlbaum Associates.

McCauley, Stewart M. and Christiansen, Morten H. 2019. Language learning as language use: A cross-linguistic model of child language development. *Psychological Review* 126(1), 1–51.

Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Joan Bybee and Paul Hopper (eds.), *Frequency Effects and the Emergence of Lexical Structure*, pages 137–157, Amsterdam: John Benjamins.

Pierrehumbert, Janet B. 2002. Word-specific phonetics. *Laboratory Phonology* VII, 101–139.

Pierrehumbert, Janet B. 2006. The next toolkit. *Journal of Phonetics* 4(34), 516–530.

Pine, Julian M., Conti-Ramsden, Gina, Joseph, Kate L., Lieven, Elena V. M. and Serratrice, Ludovica. 2008. Tense over time: testing the Agreement/Tense Omission Model as an account of the pattern of tense-marking provision in early child English. *Journal of Child Language* 35(1), 55–75.

Pinker, Steven. 1996. *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rice, Mabel L., Wexler, Kenneth and Hershberger, Scott. 1998. Tense over time: The longitudinal course of tense acquisition in children with Specific Language Impairment. *Journal of Speech, Language, and Hearing Research* 41(6), 1412–1431.

Rispoli, Matthew. 2016. Cross-morpheme facilitation: The systematic emergence of agreement in 2-year-olds. *Language Acquisition* 23(3), 293–306.

Rispoli, Matthew, Hadley, Pamela A. and Holt, Janet K. 2009. The growth of tense productivity. *Journal of Speech, Language, and Hearing Research* 52(4), 930–944.

Rissman, Lilia, Legendre, Geraldine and Landau, Barbara. 2013. Abstract morphosyntax in two-and three-year-old children: Evidence from priming. *Language Learning and Development* 9(3), 278–292.

Rowland, Caroline F. and Theakston, Anna L. 2009. The acquisition of auxiliary syntax: A longitudinal elicitation study. Part 2: The modals and auxiliary *Do. Journal of Speech, Language, and Hearing Research* 52(6), 1471–1492.

Sachs, Jacqueline. 1983. Talking about the there and then: The emergence of displaced reference in parent-child discourse. *Children's Language* 4, 1–28.

Sadler, Louisa. 1998. English auxiliaries as tense inflections. *Essex Research Reports in Linguistics* 24, 1–15.

Sarkar, Deepayan. 2008. *Lattice: Multivariate Data Visualization with R*. New York: Springer, iSBN 978-0-387-75968-5.

Selkirk, Elisabeth O. 1984. *Phonology and Syntax*. MIT Press.

Selkirk, Elisabeth O. 1996. The prosodic structure of function words. In J. Morgan and K. Demuth (eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, pages 187–214, Lawrence Erlbaum.

Seyfarth, Scott. 2014. Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition* 133(1), 140–155.

Sóskuthy, Márton and Hay, Jennifer. 2017. Changing word usage predicts changing word durations in New Zealand English. *Cognition* 166, 298–313.

Spencer, Andrew. 1991. *Morphological Theory: An Introduction to Word Structure in Generative Grammar*. Oxford: Basil Blackwell.

Spencer, Jessica. 2014. *Stochastic Effects in the Grammar: Toward a Usage-Based Model of Copula Contraction*. Stanford University: PhD dissertation.

Suppes, Patrick. 1974. The semantics of children's language. *American Psychologist* 29(2), 103–114.

Theakston, Anna L. and Lieven, Elena V. M. 2005. The acquisition of auxiliaries *Be* and *Have*: An elicitation study. *Journal of Child Language* 32(03), 587–616.

Theakston, Anna L. and Lieven, Elena V. M. 2008. The influence of discourse context on children's provision of auxiliary *Be*. *Journal of Child Language* 35(01), 129–158.

Theakston, Anna L., Lieven, Elena V. M., Pine, Julian M. and Rowland, Caroline F. 2005. The acquisition of auxiliary syntax: *Be* and *Have*. *Cognitive Linguistics* 16(1), 247–277.

Theakston, Anna L. and Rowland, Caroline F. 2009. The acquisition of auxiliary syntax: A longitudinal elicitation study. Part 1: Auxiliary *Be*. *Journal of Speech, Language, and Hearing Research* 52(6), 1449–1470.

Wescoat, Michael T. 2005. English nonsyllabic auxiliary contractions: An analysis in LFG with lexical sharing. In *Proceedings of the LFG05 Conference*, pages 468–486, online: http://web.stanford.edu/group/cslipublications/cslipublications/LFG/.

Wexler, Ken. 1994. Optional infinitives, head movement and the economy of derivations. In David Lightfoot and Norbert Hornstein (eds.), *Verb Movement*, pages 305–350, Cambridge University Press.

Wexler, Ken. 1998. Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua* 106(1), 23–79.

Wilson, Stephen. 2003. Lexically specific constructions in the acquisition of inflection in English. *Journal of Child Language* 30(1), 75–115.

Revisiting Arabic predicative structures

Maris Camilleri
University of Essex

Proceedings of the LFG'21 Conference

On-Line

Miriam Butt, Jamie Y. Findlay, Ida Toivonen (Editors)

2021

CSLI Publications

pages 51–71

http://csli-publications.stanford.edu/LFG/2021

Keywords: predicative structures, copulas, canonical locatives, inverted locatives, Arabic dialects

Camilleri, Maris. 2021. Revisiting Arabic predicative structures. In Butt, Miriam, Findlay, Jamie Y., & Toivonen, Ida (Eds.), *Proceedings of the LFG'21 Conference, On-Line*, 51–71. Stanford, CA: CSLI Publications.

Abstract

This work brings together the array of predicative structures available across the different Arabic varieties and argues in favour of an analysis that keeps locative predications apart from other vanilla predications on the basis of a number of differing (morpho)syntactic behaviours. While locatives are initially presented as a unified set of structures, they are later differentiated as canonical vs. inverted and treated as two separate constructions. The former is attributed a be_{LOC}<SUBJ,OBL> analysis and the inverted counterpart is here argued to involve a GF - θ-role remapping that renders a be_{LOC}<SUBJ,OBJ> analysis where the *theme* does not function as a SUBJ, but as an unaccusative OBJ; an analysis that is a first of its type in the literature on Arabic and one that challenges the mainstream analysis of this structure, as well as what NOM case identifies in the grammar of the Modern Standard Arabic variety. The analysis for inverted locatives being pursued here in turn predicts and diachronically motivates the otherwise synchronically *ad hoc* constraints that characterise (predicative) BE possessive structures, which are here understood to be direct descendants of inverted locative structures.

1 Introduction

The paper aims to bring to the fore lesser known facts about predicative structures in Arabic and to then focus on highlighting why predicative locative structures stand out from the rest. I will do so by first bringing together in §2 the different predicative/copular structures available across the Arabic dialectal varieties. In §3 I then briefly review the treatment of predicative structures in the LFG literature and point out how predicative locatives appear to have been singled out by the distinct treatment they have received by several proponents. Reinforced by what has already been presented in previous LFG literature, in §4 I provide arguments of both a synchronic and diachronic nature that suggest that Arabic predicative locatives also merit a separate treatment, in contrast to the previous uniform account of Arabic predicative structures in Attia (2008). In §5 I then work out an analysis of the key components of the different predicative locative structures while in §6 I summarise the contributions presented in this study.

2 The nature of predicative structures in Arabic

Predicative structures in Arabic and the interaction with the presence/absence of a copula have received ample attention, even in the typological literature (e.g., Stassen (2009), Pustet (2003)). They have been shown to take PP (1a),[1] [-DEF] AP (1b), [-DEF] NP (1c), AdvP (1d) and CP (e.g., (7b) in §3) predicates with a zero copula or an obligatory copula in non-PRESENT TENSE contexts that is expressed by one of the relevant paradigmatic forms of the copula *kān* 'be', which linearly precedes or follows the SUBJ.

[1]PP predications need not solely be locative in nature. They could for instance express a BELONG sort of clausal possessive structure, where e.g., 'bag of-me' means: 'The bag is mine'.

(1) a. (kān-u) al-awlād (kān-u) **fil-bait**
be.PFV.3-PL DEF-children be.PFV.3-PL in.DEF-house

The children were/are in the house. locative

b. iṭ-ṭālib-a **zakiy-ya**
DEF-student-SGF clever-SGF
The student (F) is clever.

c. iz-zalame **muhandis**
DEF-man engineer.SGM
The man is an engineer.

d. il-ḥafla **bukra**
DEF-party tomorrow

The party is tomorrow. Palestinian

[+DEF] NP predicates (as in (2)) render identificational or specificational predicative structures. In these contexts, some dialects allow for the optional presence of an inflecting 3rd PERSON pronominal copula form that follows the SUBJ, when available, since the SUBJ can be dropped in these structures (Li and Thompson 1977, Eid 1991, Fassi-Fehri 1993, Fassi-Fehri 2012, Ouhalla 2013, Choueiri 2016).[2]

(2) a. Amal Alamuddin **?*ø / hiyye** Amal Clooney
Amal Alamuddin COP.3SGF Amal Clooney

Amal Alamuddin is Amal Clooney. identificational

b. Sami **ø/huwwe** mudīr l-madrase
Sami COP.3SGM director.SGM DEF-school

Sami is the director of the school. specificational - Lebanese: Choueiri (2016, 102)

The use of the negative pronominal copula with which predicates can be negated may either display full agreement in PERSON, NUMBER and GENDER with the SUBJ's CONCORD feature values or take a default form, as the alternation represented via the ∼ (tilde) symbol illustrates in (3) below.

(3) hē **manhāš** ∼ **miš** marēḍ-a
she NEG.COP.3SGF NEG.COP sick-SGF

She is not sick. NEG AP predication - rural Tulkarem

Building further on Stassen (1996), Camilleri and Sadler (2019b, 2020) demonstrate that previous accounts that concentrate on the copula across the Arabic varieties do not fully capture the rich array of what is available. New grammaticalised copulas across the different varieties have emerged, which seem to first target locative predications (4a) as they later diffuse and target more generalised stage-level (4b) and individual level predicative contexts (4c).

(4) a. ti-gul huma **gāʕid-īn** fi magṭaʕ
2-say.IPFV.SGM NOM.3PL COP.3-PL in remote area

[2] In Classical/Modern Standard Arabic, APs and NPs used predicationally are NOM-marked. In the context of the copula *kāna* 'be' (and other similar elements that can partake in this structure), these are ACC-marked.

It's as though they were in a remote area. urban Hijazi: Basulaiman (2018, 32)

b. moḥammed **rā-h** b-xēr
Mohammed COP-3SGM.ACC with-good

Mohammed is well. stage-level predicate - Algerian: Tapiéro (2002, 14)

c. bənāt merdīn kθīr kwās=**ənne**
girl.PL Mardin a lot beautiful.PL=COP.3PL

The girls of Mardin are very pretty. individual-level predicate - Mḥallamiye: Retsö (1987, 221)[3]

The above data constitute instances of vanilla predicative structures. Another set of predicative/copula structures exists and has been discussed in e.g., Soltan (2007), Mohammad (2000), Alharbi (2017) and Alsaeedi (2019). These structures include: predicative locative inversions (5a), a sub-set of clausal possessive structures, which, building on Hallman's (2020) analysis (which differs from previous literature), I here refer to as BE possessives (5b)[4] and existential structures, at least in Classical/Modern Standard Arabic (5c).[5]

(5) a. (kān) ʕind əš-šajara ʕšūš
be.PFV.3SGM at DEF-tree.SGF nest.PL

Near the tree were/are nests. inverted locative - urban Palestinian: Boneh and Sichel (2010, 18)

b. ʕand karīm ḥsāb bəl-bank
at Karim account in.DEF-bank

Karim has a bank account. BE possessive - Syrian: Hallman (2020, 2)

c. hunāka turuq-un kaθīr-a
there way.PL-NOM.INDEF a lot-SGF

There are a lot of ways. existential - Modern Standard Arabic: arabiCorpus

This data set brings together the syntactically predicative or copular structures that are available in Arabic. Their grouping here does not imply that they call for a uniform analysis. Rather, I want to next demonstrate how predicative locatives stand out from the rest of the vanilla predicative structures and that in properly understanding and analysing these structures in the first place, we will then be in a position to better analyse the constructions that have diachronically developed out of them. Before progressing any further I will first in §3 provide an overview of

[3]The enclitic pronominal copula in *qəltu* Arabic dialects has developed as a post-predicative copula influenced by contact with Neo-Aramaic. The grammar of these dialects differs from more mainstream non-*qəltu* dialects and remains heavily underdescribed. The analysis to follow in §5 will unfortunately not incorporate an account of post-predicative copulas.

[4]Without going into much detail here, it suffices to point out that BE possessives are distinguished from HAVE possessives across the Arabic dialects. The latter are not predicative, but transitive in nature and are predicated of a (grammaticalised) verbal element (Comrie 1991, Stassen 2009, Camilleri 2019, Hallman 2020).

[5]While existential structures may be deemed predicational in Modern Standard Arabic, this is not so in the dialects. For this reason, a discussion of existential data will not figure here.

the core literature on the treatment of predicative and copular structures in LFG, on the basis of which I will then in §4 be able to carve out the most adequate analysis for the primary data of interest here.

3 The treatment of predicative structures in LFG

Predicative structures have been given a fair share of attention in LFG. I here first consider the important c-structure considerations to bear in mind and then proceed to f-structure concerns central to predicative structures.

Mainstream LFG is not characterised by pieces of empty syntax at the c-structure level. This does not equate to saying that the absence of such precludes information from still reaching the syntax in one way or another. This can for example be observed in the context of (subject) *pro*-drop and its analytical treatment, where the c-structure does not associate with any piece of tree that stands in for any covert SUBJ element. A similar scenario holds in the context of copulaless structures.

In Arabic and in other languages (see e.g., Stassen (1997), Nordlinger and Sadler (2007)), the absence of a copula *often* contributes morphosyntactic and morphosemantic information associated with the PRESENT TENSE as well as POSITIVE POLARITY values. This information is not accounted for via the lexical entry, unlike the treatment of *pro*-drop. Rather, it is constructionally-specified, i.e. specified via the annotation on the phrase structure rule. Given a sample phrase structure rule such as (6), it is specifically the epsilon (ϵ) notation (Dalrymple 2001) that hosts the information that is realized by the construction in the absence of a c-structure correspondence in I, which is then what gets fed into the f-structure. The epsilon notation is in an either-or relation with the presence of an I node, which in Arabic can be filled by the copula *kān* ‘be’ or the pronominal copula (which fully inflects when expressing negation). The XP following the copula in (6) is meant to refer to any underspecified phrasal category that features as a predicate, including CPs, NPs, APs, PPs and AdvPs.

(6) $$\bar{\mathrm{I}} \rightarrow \left\{ \begin{array}{c} I^{o} \\ (\uparrow \textsc{tense}) = \textsc{non-pres} \\ ((\uparrow \textsc{subj pred}) = \text{‘}\textsc{pro}\text{’}) \end{array} \middle| \begin{array}{c} \epsilon \\ (\uparrow \textsc{tense}) = \textsc{pres} \\ (\uparrow \textsc{pol}) = \textsc{pos} \end{array} \right\} \begin{array}{c} XP \\ \uparrow = \downarrow \big| (\uparrow \textsc{gf}) = \downarrow \end{array}$$

The XP has been here annotated with what reads as an analytical choice between a GF, which would entail that the head of the XP functions as the lexical head/PRED of the GF’s f-structure, or a co-head function. The latter analysis is a possibility based on the fact that the XP in the c-structure functions as a complement to a functional category, namely I (Bresnan 2001).

The ambiguity that characterises the XP annotation draws from the varied analyses predicates or (non-SUBJ) postcopular items have been attributed in LFG. The different analyses can be collapsed into a distinction based on whether the predicative part of the structure (i.e. the XP in (6)) functions as the f-structure’s PRED, i.e. the (lexical) head of the construction, with the copula functioning as a co-head, bearing grammatical, rather than lexical information, or whether it is the copula, irrespective of whether it is present or not, that functions as the f-structure’s PRED.

The former analysis is referred to as the single-tier analysis, as there is no additional f-structure internal to the larger/outer f-structure that would host the head of the predicative phrase separately. Under this analysis, the copula, whether covert or not, solely contributes grammatical information to the structure (Nordlinger and Sadler 2007). On the other analysis, the copula, be it overt or not is taken to function as the structure's PRED, whereby the postcopular XP in (6) functions as a complement to the copula (Rosén 1996, Butt et al. 1999, Dalrymple et al. 2004, Attia 2008). Under this broader characterisation of the copula as the f-structure's PRED, the GF that associates with the non-SUBJ argument of the copula has been attributed varied analyses. Most prominent of these is the distinction between an open vs. closed argument, which translates into the predicate being attributed with an XCOMP or a PREDLINK GF (Dalrymple et al. 2004). While the PREDLINK is a GF that specifically maps onto the predicative complements of copulas, the XCOMP is a non-core GF used elsewhere in the grammar. The distinct nature of the two analyses is meant to account for the differences observed in representative data such as (7). In (7a), the predicative AP displays agreement with the SUBJ, implicative of the functional relation that associates the agreement on the predicate with the f-structure's SUBJ as though the SUBJ is its own. On this analysis, the copula is viewed as a raising predicate, where it does not subcategorise for its own SUBJ. This open complement analysis is however unable to account for the data in (7b), since the SUBJ of the matrix structure differs from the SUBJ within the complement, which hosts a free relative clause. There is thus no functional relation between any of the GFs in the different f-structures.

(7) a. el-bent kān-at nāym-e
DEF-girl be.PFV-3SGF asleep-SGF

The girl was asleep. — adjectival predication
rural Galilean: Mohammad (1998, 4)

b. inti mantīš (i)lli min tūnis
you NEG.COP.2SG COMP from Tunis

You are not the one from Tunis. — SUBJ of matrix ≠ SUBJ of complement
Rammun: Awwad (1987, 116)

Different predicative structures can easily be collapsed under the PREDLINK double-tier analysis as Attia (2008) does when analysing vanilla predicative structures in Arabic, circumventing issues that have to do with the inability to assign an XCOMP GF to the postcopular item without analytically differentiating amongst different predicative structures. There is however one analysis that aligns with the double-tier set of analyses that stands out in accounting solely for (canonical) predicative locatives, and that is: be_{LOC}<SUBJ,OBL$_\theta$>. This analysis has been presented for locative predications in Bresnan and Kanerva (1989), Bresnan (1989, 1994, 2001), Falk (2004), Bresnan et al. (2015) and Sulger (2015) for Bantu, English, Hebrew and Hindi/Urdu, respectively, as well as for locative and existential structures in Hungarian (Laczkó 2012).

This brief summary of both the analyses of predicative structures in LFG and an overview of their different treatments and in which sort of literature provides a snapshot of the fact that predicative locatives in certain LFG analyses have been provided with a distinct analysis that distinguishes them from other sorts of predicative structures. Building on Falk (2004), Camilleri and Sadler (2020) suggest that locatives might in effect be special in Arabic too. Here I will take this proposal a step further as I provide arguments why this is the case and work out its details. I specifically go for a uniform f-structure treatment of predicative locatives that is independent of the absence/presence of the copula. This renders a treatment that differs slightly from Falk's account of the Hebrew counterparts, since his account resorts to a single-tier analysis of predicative locatives in copulaless contexts. As I work my way through the different arguments as to why predicative locatives in Arabic should be analysed differently from other standard predicative structures in §4, I will demonstrate how this ends up predicting the two sorts of locative construals available, based on the structures' varied GF - θ-role mappings associated with the copula *be*$_{\text{LOC}}$ as well as diachronic developments out of such predicative structures.

4 Singling out predicative locatives

In this section I explore certain grounds on the basis of which one could argue that in Arabic too, there is scope to analytically single out locative predications from other standard predicative/copular structures and that the apt analysis is one along the lines of: *be*$_{\text{LOC}}$<SUBJ,OBL$_\theta$> for canonical locatives and *be*$_{\text{LOC}}$<SUBJ,OBJ> for their inverted counterparts.

I here present five different behaviours which distinguish locative predications from the rest of the vanilla predicative structures. These are: a) variation in the SUBJ's DEFINITENESS constraints, b) NEG realization, c) varied copula agreement behaviours, d) variation in the resolution facts that accompany coordinate PP SUBJs, and finally e) a diachronic-based argument that has to do with the fact that locatives primarily stand out as the first targets for emergent copula structures and the fact that they are the only predicative structures that function as precursor structures and bases for further grammatical developments.

Definiteness. The vanilla predicative structures presented in (1) all involved [+DEF] SUBJs. This is in fact the only sort of SUBJ type that is available for them, as illustrated through the ungrammaticality of both an unmodified [-DEF] SUBJ (8a) and a modified one (8b).

(8) a. *binit ḥilw-a
girl sweet-SGF

Intended: A girl is sweet.

b. *binit zɣīr-a ḥilw-a
girl little-SGF sweet-SGF

Intended: A small girl is sweet.[6]

In contrast, the SUBJ of locatives *can* be a [-DEF] (modified or unmodified) SUBJ, yet an unmodified indefinite subject cannot be sentence initial, as the ungrammaticality of (9) illustrates.[7]

[6]The only possible reading available for both the structures in (8) is that of an attributive use of the adjective(s).

[7]The impossibility of a [-DEF] SUBJ appearing in sentence-initial position is a fact that holds true of verbal

A *modified* [-DEF] counterpart *can* however function as a SUBJ of a predicative locative in a sentence initial position at least in certain dialects, as illustrated in (10).

(9) *binit fil-bait
girl in.DEF-house

Intended: A girl is in the house.

(10) **binit zɣīr-a** (qāʕd-a) fil-bait
girl small-SGF COP-3SGF in.DEF-house

A small girl is in the house. rural Tulkarem

Circumventing the constraint that prohibits (unmodified) [-DEF] SUBJs in sentence initial position is easily done in non-PRESENT TENSE structures. Therein, as in (11), the SUBJ ends up appearing in yet another canonical SUBJ position; following the copula – an output that would *still* be deemed ungrammatical in the context of non-locative predicative structures.

(11) **kaan-at** binit fil-bait
be.PFV.3-SGF girl in.DEF-house

A girl was in the house. PAST TENSE

In the PRESENT TENSE, however, the different dialects appear to have at most two possibilities with which to rectify the situation. The first is to maintain the linear order of the constituents where NP $\prec$ PP but where a grammaticalised (erstwhile PP) element *fīh* (or its counterparts in the different dialects) literally meaning 'in-3SGM.GEN' precedes the SUBJ, as in (12). This functions as one of the most common repair strategies across the different Arabic varieties with which to license/salvage a [-DEF] SUBJ in a locative predication.[8] No similar strategy occurs in the context of the other vanilla predicative structures.

(12) **fīh** binit fil-bait
FĪH girl in.DEF-house

A girl is in the house. FĪH insertion

---

predications as well, as illustrated through the ungrammaticality of (i). [-DEF] DFs, represented in (ii) in small caps, understood to sit at the left-periphery of the structure in some SpecCP position are on the other hand accepted. See Fassi-Fehri (1993) and Ouhalla (1997, 1999) for further details on the Modern Standard Arabic data facts.

i *wlād bi-ḥibb-u yi-lʕab-u fuṭbūl
boy.PL BI.3-love.IPFV-PL 3-play.IPFV-PL football
Intended: Boys love playing football.

ii BINIT (kān-at) fil-bait, miš walad
girl be.PFV.3-SGF in.DEF-house NEG boy
A GIRL was/is in the house, not a boy. Palestinian

[8] Albeit somewhat redundant due to its original function, this strategy has with time also infiltrated non-PRESENT TENSE locative structures such as those in (11) and has in some dialects even ended up becoming obligatory. It thus ended up changing its function from one that allowed [-DEF] SUBJs in the absence of any other item that could precede it, to one that more generically licenses the presence of a [-DEF] SUBJ within a locative predication.

The alternative remedy which the Arabic varieties have at their disposal is to change the structure completely, rendering an inverted locative structure as in (13) repeated from (5a). This construction goes part and parcel with the presentational effect it renders, where it involves *in-situ* informational focus that presents the [-DEF] *theme* as new information in the discourse context, with the PP *locative* functioning as the topic, i.e. presupposed/known information.[9] Just as Bresnan and Kanerva (1989) demonstrate for Chicheŵa, in Arabic we similarly find that the consequences of this discourse effect include a [-DEF] restriction on the *theme*, a correlated inability for the *theme* to be expressed as a pronoun and the *theme*'s possibility to be contrastively focussed (14). In this structure (as also happens in the case of structures such as (11) *cf.* ftn. 8) one observes the infiltration of *fīh* (or its equivalents). Depending on the dialect in question, its presence may be obligatory or optional and can precede or follow the PP *locative* so long as it always precedes the NP *theme*.

(13) (kān) ʕind əš-šajara ʕšūš
be.PFV.3SGM at DEF-tree.SGF nest.PL

Near the tree are/were nests. Inverted LOC - urban Palestinian: Boneh and Sichel (2010, 18)

(14) fil-bait WALAD miš binit
in.DEF-house boy NEG girl

In the house there's a boy, not a girl. rural Tulkarem

NEG **realization.** Concomitant with the availability of [-DEF] *themes* in predicative locatives is the morphosyntactic realization of negation in the structure. In non-PRESENT TENSE contexts, the realization of sentential negation ((↑ ENEG) = + (Przepiórkowski and Patejuk 2015)) is across the different copular structures in the vernacular Arabic varieties uniformly expressed via a NEG-realizing inflectional form of the copula *kān* 'be'. In PRESENT TENSE copular structures with [+DEF] *themes*, as illustrated through (3) in §2 and (15) below, sentential negation is expressed via a negative pronominal copula, which, depending on the particulars of the different dialects may involve the use of either default or inflecting forms.

(15) il-binit (lissat-ha) **miš** ∼ **manhāš** fil-bait/mara
DEF-girl still-3SGF.GEN NEG.COP NEG.COP.3SGF in.DEF-house/woman

The girl is not yet in the house/a woman. NEG pronominal copula - rural Tulkarem

Pronominal negation is however *not* available in the context of [-DEF] *themes*, i.e., in the context of canonical and inverted locative structures (and by extension BE possessives). Rather, (↑ ENEG) = + is expressed via the NEG-realizing inflectional counterpart of *fīh*, which takes the form of: *mā fī(š)*, *fīš* or *fišš*, depending on the dialect, at the exclusion of e.g., *miš*, as in (16).[10]

[9]Note that the use of 'topic' here should not be understood as the grammaticalised/f-structure DF label, as in this context the PP bears no DF role. Rather, its postcopular position in Arabic is a canonical GF position. Reference to topic here aligns with the information-structure TOPIC which is composed out of the [-NEW] [+PROM] feature values in Butt and King's (1996) geometry of information features. The *theme* in this structure's presentational nature takes on the [+NEW] [+PROM] FOCUS feature set.

[10]Without going into much details here, but mostly following a particular segment of the literature (given that

(16) a. **fī-š** / *miš binit fil-bait
FĪH-NEG / NEG.COP girl in.DEF-house

A girl is not in the house. NEG canonical locative

b. fil-bait **fī-š** / *miš binit
in.DEF-house FĪH-NEG / NEG.COP girl

In the house there isn't a girl. NEG inverted locative - Palestinian

Copula agreement. When it comes to copula agreement, key to our data is that as illustrated through (1a), for instance, the verbal copula fully agrees in PERSON, NUMBER and GENDER with the [+DEF] SUBJ's CONCORD feature values. The verbal copula in locatives with a [-DEF] *theme*, whether inverted or not, displays either full or default 3SGM agreement, depending on the dialect. Default agreement is the most widespread strategy across the dialects. The paradigmatic data set in (17) comes from rural Galilean, which happens to be one of those few dialects that still allow for full copula agreement with the [-DEF] *theme*. The data illustrate two word order variations of the canonical locative predication and demonstrate additional agreement nuances therein. In (17b) we further observe how in this particular dialect the [-DEF] *theme* can precede the copula (so long as the *theme* is itself preceded by *fīh*) and when this is the case, only full agreement is possible on 'be'.

(17) a. kān-u ∼ kān fīh xams zlām bed-dār
be.PFV.3-PLM be.PFV.3SGM FĪH five man.PL in.DEF-house

Five men were in the house. p. 50

b. fīh xams neswān kān-en / *kān bed-dār
FĪH five woman.PL be.PFV.3-PLF be.PFV.3SGM in.DEF-house

Five women were in the house. canonical locative - p. 51

c. kān-u ∼ kān ʕen-na xams zlām
be.PFV.3-PLM be.PFV.3SGM at-1PL.GEN five man.PL

Five men were at our place. inverted locative - rural Galilean: Mohammad (1998, 52)

PP coordinate conjuncts and resolution. The next varied sort of morphosyntactic behaviour has to do with the observation that coordinated PPs display distinct behaviours in locative vs. other structures (whether predicative or verbal). The data to be presented serves a dual function in that it also ends up rendering itself as a test for PP subjecthood in Arabic, which is essential in the analysis of inverted locatives.

Testing the subjecthood of PPs in Arabic is possible by for instance observing their behaviour in raising structures; a test that has recently become available for use in Arabic following the analysis of a number of relevant structures in ElSadek and Sadler (2015) and Camilleri and Sadler (2019a). I here make use of one of their predicates – *šakl*, whose literal meaning is 'form,

there are varied treatments of *fīh*), namely Halila (1992), Eid (1993) and Hallman (2020), *fīh* is essentially treated as a (vacuous) verbal element whose grammaticalised verbal status is best evinced and reinforced through its ability to realize NEG, as in (16). An analytical treatment of *fīh* within LFG will be pursued in §5.

shape', but has grammaticalised a verbal function with the meaning 'seem, appear'. In (18), *šakl* 'seem, appear' heads the matrix clause which embeds a PAST TENSE locative predication in its complement and the embedded clause's locative PP surfaces in the matrix in a preverbal position. (A post-verbal position would have been just as appropriate). The inflection on *šakl* is the default 3SGM, as is the marking on the copula *kān* 'be' in the embedded clause, which in turn provides additional support that the embedded locative predication is an inverted one. Full agreement on the copula would have been expressed by the 3SGF *kān-at*, since the *theme* is an inanimate PL NP.

(18) fuq il-xizane šakl-u [kān flūs kθiyr
on DEF-wardrobe.SGF seem-3SGM.GEN be.PFV.3SGM money a lot
mu-xbiy-ya]
PASS.PTCP-hide-SGF

On the wardrobe seems to have been a lot of hidden money. rural Tulkarem

To further determine that a locative PP can indeed function as a SUBJ, including the SUBJ of a raising structure, as in (18), I demonstrate a more transparent structure involving coordinated PP locative arguments, especially in order to further determine that the 3SGM marking on the matrix in (18) is not meant to imply that the structure should be interpreted as an *it*-expletive type of construction (and hence not involving raising at all). Within the adjectival predication (19a) and the equative predication (19b) below, we find 3PL resolution both on the matrix raising predicate as well as on the PAST TENSE 'be' (19a) and pronominal (19b) copulas within the embedded clause.

(19) a. [fuq il-xezāne]$_i$ u [fi qāʕ il-bīr]$_j$ šakla-hum$_{i+j}$
on DEF-wardrobe.SGF CONJ in bottom.SGM DEF-well.SGM seem-3PL.GEN
[kān-u$_{i+j}$ malyan-ēn flūs]
be.PFV.3-PL full-PL money.PL

On the wardrobe and in the bottom of the well seem to have been full of money.

b. [fūq il-xezāne]$_i$ u [fī qāʕ il-bīr]$_j$ šakla-hum$_{i+j}$
on DEF-wardrobe.SGF CONJ in bottom DEF-well seem-3PL.GEN
hummi$_{i+j}$/*hu ʔaħsan taxmēn il-i weyn li-flūs
COP.3PL/COP.3SGM good.ELAT guess.SGM to-1SG.GEN where DEF-money.PL
mumkin t-kūn t-xabb-at
perhaps 3F-be.IPFV.SG PASS-hide.PFV-3SGF

On the wardrobe and in the bottom of the well seem to be my best guess as to where the money may be hidden. rural Tulkarem

In contrast to the 3PL resolution observed in the context of coordinated PP SUBJs in (19), a counterpart to the locative predication in (18) involving coordinated PPs, as in (20), does not result in a similar behaviour. Rather, the matrix raising predicate and embedded copula maintain a 3SGM default form, as in (18).

(20) [fuq il-xezāne]$_i$ u [fi qāʕ il-bīr]$_j$ šakl-u / *šakla-hum
on DEF-wardrobe.SGF CONJ in bottom DEF-well seem-3SGM.GEN / seem-3PL.GEN
kān / *kān-u flūs kθiyr mu-xbiy-ya
be.PFV.3SGM / be.PFV.3-PL money.PL a lot PASS.PTCP-hide-SGF

On the wardrobe and in the bottom of the well, there seem to have been a lot of hidden money. rural Tulkarem

The above data demonstrate that PP *locatives can* function as SUBJs in Arabic. They additionally shed light on a contrast that holds between PPs as SUBJs of an inverted locative and PPs as SUBJs in other predicative clauses: The latter clearly trigger agreement, as evinced through the 3PL resolution in (19) in the context of raised coordinated SUBJs, while PPs in inverted locatives *do not*, as the ungrammaticality of the resolved argument in (20), demonstrates. I take this to suggest that the 3SGM agreement in inverted locatives results from the non-canonical mismatch that results, whereby the logical subject, i.e., the highest thematic argument does not map onto the highest GF in the structure. This argumentation also extends to the copula agreement facts presented in (17c) above, given that the highest GF in the inverted locative, i.e., the SUBJ does not happen to map onto the highest θ-role, i.e., the *theme* argument.

Diachronic-oriented motivations. The final points of divergence that distinguish locative predications from other vanilla predications are diachronic in nature. The first has to do with the fact alluded to in §2, where somehow, the emergence of new copulas across the different Arabic dialects, independent of the type of grammaticalised copula strategy that is involved, has targeted locative structures across the board. While the copula has also infiltrated other predicative structures, in particular ones with stage-level predications, this is only true of certain dialects (Camilleri and Sadler 2019b). Locative predications thus clearly stand out as earlier targets for copula emergence. The second diachronic point to be made is the fact that locative predications turn out to be the *only* (non-grammaticalised) predicative structures that have led to further grammaticalisations, yielding the development of existential and possessive structures.[11]

I take the above presented set of arguments to provide us with ample grounds on the basis of which to suggest that predicative locative structures merit their own separate analysis in Arabic. Beyond that, however, there are a number of further ramifications on the grammar at large, particularly if we were to concentrate on both the synchronic and diachronic syntax of existential and possessive structures in Arabic. Space and scope constraints restrict me from engaging into this in any detail, yet it suffices to state here that the analysis of inverted locative structures along the lines being argued for here predict and determine, without any need to resort to *ad hoc* constraints, both the syntax of, and the morphosyntactic conditions on BE possessives such as (5b), which are predicational structures, and which I take to be direct developments specifically out of inverted locatives.

[11] None of these grammaticalised structures make use of any of the newer-type copulas that have targeted locative predications across the larger Arabic macrosystem. This further supports the view that copula emergence has taken place at a much later stage in the system.

5 Working out an analysis

For canonical locative structures, the analysis being argued for here is one where the copula functions as a two-place predicate with both its arguments, i.e. the NP *theme* and the PP *locative* mapping onto two core GFs. Couched within standard Lexical Mapping assumptions (Bresnan and Kanerva 1989, Bresnan and Zaenen 1990) that couple an argument ranking hierarchy with the ranking of the [-/+r(estricted)]/[-/+o(bjective)] feature values that compose the core GFs, the *theme* gets intrinsically identified as a [-r] argument, while the *locative* is identified as [-o]. Well-formedness constraints result in the *theme*'s mapping onto the SUBJ GF, as represented in Table 1. Since I am here assuming a uniform analysis of predicative locative structures that is independent of a copula in the structure, Table 1 also incorporates a representation of null-be_{LOC}.

be$_{\text{LOC}}$/null-***be***$_{\text{LOC}}$	<	**arg 1** *theme*	**arg 2** *locative*	>
		[-r]	[-o]/[+r]	
		SUBJ	OBL	

Table 1: The θ-role - GF mapping in **canonical predicative locatives**

When compared with the analysis for canonical counterparts in Table 1, accounting for the inverted locative facts as they stand for Arabic (which find parallels in non-predicative counterparts too) constitutes a transparent instance of a θ-role - GF mapping reversal; something which is not an obvious possibility were we to analyse predicative locatives as involving a closed double-tier PREDLINK analysis as previous work has done, thus resulting in the loss of generalisations over locative structures at large. Following Kibort (2007) and her analysis grounded in the markedness hierarchy of the decomposition feature values, the *theme* argument in non-canonical locative structures, while maintaining its inherent [-r] value gets assigned a [+o] (see Table 2), which in turn functions as a 'mechanism of increasing markedness' (p. 267) and thus gets mapped onto an OBJ. This is in line with its unaccusative OBJ status in the grammar; i.e., an OBJ that can alternate with a SUBJ function in certain intransitive contexts. It also aligns with the added information-structure load which the inverted locative expresses when compared to its canonical counterpart. The *locative* is then available to map onto the SUBJ function, which constitutes the highest (and least marked) compatible function. The alternation these two locative structures display illustrates how in Arabic, there are multiple BE lexical entries. More specifically, there are two different mappings available in the context of the BE_{LOC} copula; each with its different requirements, as will be shown in their respective lexical entries in (31) and (33).

The OBJ function the *theme* ends up associating with in inverted locative structures is by no means the usual or canonical one. For starters, since the structure also happens to express presentational focus, as made reference to in §4, this particular OBJ must be [-DEF] and non-pronominal. Unlike canonical OBJs it cannot be passivised or relativised upon either. Although more work needs to be done, a preliminary investigation of the Arabic data suggests that such behaviours hold true of unaccusative OBJs in structures involving inverted locatives in general.

be$_{\text{LOC}}$/null-***be***$_{\text{LOC}}$	<	**arg 1**	**arg 2**	>
		theme	*locative*	
		[-r]	[-o]	
		[+o]		
		OBJ	SUBJ	

Table 2: The θ-role - GF mapping in **inverted locative predications**

What unifies the predicates in such syntactic contexts is their unaccusative nature. At this juncture it is worth making reference to data from Classical/Modern Standard Arabic to ensure that all potential issues are dealt with, in the hope of reaching a true comprehensive understanding, especially since predicative inverted locatives in the Arabic literature have not been treated in the way they are being analysed here. In the varied analyses provided, the PP is treated as having scrambled into a position that precedes the *theme* from its usual position in canonical locative structures, but where importantly, the *theme* is nonetheless deemed as maintaining its SUBJ function within the structure (Soltan 2007, Alharbi 2017, Alsaeedi 2019). Key to the data is the fact that in non-vernacular Arabic, NPs are CASE-marked, and as observed in (21) below, the *theme* in the inverted locative maintains the NOM-marking as otherwise present on the *theme* in the canonical counterpart. It has been this NOM-marking (even within the context of a *kāna* 'be') that appears to have led to this seemingly uncontroversial/unchallenged analysis of the *theme* as the structure's SUBJ, even if the agreement facts observed on the copula, for instance, are not consistent with a context in which the *theme* is the structure's SUBJ.

(21) kāna ∼ kān-at fī ʔal-bayt-i ʔimraʔat-**un**
be.PFV.3SGM be.PFV.3-SGF in DEF-house.SGF-GEN woman.SGF-NOM.INDEF

A woman was in the house. Modern Standard Arabic: Soltan (2007, 111)

To be able to challenge the previous literature is to first determine that PPs can function as SUBJs in Arabic. This has been evinced in §4 through their ability to partake in structure-sharing within SUBJ-to-SUBJ raising constructions and their linear positioning in canonical pre- and post-verbal SUBJ positions. Secondly, the revisiting of something more basic is required, and that is: the function of NOM CASE in Arabic. That CASE does not always align in a one-to-one relation with any one given GF is well-known (e.g., Mohanan (1982)), and in effect this is quite clear in the Arabic dialectal system at large, where e.g., SUBJs can be cross-referenced by ACC and DAT pronominal forms incorporated on the verb. The proposal being put forward here is that NOM CASE in Arabic may be either informationally-grounded or assigned to the highest available nominal GF. The former is illustrated through (22), where the grammatical TOPIC is NOM-marked yet then bound by an ACC resumptive pronoun functioning as the OBJ. That NOM happens to align with the SUBJ GF is itself an artifact of the SUBJ's prototypical expression as a NP and which NP happens to additionally function as a DF of sorts (Bresnan 2001).

(22) ʔal-riwāyat-**u**$_i$ ʔallaf-at-ha$_i$ zaynab-u
DEF-novel.SGF-NOM write.PFV.3-SGF-3SGF.ACC Zaynab-NOM

(As for) the novel, Zaynab wrote it. Modern Standard Arabic: Ouhalla (1997, 12)

Under the proposal being made here, in the context of inverted locatives (21), as is also the case in BE possessives (23a) and structures that take PP *experiencers* (23b), since the SUBJ happens to be non-canonically expressed by a PP, NOM-marking simply gets assigned to the next available NP in the structure, which happens to be the OBJ, resulting in the type of copula (and verbal) agreement mismatches discussed earlier in §4 in light of (17c).

(23) a. kāna ∼ kān-at ʕinda ʔal-ʔawlād-i sayyarat-**un**
be.PFV.3SGM be.PFV.3-SGF at DEF-boy.PL-GEN car.SGF-NOM.INDEF

The boys had a car. BE possessive - p. 111

b. ya-ğibu ∼ ta-ğibu ʕala ʔal-muʔmin-īn
3-must.INDIC.SGM 3F-must.INDIC.SG on DEF-believer-PLM
ʔaṣ-ṣalāt-**u**
DEF-praying.SGF-NOM

The believers have to pray. PP *experiencer* $\prec$ NP *theme* - Modern Standard Arabic: Soltan (2007, 109)

With that additional bit of background in place, we move on to account for other bits of structure within locative predications. The first is to cater for *fīh*. *fīh* is essentially syntactically required in the string in the context of a [-DEF] *theme* yet bears no semantic contribution, unlike its NEG-realizing counterpart, and will solely be associated with a FORM feature (Bresnan 1982).[12] The f-structure associated with the canonical locative in (24) (which is essentially a PAST TENSE version of (12) above) is presented below.

(24) kān **fīh** binit
be.PFV.3SGM FĪH girl.SGF
fil-bait
in.DEF-house
A girl was in the house.

$$\left[\begin{array}{ll} \text{PRED} & \text{`BE<SUBJ,OBL}_\theta\text{>'} \\ \text{TENSE} & \text{PAST} \\ \text{FORM} & \text{FĪH} \\ \text{SUBJ} & \left[\begin{array}{ll} \text{PRED} & \text{`GIRL'} \\ \text{DEF} & - \end{array}\right] \\ \text{OBL}_\theta & \left[\begin{array}{ll} \text{PRED} & \text{`IN<OBJ>'} \\ \text{OBJ} & \left[\begin{array}{ll} \text{PRED} & \text{`HOUSE'} \\ \text{DEF} & + \end{array}\right] \end{array}\right] \end{array}\right]$$

The *abridged* set of rules presented in (25-29) below are laden with analytical stock which I will unpack here. [13] *fīh* is instantiated as a V that is attributed with a FORM feature (as illustrated in

[12] *fīh* occurs in the context of other [-DEF] *theme*-taking structures including possessive constructions, unaccusative intransitives and non-agentive transitives such as psychological and experiencer verbs; the latter two contexts have not been previously mentioned in the literature, when the distribution of *fīh* is discussed. In all these contexts, unlike in locative predications, *fīh* is by no means obligatory. The NEG counterpart, on the other hand, displays a distinct distribution. As alluded to in §2, *fīh* in existential structures should not be conflated with the function of *fīh* in any of the structures mentioned above and ought to be analysed in its own right.

[13] To be more explicit, the following are the key bits that have been left out from the set of phrase structure rules presented here: 1. The VP rule in (27) lacks reference to optional material that may linearly follow the $\bar{\text{V}}$, 2. The $\bar{\text{I}}$ and $\bar{\text{V}}$ rules in (26) and (28), respectively, are here being represented without reference to the fact that an optional $\widehat{NEG}$ may precede the I and V nodes, 3. The $\bar{\text{I}}$ rule in (26) lacks reference to $\underset{\uparrow = \downarrow | (\uparrow \text{GF}) = \downarrow}{XP}$ material (present

its corresponding lexical entry presented in (34) below). The *fīš* counterpart is treated as a NEG FORM and its lexical entry comes along with the existential constraint (← ENEG) = +, which makes reference to the fact that in the f-structure where the NEG FORM feature is, ENEG is also an attribute therein with value +. (↑ ENEG) = + in the structure is then expressed either by *fīš* itself or in tandem with other pieces of syntax, e.g. *mā* as part of a bi-partite NEG realization, depending on the dialect (Camilleri and Sadler 2017). As the phrase structure rules demonstrate, *fīh* is allowed to co-occur with *kān* 'be', yet in a context where (↑ ENEG) = + is expressed by the copula (pronominal or verbal), the presence of a NEG FORM is excluded.

The I node is in a complementary distribution with the ϵ and in the absence of I, the TENSE value can only be PRESENT (see e.g., Nordlinger and Sadler (2007)). The absence of a copula in I implies other things in Arabic. As the data presented in this study illustrate, the availability of a [-DEF] SUBJ in such contexts obligatorily requires the presence of *fīh*. What the absence of a copula *does not* imply, in Arabic, despite a number of previous claims in the literature, is that the structure is POL = + (since the negative pronominal copular form is assumed to occupy a position in I when available). It has here been demonstrated through the data contrasts presented in (15) and (16) in §4 that (↑ ENEG) = + can *still* be expressed, even within a copulaless structure. It is thus for this reason that (↑ ENEG) = – is represented only as an *optional* possibility under the ϵ. In a context where a copulaless structure *does* express (↑ ENEG) = +, then this must obligatorily be a context where a [-DEF] SUBJ or OBJ (generalised as (↑ MINUSR)) is present as well as the NEG FORM *fīš*.[14] Finally, the V node in (28), which includes *fīh*, replicates the information otherwise available in the lexical entry. The constraint that determines the distribution of *fīh* in its use in canonical and inverted locative predications (and by extension BE possessives) makes reference to the a-structure - f-structure correspondence assumption in Butt et al. (1997).[15] The rule once again generalises over the SUBJ and OBJ GFs and as dictated perhaps more clearly in the lexical entry in (34), the presence of *fīh* is part and parcel of a structure that must involve a [-DEF] (↑ MINUSR) and that this GF must in turn correspond with a *theme* argument.

(25) $\text{IP} \rightarrow \left(\begin{array}{c} \{NP \mid PP\} \\ (\uparrow \text{SUBJ}) = \downarrow \end{array} \right) \quad \begin{array}{c} \bar{I} \\ \uparrow = \downarrow \end{array}$

(26) $\bar{I} \rightarrow \left\{ \begin{array}{c} I \\ ((\uparrow \text{TENSE}) = \text{NON-PRES}) \\ ((\uparrow \text{ENEG}) = +) \\ \neg(\uparrow \text{NEG FORM}) \\ \end{array} \middle| \begin{array}{c} \epsilon \\ (\uparrow \text{TENSE}) = \text{PRES} \\ ((\uparrow \text{ENEG}) = -) \\ (\uparrow \text{SUBJ DEF}) = - \rightarrow (\uparrow \text{FORM}) =_c \text{FĪH}\_ \\ \left[\begin{array}{c} (\uparrow \text{MINUSR DEF}) = - \\ (\uparrow \text{ENEG}) = + \end{array} \right] \rightarrow (\uparrow \text{NEG FORM}) =_c \text{FĪŠ(Š)}\_ \end{array} \right\}$

in rule (28)) that follows I, and 4. The V node in (28) does not represent the otherwise additional availability of the neg-counterpart of $f\bar{\imath}h$.

[14]This constraint kills two bird with one stone and holds not only true of predicative locatives with a [-DEF] *theme*, which depending on the canonical vs. inverted nature of the predication, map onto a SUBJ or OBJ GF, respectively, but also of BE possessives, which as alluded to in the end of §4 are here analysed as direct developments out of inverted locatives and similarly involve the mapping of a [-DEF] *theme/possessed* argument onto an OBJ. The constraint also holds true of the distribution of *fīh* in generalised unaccusative verbal contexts, be they in/transitive. For those dialects in which negation is solely expressed by *mā* along with *fīh* without the use of any designate NEG FORM, modifications in the stipulation of the rules and the lexical entry would have to follow accordingly.

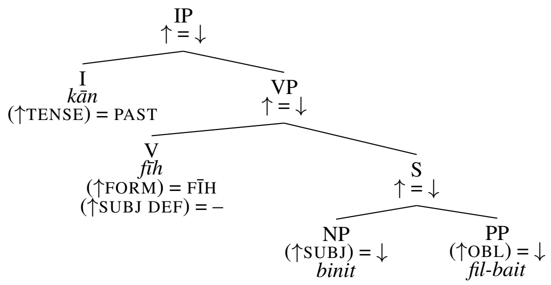
[15]If alternative correspondences in e.g., Asudeh and Giorgolo (2012) and Findlay (2017) were to be employed, while the nature of how things are stated would be somewhat different, the morphosyntactic conditions that underpin the distribution of *fīh* would however remain the same.

(27) $VP \rightarrow \underset{\uparrow = \downarrow}{\bar{V}}$

(28) $\bar{V} \rightarrow \underset{(\uparrow \text{FORM}) = \text{FĪH\_}}{V} \rightarrow \left[\begin{array}{c}(\uparrow \text{MINUSR DEF}) = - \\ (\uparrow \text{MINUSR})_{\lambda}{}^{-1} =_c \text{THEME}\end{array}\right] \quad \underset{\uparrow = \downarrow | (\uparrow \text{GF}) = \downarrow}{XP}$

(29) $\text{S} \rightarrow \left\{\begin{array}{c} NP \mid PP \\ (\uparrow \text{SUBJ}) = \downarrow \end{array}\right\} \left(\begin{array}{c} YP \\ \uparrow = \downarrow | (\uparrow \text{GF}) = \downarrow \end{array}\right)$

On the basis of the above rules, the c-structure associated with (24) is provided below.



(30)

For completeness, the different lexical entries associated with *kān* when this functions as a *be*$_{\text{LOC}}$ copula are provided in (31) and (33) below.[16] Since the requirements of *fīh* in canonical locative structures differ from dialect to dialect, the last constraint in (31) might need to be further refined accordingly, whereby resort to the *fīh* strategy in the structure is only necessary if the SUBJ is not modified (32).[17]

(31) *kān: I* (↑ PRED) = ‘be$_{\text{LOC}}$ <SUBJ, OBL>’
(↑ TENSE) = PAST
(↑ SUBJ DEF) = – → (↑ FORM) $=_c$ FIH_

(32) $\left[\begin{array}{c}(\uparrow \text{SUBJ DEF}) = - \\ (\uparrow \text{SUBJ}) \\ \neg(\rightarrow \text{ADJ})\end{array}\right] \rightarrow (\uparrow \text{FORM}) =_c \text{FIH\_}$

Similarly, in (33), the lexical entry of BE$_{\text{LOC}}$ in inverted locative contexts, the optionality of *fīh* in the structure is once again dependent on the dialect in question and may additionally be determined by the structure’s TENSE value. Here BE$_{\text{LOC}}$ is specified as taking a SUBJ of a PP c-structure category. The constraint stipulating the NP within the SUBJ PP to be [+DEF] is an important constraint that characterises PP SUBJs in Arabic. It differentiates them from predicative PP functions, in which the NP complement can be [+/-DEF]. (34) represents the lexical entry for the grammaticalised *fīh* as employed in predicative structures (and beyond).

[16] The lexical entries do not make reference as to how agreement gets worked out. This will heavily depend on the variety involved. In non-default inflecting *kān* contexts, there is a canonical display of agreement with the SUBJ’s CONCORD feature values. In the context of full agreement within inverted locative structures, however, then agreement in that context must be stipulated in the relevant lexical entry as involving agreement with the OBJ’s CONCORD feature values, at least in the case of those varieties that still display full agreement with the *theme*.

[17] To account for the differences between locative predications and e.g., adjectival predications requires either the assumption that in the latter structures the copula is solely a feature-bearer and the adjective functions as the f-structure’s PRED or that the copula similarly functions as the f-structure’s head yet associates with (yet another) distinct subcategorisation frame (and hence, lexical entry). Instead of an OBL, the copula would take a PREDLINK, with the adjective functioning as the latter’s head.

(33) *kān: I* (↑ PRED) = 'be$_{LOC}$ <SUBJ, OBJ>'
(↑ TENSE) = PAST
CAT((↑ SUBJ), {PP})
(↑ SUBJ OBJ DEF) = +
(↑ OBJ DEF) = –
((↑ FORM) $=_c$ FIH$_{-}$)

(34) *fīh: V* (↑ FORM) = FIH
((FORM ↑) MINUSR DEF) = –
(↑ MINUSR)$_{\lambda}^{-1}$ $=_c$ THEME

6 Conclusion

As I revisited predicative locatives in Arabic within the realms of LFG I presented arguments as to why locative predications may best be analysed distinctly from other vanilla counterparts. It adds to the LFG literature that treats canonical locatives as structures involving a SUBJ and an OBL. No distinction was made here between locatives with/without a copula. The analysis provided for canonical locatives has opened a much needed quest to tackle the analysis of locative inversions in Arabic. I have here solely provided initial grounds on the basis of which one can argue that we are indeed dealing with a predicative construction that must be intrinsically differentiated from its canonical counterpart. Evidence for the SUBJ function of PPs was provided via reference to the agreement facts, which I take to be a display of a mismatch between the highest thematic role in the structure and the highest GF ($\hat{\theta} \neq \widehat{\mathrm{GF}}$), as well as the important raising facts. In challenging previous literature on Arabic when it came to the analysis of inverted locatives, I have demonstrated how NOM CASE in Modern Standard Arabic need not always be understood to align with a SUBJ GF. Rather, in this case and in similar copular and verbal constructions that take PP SUBJs, NOM-marking simply gets assigned to the next available nominal GF in the structure, thus non-canonically appearing on OBJ GFs.

Nailing down the analysis of Arabic locatives ends up having important ramifications on the grammar, when one considers the diachronic developments of existential and possessive structures out of locative predications. The analysis of inverted locatives here provides diachronic weight to Hallman's (2020) synchronic account of BE possessives in Syrian Arabic. It in turn challenges previous typological literature that has solely stated that Arabic predicative locatives serve as precursors of (HAVE) possessives (by presenting *canonical* locative examples to illustrate their point). On the basis of the chained nature of the argument being developed here and concomitant with the analysis provided here for inverted locatives, it has specifically been *inverted* locatives that functioned as precursors to possessive structures, and the first to have developed were BE possessives. (HAVE counterparts only developed from BE ones later on). It is with this background that one can come to appreciate the edge that a given treatment may end up attaining, when the analysis of a particular structure is not solely viewed narrowly in and of itself, but rather informed by, and analysed as part of a tapestry of intertwined synchronic facts and diachronic developments within the grammar.

References

Alharbi, Bader Yousef. 2017. *The Syntax of Copular Clauses in Arabic*. Ph. D.thesis, University of Wisconsin-Milwaukee.

Alsaeedi, Mekhlid. 2019. *Existential and Negative Existential Constructions in Arabic: Typology and Syntax*. Ph. D.thesis, Arizona State University.

Asudeh, Ash and Giorgolo, Gianluca. 2012. Flexible composition for optional and derived arguments. In Miriam Butt and Tracy H. King (eds.), *Proceedings of the LFG12 Conference*, pages 64–84.

Attia, Mohammed. 2008. A unified analysis of Copula Constructions in LFG. In Miriam Butt and Tracy H. King (eds.), *Proceedings of LFG08*, pages 89–108, Stanford: CSLI Publications.

Awwad, Mohammad A. 1987. Free and bound pronouns as verbs in rural palestinian colloquial arabic. *Zeitschrift für arabische Linguistik* 16, 108–118.

Basulaiman, Eiman. 2018. *The Grammaticalisation of the Posture Verb* Sit *as a Progressive Marker in UHA: A Corpus-based Analysis*. Masters Thesis, University of Essex.

Boneh, Nora and Sichel, Ivy. 2010. Deconstructing possession. *Natural Language and Linguistic Theory* 28, 1–40.

Bresnan, Joan. 1982. The Passive in Lexical Theory. In Joan Bresnan (ed.), *The Mental Representation of Grammatical Relations*, pages 3–86, Cambridge: MIT Press.

Bresnan, Joan. 1989. The syntactic projection problem and the comparative syntax of locative inversion. In Chu-Ren Huang and Keh-Jiann Chen (eds.), *Proceedings of Rocling II Computational Linguistics Conference II*, pages 375–396, Nantou: The Association for Computational Linguistics and Chinese Language Processing (ACLCLP).

Bresnan, Joan. 1994. Locative Inversion and the Architecture of Universal Grammar. *Language* 70(1), 72–131.

Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Oxford: Blackwell Publishers.

Bresnan, Joan, Asudeh, Ash, Toivonen, Ida and Wechsler, Stephen. 2015. *Lexical-functional Syntax*. Oxford: Blackwell Publishers.

Bresnan, Joan and Kanerva, Jonni M. 1989. Locative Inversion in Chicheŵa: A Case Study of Factorization in Grammar. *Linguistic Inquiry* 20, 1–50.

Bresnan, Joan and Zaenen, Annie. 1990. Deep Unacccusativity in LFG. In Katarzyna Dziwirek, Patrick Farrell and Errapel Mejías-Bikandi (eds.), *Grammatical Relations: A Cross-Theoretical Perspective*, pages 45–58, Stanford: CSLI Publications.

Butt, Miriam, Dalrymple, Mary and Frank, Anette. 1997. An architecture for linking theory in LFG. In Miriam Butt and Tracy H. King (eds.), *Proceedings of LFG97*, Stanford: CSLI Publications.

Butt, Miriam and King, Tracy H. 1996. Structural topic and focus without movement. In Miriam Butt and Tracy H. King (eds.), *Proceedings of LFG96*, Stanford: CSLI Publications.

Butt, Miriam, King, Tracy Holloway, Niño, María-Eugenia and Segond, Frédérique. 1999. *A Grammar Writer's Cookbook*. Stanford: CSLI Publications.

Camilleri, Maris. 2019. The development of the universal perfect in Arabic. *Language* 95(4), 683–710.

Camilleri, Maris and Sadler, Louisa. 2017. Negative Sensitive Indefinites in Maltese. In Miriam

Butt and Tracy H. King (eds.), *Proceedings of LFG17*, pages 146–166, Stanford: CSLI Publications.

Camilleri, Maris and Sadler, Louisa. 2019a. The formation of non-canonical SUBJ-to-SUBJ lexical raising predicates in Arabic. In Miriam Butt, Tracy H. King and Ida Toivonen (eds.), *Proceedings of LFG19*, pages 90–110, Stanford: CSLI Publications.

Camilleri, Maris and Sadler, Louisa. 2019b. The grammaticalisation of a copula in vernacular Arabic. *Glossa* 4(1), 137.

Camilleri, Maris and Sadler, Louisa. 2020. The grammaticalisation of an auxiliary and a copula: The Arabic 'sit' participle. *Journal of Historical Syntax* 4(6), 1–60.

Choueiri, Lina. 2016. The Pronominal Copula in Arabic. *Brill's Journal of Afroasiatic Languages and Linguistics* 8(1), 101–135.

Comrie, Bernard. 1991. On the importance of Arabic for general linguistic theory. In Bernard Comrie and Mushira Eid (eds.), *Perspectives on Arabic Linguistics III*, pages 3–30, Amsterdam/Philadelphia: John Benjamins.

Dalrymple, Mary. 2001. *Lexical Functional Grammar*, volume 34 of *Syntax and Semantics*. New York: Academic Press.

Dalrymple, Mary, Dyvik, Helge and King, Tracy Holloway. 2004. Copula Complements: Closed or Open? In M. Butt and T. H. King (eds.), *Proceedings of LFG04*, pages 188–198, Stanford: CSLI Publications.

Eid, Mushira. 1991. Verbless sentences in Arabic and Hebrew. In Bernard Comrie and Mushira Eid (eds.), *Perspectives on Arabic Linguistics III*, pages 31–61, Amsterdam/Philadelphia: John Benjamins.

Eid, Mushira. 1993. Negation and predicate heads. In Mushira Eid and Greg Iverson (eds.), *Principles and Predication: The analysis of natural language*, pages 135–151, Amsterdam/Philadelphia: John Benjamins.

ElSadek, Shaimaa and Sadler, Louisa. 2015. Egyptian Arabic perceptual reports. In Miriam Butt and Tracy H. King (eds.), *Proceedings of LFG15*, pages 84–102, Stanford: CSLI Publications.

Falk, Yehuda. 2004. The Hebrew present tense copula as a mixed category. In Miriam Butt and Tracy H. King (eds.), *Proceedings of LFG04*, pages 226–246, Stanford: CSLI Publications.

Fassi-Fehri, Abdelkader. 2012. *Key Features and Parameters in Arabic Grammar*. Amsterdam: John Benjamins.

Fassi-Fehri, Abdulkader. 1993. *Issues in the structure of Arabic clauses and words*. Dordrecht: Kluwer Academic Publishers.

Findlay, Jamie Y. 2017. Mapping theory without argument structure. *Journal of Language Modelling* 4(2), 293–338.

Halila, Hafedh. 1992. *Subject specificity effects in Tunisian Arabic*. Ph. D.thesis, University of Southern California.

Hallman, Peter. 2020. Head and Dependent Marking in Clausal Possession. *Linguistic Inquiry* 4, 1–30.

Kibort, Anna. 2007. Extending the applicability of Lexical Mapping Theory. In Miriam Butt and Tracy H. King (eds.), *Proceedings of the LFG07*, pages 250–270, Stanford: CSLI Publications.

Laczkó, Tibor. 2012. On the (un)bearable lightness of being an LFG style copula in Hungarian.

In Miriam Butt and Tracy H. King (ed.), *Proceedings of the LFG12 Conference*, pages 341–361, Stanford: CSLI Publications.

Li, Charles N. and Thompson, Sandra A. 1977. A mechanism for the development of copula morphemes. In Charles N. Li (ed.), *Mechanisms of Syntactic Change*, pages 419–444, Austin: University of Texas Press.

Mohammad, Mohammad A. 1998. The syntax of indefinite subjects in equative sentences in Palestinian Arabic, ms. University of Florida.

Mohammad, Mohammad A. 2000. *Word order, agreement and pronominalization in Standard and Palestinian Arabic*. Amsterdam: John Benjamins.

Mohanan, K. P. 1982. Grammatical Relations and Clause Structure in Malayalam. In Joan Bresnan (ed.), *The Mental Representation of Grammatical Relations*, pages 503–589, Cambridge: MIT Press.

Nordlinger, Rachel and Sadler, Louisa. 2007. Verbless Clauses: Revealing the Structure within. In Jane Grimshaw, Joan Maling, Chris Manning, Joan Simpson and Annie Zaenen (eds.), *Architectures, Rules and Preferences: A Festschrift for Joan Bresnan*, pages 139–162, Stanford: CSLI Publications.

Ouhalla, Jamal. 1997. Remarks on focus in Standard Arabic. In Mushira Eid and Robert R. Ratcliffe (eds.), *Perspectives on Arabic Linguistics X*, pages 9–46, Amsterdam/Philadelphia: John Benjamins.

Ouhalla, Jamal. 1999. Focus and Arabic clefts. In G. Rebuschi and L. Tuller (eds.), *The Grammar of Focus*, pages 335–359, Amsterdam/Philadelphia: John Benjamins.

Ouhalla, Jamal. 2013. Agreement unified: Arabic. In Lisa L. Cheng and Norbert Corver (eds.), *Diagnosing Syntax*, pages 314–333, Oxford: OUP.

Przepiórkowski, Adam and Patejuk, Agnieszka. 2015. Two representations of negation in LFG: Evidence from Polish. In Miriam Butt and Tracy H. King (eds.), *Proceedings of LFG15*, pages 322–336, Stanford: CSLI Publications.

Pustet, Regina. 2003. *Copulas: Universals in the Categorization of the Lexicon*. Oxford: OUP.

Retsö, Jan. 1987. Copula and double object pronominal objects in some Semitic languages. *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 126, 219–45.

Rosén, Victoria. 1996. The LFG architecture and "verbless" syntactic constructions. In Miriam Butt and Tracy H. King (eds.), *Proceedings of LFG 1996*, pages 1–9, Stanford: CSLI Publications.

Soltan, Usama. 2007. *On Formal Feature Licensing in Minimalism: Aspects of Standard Arabic Morphosyntax*. Ph. D.thesis, University of Maryland.

Stassen, L. 1996. "The switcher's paradise: Nonverbal predication in Maltese". *Rivista di Linguistica* 8(1), 275–300.

Stassen, Leon. 1997. *Intransitive Predication*. Oxford: OUP.

Stassen, Leon. 2009. *Predicative Possession*. Oxford: OUP.

Sulger, Sebastian. 2015. *Modeling nominal predications in Hindi/Urdu*. Ph. D.thesis, Universität Konstanz.

Tapiéro, Norbert. 2002. *Manuel d'arabe algérien moderne: supplément de 15 dialogues avec traduction*. Paris: Klincksieck.

Copala Triqui's syntactic causative: Reconsidering clause linkage in LFG

Rebecca Dinkel
University at Albany, SUNY

Proceedings of the LFG'21 Conference

On-Line

Miriam Butt, Jamie Y. Findlay, Ida Toivonen (Editors)

2021

CSLI Publications

pages 72–92

http://csli-publications.stanford.edu/LFG/2021

Keywords: clause linkage types, mismatch approach, subordination, coordination, cosubordination

Dinkel, Rebecca. 2021. Copala Triqui's syntactic causative: Reconsidering clause linkage in LFG. In Butt, Miriam, Findlay, Jamie Y., & Toivonen, Ida (Eds.), *Proceedings of the LFG'21 Conference, On-Line*, 72–92. Stanford, CA: CSLI Publications.

1 Introduction

Copala Triqui (CT) is an Otomanguen language originally spoken in San Juan Copala, Oaxaca, Mexico.[1, 2] There is also a large diasporic community of CT speakers in the Capital Region of New York, where this study takes place. CT is of interest because it has several atypical clause linkage types that fall outside the scope of canonical subordination and coordination. This paper focuses on CT's syntactic causative which does not display all of the properties of canonical subordination nor canonical coordination due to its complement initial order, seen in example (1).

(1) **Qui-xra'** **xruj** qui-'yaj nana̲.
    CMPL-break pot CMPL-make wind
    'The wind made the pot break.'

[1] Acknowledgements: Thank you to Román Vidal López, Monica De Jesús Ramírez, Aaron Broadwell and the rest of the Albany Copala Triqui Working Group, and the DeCormier and Buhrmaster families and the University at Albany for grants supporting this research**.** Thank you to the LFG conference participants and reviewers who provided feedback on this work and the LFG conference organizers for facilitating these discussions.

[2] Examples are transcribed in the Triqui orthography developed by Barbara and Bruce Hollenbach of the Summer Institute of Linguistics for translation of the New Testament. This orthography is the same as IPA except for the following consonants: <x> = [ʃ], <xr> = [ʂ] (a retroflex alveopalatal sibilant), <ch> = [tʃ], <chr> = [tʂ], <c> = [k] (before front vowels), <qu> = [k] before back vowels, [v] = [β] and <j> = [h]. <Vn> transcribes a nasalized vowel, and an <h> is unpronounced but represents a syllable break wherein two vowels are adjacent to each other. Long vowels are indicated by <VV>. CT has eight tones that are divided into an upper (tones 5, 4, 3, 32, and 31) and lower register (tones 2, 1, and 13) with most verbs in CT having an upper and lower register stem. Verbal stems in continuative and completive aspect use their upper register stem, and in potential aspect their lower register stem. When negated, stems in completive and potential aspect flip to their lower and upper register, respectively (Broadwell 2019, 2014; Hollenbach 2005, 1984). High tones (tones 4 and 5) are indicated by accents and low level tones (tones 1 and 2) with an underscore while the mid tone (tone 3) is unmarked, for example: tone 5 <V́V́> <V́>; tone 4 <VV́>, <V́>; tone 3 <VV>, <V>; tone 2 <V̲V̲>, <V̲>; tone 1 <V̲V̲>, <V̲>; contour tone 3 2 <VV>, <V>; contour tone 3 1 <VV̲>, <V̲>; contour tone 1 3 <V̲V>, <V̲>. Though this transcription does not fully mark all tone distinctions, it is the easiest and most popular to use amongst Triqui speakers.

Abbreviations used in this paper are: 1, 2, 3=first, second, and third person; CMPL=completive; COMP= complementizer; CON=continuative; CONJ=conjunction; DEC=declaration; F=feminine; FAM=familiar; IP=inflectional phrase; M= masculine; N=noun; NEG=negative; NegP=negative phrase; NP=noun phrase; PART=particle; PL=plural; POT=potential; PP=prepositional phrase; PREP=preposition; PRO=pronoun; OBJ=object; S=singular; S=sentence (in syntactic tree); SUBJ=subject; V=verb.

In contrast, most verbs with clausal complements have a complement final order in CT, which this paper shows are canonically subordinate for CT. CT's syntactic causative also differs from canonical coordinate constructions in CT.

Work in LFG on atypical clause linkage types, like CT's syntactic causative, has just begun. For example, Belyaev (2014) argues that atypical clause linkage types are the result of systematic 'mismatches' between coordination and subordination at the c(onstituent)-structure, f(unctional)-structure, and s(emantic)-structure. Accounts of atypical clause linkage types outside of LFG include work in Role and Reference Grammar (RRG) on cosubordination (Van Valin & La Polla 1997). RRG defines cosubordination as a third kind of clause linkage where a non-embedded clause is grammatically dependent on another as demonstrated by operator scope and dependency.[3] This paper demonstrates that the 'mismatch' approach does not fully account for the properties of CT's syntactic causative. The c-structure and f-structure of CT's syntactic causative are not clearly diagnosable as either subordinate or coordinate but should be in a 'mismatch' account because CT's syntactic causative can be modeled in LFG, as seen in Figure (1).[4]

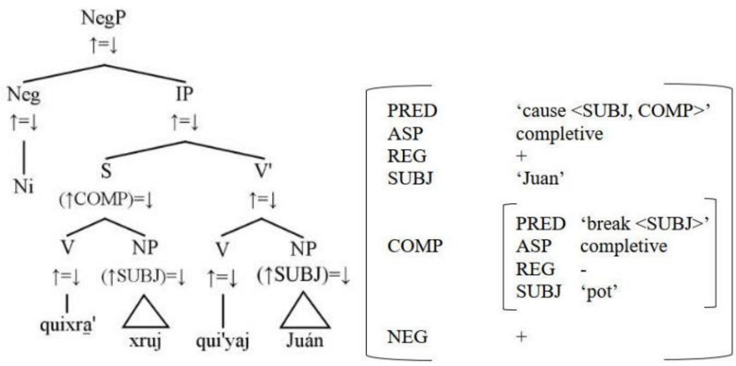


Figure 1: C-structure (left) & f-structure (right) of CT's syntactic causative

Instead, this paper demonstrates that CT's syntactic causative meets the definition of cosubordination. Figure (1) shows that CT's syntactic causative

[3] In RRG, operators are similar to 'functional categories' in other linguistic frameworks, and includes forms expressing negation, TAM, modality, illocutionary force, and directionals (Bohnemeyer & Van Valin 2017:150).

[4] In Figure (1) the feature 'REG' refers to 'tonal register' discussed in section (4), while its value of '+' refers to the verb being in its upper register stem and the value of '-' refers to the verb being in its lower register stem. As noted in footnote (2), verbal stems in completive and potential aspect flip to their lower and upper stems, respectively, when negated. In Figure (1), both verb stems are in completive aspect while only the verb stem *quixra̲'* is in its lower tonal register (Broadwell 2019, 2014; Hollenbach 2005, 1984).

consists of two unembedded clauses that exhibit operator dependency. However, there are problems with creating a third type of clause linkage, since doing so still may not capture all clause linkage variation cross-linguistically (Belyaev 2014:6). This paper aims to simply expand the description of clause linkage types in LFG by considering alternative approaches, such as RRG's concept of cosubordination. This paper thus uses definitions and diagnostics from both 'mismatch' and cosubordinate approaches as well as developing some language-internal diagnostics for subordination and coordination in CT, as explained throughout the remainder of this paper.

This paper is organized as follows. Section (2) provides an overview of relevant grammatical features of CT and section (3) of CT's syntactic causative. Sections (4-5) demonstrate that the presented constructions meet the definitions for canonical subordination and canonical coordination at different levels of grammar, respectively, and that canonical subordinate constructions exhibit operator dependency, but canonical coordinate constructions do not. Sections (4-5) do so while also developing language internal diagnostics for these properties in CT. Section (6) thus provides evidence of subordination and coordination at different levels of grammar and of operator dependency for CT's syntactic causative. Section (7) argues for the model of CT's syntactic causative in LFG presented in Figure (1) and a reconsideration of clause linkage types in LFG. Section (8) provides a conclusion.

2 Overview of Grammatical Features of CT

CT has a VSO order and is prepositional (Hollenbach 1992:187), as seen in example (2).

(2)	Qui-na'nu'	xnii	rihaan	mesá.
	CMPL-clean	boy	PREP	table.
	'The boy cleaned the table.'			

However, example (3) shows an SVO order is possible when the subject is focused (Hollenbach 1992:206).

(3)	**Juán**	qui-na'nu'	rihaan	mesá.
	Juan	CMPL-clean	PREP	table.
	'**Juan** cleaned the table.'			

Example (4) shows negative particles occur before the verb and declarative particles can optionally be used sentence finally (Hollenbach 1992:240-241).

(4)	**Ni**	güe̱j	Miguél	xráá	yahij	(**ma'**).
	NEG	CMPL.jump	Miguel	PREP	rock	DEC
	'Miguel didn't jump over the rock.'					

Most adverbs have relatively free distribution and can occur after the subject, object, or oblique, but not between the verb and the subject, as seen in examples (5 a-d).

(5) a.

Aga'	**'un'**	qui-na'nu'	Juán	rihaan	mesá.
o'clock	five	CMPL-clean	Juan	PREP	table.

‘Juan cleaned the table at five o’clock.’

b.

Qui-na'nu'	Juán	**aga'**	**'un'**	rihaan	mesá.
CMPL-clean	Juan	o'clock	five	PREP	table.

‘Juan cleaned the table at five o’clock.’

c.

Qui-na'nu'	Juán	rihaan	mesá	**aga'**	**'un'.**
CMPL-clean	Juan	PREP	table	o'clock	five.

‘Juan cleaned the table at five o’clock.’

d.

*Qui-na'nu'	**aga'**	**'un'**	Juán	rihaan	mesá.
CMPL-clean	o'clock	five	Juan	PREP	table.

‘Juan cleaned the table at five o’clock.’

3 Overview of CT’s Syntactic Causative

CT’s syntactic causative has similar properties to other syntactic causatives in the world’s languages. CT’s syntactic causative marks a CAUSE or ‘precipitating’ event which includes the CAUSER and an EFFECT or ‘result’ event, which includes the CAUSEE. CT’s syntactic causative is also formed through addition of the argument of the CAUSER to another clause (i.e. Comrie 1996; Dixon 2000). CT’s syntactic causative is formed with the verb *'yaj* ‘do, make, cause’ which can be used transitively with a normal VSO order, in its basic sense (Hollenbach 1992:204), as seen in example (6).

(6)

Qui-'yaj	Juán	ve'.
CMPL-make	Juan	house

‘Juan made the house.’

When *'yaj* ‘do, make, cause’ is used in its causative sense it has a complement, or EFFECT, clause initial order, though internally each clause follows a normal VSO order (Broadwell 2012), as seen in example (7).

	[[EFFECT	**EVENT]**	**[CAUSE**	**EVENT]]**
	EFFECT	**CAUSEE**	**CAUSE**	**CAUSER**
(7)	Qui-xra'	xruj	qui-'yaj	nana.
	CMPL-break	pot	CMPL-make	wind

‘The wind made the pot break.’

CT's syntactic causative may have once been part of a class of complement initial verbs, distinct from typical complement final verbs in CT. One verb, *rá* 'think', may no longer be fully productive, while *taj* 'say' shows differences from CT's syntactic causative currently. Constructions with the verb *taj* 'say' can displace the second clause, giving a complement final word order that maintains a VSO order clause internally, as seen in example (8 a-b).[5]

(8) a. **Se** **naca'** **so'** ca-taj so'
NEG POT.sweep 3S.M.PRO CMPL-say 3S.M.PRO
rihaan=j.
PREP=1S
'He told me he did not sweep.'

b. Ca-taj so' rihaan=j **se** **naca'**
CMPL-say 3S.M.PRO PREP=1S NEG POT.sweep
so'.
3S.M.PRO
'He told me he did not sweep.'

At an earlier stage, CT's syntactic causative allowed a complement initial order with the complementizer *se vaa* (Hollenbach 1992:220), but this is no longer acceptable for Triqui speakers with or without the complementizer, as seen in example (9 a-b).

(9) a. ***Qui-'yaj** **Juán** **se** **vaa** qui-xra'
CMPL-make Juan COMP CON.exist CMPL-break
xruj.
pot
'Juan made the pot break.'

b. ***Qui-'yaj** **nana** xra' xruj.
CMPL-make wind CON.break pot
'The wind made the pot break.'

4 Diagnosing Subordination at Different Levels of Grammar in CT

This section examines a class of verbs with a complement final order in CT, known as 'control verbs', and shows they are subordinate at different levels of grammar. Their features help create language internal diagnostics for canonical subordination for CT that can be compared to CT's syntactic causative, since they also do not take complementizers. The verb *taj* 'say' discussed in section

[5] The negative particle *se*, as opposed to *ni*, is used when the following verb is in potential aspect.

(3) is not a candidate for canonical subordination since its complement clause can be displaced and it does not exhibit properties of control. Features of control in CT include (1) copy control and (2) register control, given that CT does not have any true infinitival verbs (Broadwell 2019:17; Broadwell 2014:16). Copy control is when the controlled argument is expressed, as opposed to being omitted, as is true with languages with infinitival verbs (Broadwell 2019; Polinksy & Postdam 2006), as seen in example (10).[6,7]

(10)	Me	rá	**Juán**	cha	**Juán**	chraa.
	want	PART	Juan	POT.eat	Juan	tortilla

‘Juan wants to eat tortilla(s).’

Register control is where the control verb controls the tonal register of the verb of its complement (Broadwell 2019, 2014; Hollenbach 2005, 1984), though this topic is not discussed further due to space.

4.1 C-subordination

C-subordination is defined as when a constituent occupies the complement, adjunct, or specifier positions of a maximally projecting dominating node and is embedded (Belyaev 2014:42). Copy-control and the ungrammaticality of a complement initial order show control verbs are c-subordinate. Copy-control is ungrammatical when the controller occurs sentence initially in the focus position, as seen in example (11 a-b).

(11)	a.	**Juán**	me	rá	qui-na'nu'	rihaan	mesá.
		Juan	want	PART	POT-clean	PREP	table

‘Juan wants to clean the table.

b.	***Juán**	me	rá	qui-na'nu'	**Juán**	rihaan
	Juan	want	PART	POT-clean	Juan	PREP
	mesá.					
	table					

‘**Juan** wants to clean the table.’

Displacement of the second clause, giving a complement initial order, is also ungrammatical, regardless of whether an example exhibits copy-control, as seen in example (12 a-b).

---

6 The controlled copy can be a total repetition of the DP controller or a pronoun that agrees with the DP but, must be a pronoun if the controller is a pronoun. CT exhibits subject, object, and oblique control (Broadwel1 2019:31, 2014:20-21). This is not discussed further due to space.

7 The verb *me rá* ‘want’ does not change to show aspect and is glossed as such.

(12) a. *__Qui-na'n<u>u</u>'__ **Juán** **rihaan** **mesá** me rá
POT-clean Juan PREP table want PART
Juán.
Juan
'Juan wants to clean the table.'

b. *__Qui-na'n<u>u</u>'__ **Juán** **rihaan** **mesá** me rá.
POT-clean Juan PREP table want PART
'Juan wants to clean the table.'

Examples (11-12) demonstrate that control verbs syntactically dominate their complement clause, which must be embedded. This is because when the controller occurs outside of its normal position it affects the expression of control, disallowing the expression of its controlled copy. Further, the complement clause cannot occur outside of its subordinate position.

4.2 F-subordination

In f-subordination, a constituent of a construction fulfills a grammatical function of another constituent (Belyaev 2014:46). Control verbs are f-subordinate because they require a complement clause. For example, searches of a corpus developed by Broadwell and the Albany Copala Triqui Working Group (n.d.) show that for the control verb *me rá* 'want', the control clause never occurs on its own but, always with a complement clause.

4.3 S-subordination

Belyaev (2014:49-51) simply defines s-subordination as not exhibiting s-coordination, which is defined as any construction where two or more speech act discourse references are linked by a rhetorical relation. This is because it is not clear if s-subordination is a homogenous class and thus its formal definition (Belyaev 2014:49-51). At the least, s-subordination involves two clauses in the same speech act, in which a predicate links their propositional content (Belyaev 2014:49-51). An s-subordinate construction can be diagnosed by scoping negation or modal adverbs and the ability to be focused (Belyaev 2014:49-51). Further, one clause is also always presupposed with s-subordination, whereas this is not the case with s-coordination (Belyaev 2014:49-51).

Given that s-subordination might not be a homogenous class, this paper also uses Bohnemeyer and Van Valin's (2017) Macro Event Property (MEP) for diagnosing s-subordination. The MEP is present when complex events are described as referencing one event despite containing possible subevents (Bohnemeyer & Van Valin's 2017:147). Explicit diagnostics for if the MEP is present in a given construction are the use of a single time-positional adverb or with noncontradictory time-positional adverbs of a more specific meaning, since a single event cannot occur at two different times or places. Some cases of s-subordination based on Belyaev's (2014) diagnostics may also contain the MEP, but other cases may take more than one time-positional adverbial of contradictory meanings. If a construction has the MEP, it may be said that it is definitively s-subordinate, whereas the reverse may not be true.

Control verbs can take two temporal adverbs of noncontradictory meaning, as seen in example (13).[8]

(13)	Me	rá	Juán	**quii**	qui-na'nu'	Juán
	want	PART	Juan	yesterday	POT-clean	Juan
		taxrej		rihaan	mesá.	
		early.morning		PREP	table	
	'Juan wanted to clean the table yesterday in the early morning.'					

However, control verbs cannot take two temporal adverbs when they have contradictory meanings, as seen in example (14).

(14)	*Me	rá	Juán	**aga'**	**vij**	qui-na'nu'
	want	PART	Juan	o'clock	two	POT-clean
		Juán	**aga'**	**'un'**	rihaan	mesá.
		Juan	o'clock	five	PREP	table
	'At two o'clock, Juan wanted to clean the table at five o'clock.'					

Control verbs thus reference one event that takes place at a distinct place and time, despite containing a subevent, have the MEP, and are s-subordinate.

Control verbs also exhibit operator scope and dependency, which can occur with subordinate structures in RRG and is diagnostic of s-subordination in LFG. Independent negation of each clause, and consequently two instances of negation, are disallowed, demonstrating that control verbs have scoping negation, as seen in example (15 a-c).

(15)	a.	**Ni**	me	rá	Juán	qui-na'nu'	Juán	rihaan
		NEG	want	PART	Juan	POT-clean	Juan	PREP
			mesá.					
			table					
		'Juan doesn't want to clean the table.'						

[8] This is true regardless of the position of the adverbs in each clause for examples (14-15).

b. *Me rá Juán **se** qui-na'nu' Juán rihaan
want PART Juan NEG POT-clean Juan PREP
mesá.
table
‘Juan wants to not clean the table.’

c. ***Ni** me rá Juán **se** qui-na'nu' Juán
NEG want PART Juan NEG POT-clean Juan
rihaan mesá.
PREP table
‘Juan doesn’t want to not clean the table.’

Differences in the grammaticality of different declarative particles affirm that negation scopes over both clauses, as seen in example (16 a-b). That these particles are diagnostic of the scope of negation is developed in section (5).

(16) a. **Ni** me rá Juán qui-na'nu' Juán rihaan
NEG want PART Juan POT-clean Juan PREP
mesá **ma'**.
table DEC
‘Juan doesn’t want to clean the table.

b. ***Ni** me rá Juán qui-na'nu' Juán rihaan
NEG want PART Juan POT-clean Juan PREP
mesá **a**.
table DEC
‘Juan doesn’t want to clean the table.’

5 Diagnosing Coordination at Different Levels of Grammar in CT

This section examines canonical coordinate constructions that can take the conjunction *ne* ‘and’ in CT, and shows they are coordinate at different levels of grammar. Their features help create language internal diagnostics for canonical coordination that can be compared to CT’s syntactic causative. Canonical coordinate constructions that take the conjunction *ne* ‘and’ are relevant because at an earlier stage these constructions could be covertly coordinated and omit the conjunction *ne* ‘and’ in some cases, which might have also been the case with CT’s syntactic causative. An example is seen in (17).

(17) Chá Juán (**ne**) co-'o so' a.
CMPL.eat Juan CONJ CMPL-drink 3S.M.PRO DEC
‘Juan ate and he drank’.

5.1 C-coordination

C-coordination is defined as when a construction's sister nodes and their immediately dominating node are of the same phrasal category and thus unembedded (Belyaev 2014:41; Yuasa & Sadock 2002; Haspelmath 2004).[9] Free placement of the second clause shows canonical coordinate constructions that take the conjunction *ne̲* 'and' are c-coordinate. Either clause can occur sentence initially or sentence finally as seen in example (18 a-b).

(18) a. **Chá** **so'** ne̲ co-'o so' a.
CMPL.eat 3S.M.PRO CONJ CMPL-drink 3S.M.PRO DEC
'He ate and he drank'.

b. Co-'o so' ne̲ **chá** **so'** a.
CMPL-drink 3S.M.PRO CONJ CMPL.eat 3S.M.PRO DEC
'He drank and he ate'.

Example (18 a-b) thus demonstrates that one clause is not dominated, or subordinate, to another.

5.2 F-coordination

Constituents that are f-coordinate are defined as being members of a set and do not fulfill any necessary grammatical function of another constituent (Belyaev 2014:46). These constituents can stand on their own without the other, as seen in example (19 a-b), in contrast to together in example (20).

(19) a. Qui-ra'ánj Miguél.
CMPL-dance Miguel
'Miguel danced.'

[9] Przepiórkowski and Patejuk (2021) propose analyzing coordinate structures without reference to syntactic categories, in response to previous analyses of unlike category coordination. Since this paper uses syntactic/phrasal categories in its analysis, it adopts this specific definition of coordination, given that coordinate structures can broadly be defined as structures that combine units of the same 'type' (Haspelmath 2004:34). This paper shows that the units of CT's syntactic causative are not truly of the same 'type' in addition to occupying different syntactic/phrasal categories, though it acknowledges that unlike category coordination of different syntactic/phrasal categories is a genuine phenomenon.

	b.	C-achráá	Juán	ya'ánj.
		CMPL-sing	Juan	instrument

‘Juan played the instrument.’

The Coordinate Structure Constraint (CSC) may diagnose f-coordination and stipulates that elements of a conjunct cannot be extracted (Belyaev 2014: 46-47; Ross 1967). For canonical coordinate constructions that take the conjunction *ne̱* ‘and’ only arguments of the first conjunct can be focused, as seen in example (20 a-d).

(20)	a.	C-achráá	Juán	ya'ánj	ne̱	qui-ra'ánj	Miguél
		CMPL-sing	Juan	instrument	CONJ	CMPL-dance	Miguel

‘Juan played the instrument and Miguel danced.’

b.	**Juán**	c-achráá	ya'ánj	ne̱	qui-ra'ánj	Miguél
	Juan	CMPL-sing	instrument	CONJ	CMPL-dance	Miguel

‘**Juan** played the instrument and Miguel danced.’

c.	**Ya'ánj**	c-achráá	Juán	ne̱	qui-ra'ánj	Miguél
	instrument	CMPL-sing	Juan	CONJ	CMPL-dance	Miguel

‘Juan played **the instrument** and Miguel danced.’

d.	***Miguél**	c-achráá	Juán	ya'ánj	ne̱	qui-ra'ánj
	Miguel	CMPL-sing	Juan	instrument	CONJ	CMPL-dance

‘Juan played the instrument and **Miguel** danced.’

5.3 S-coordination

A construction may not be s-subordinate if it does not exhibit the MEP and each clause can take a time-positional adverb of contradictory meaning to the other. This is the case for canonical coordinate constructions that take the conjunction *ne̱* ‘and’, as seen in example (21).[10]

(21)	Qui-ra'a̱nj	Miguél	**a'yuj**	ne̱	c-achráá	Juán
	POT-dance	Miguel	tomorrow	CONJ	CMPL-sing	Juan
	ya'ánj	**quii**.				
	instrument	yesterday				

‘Miguel will dance tomorrow and Juan played the instrument yesterday.’

Canonical coordinate constructions with *ne̱* ‘and’ thus reference more than one event that can occur at different places and times and do not have the MEP.

[10] This is true regardless of adverb placement in each clause for example (21).

Canonical coordinate constructions with *ne* 'and' can be affirmed to be s-coordinate because they also do not exhibit operator scope or dependency. Each conjunct can be independently negated, and consequently, two instances of negation are allowed. Different patterns of negation allow different declarative particles to be used, as seen in example (22 a-f).

(22) a. **Ni** c-achraa Miguél ne qui-ra'ánj
NEG CMPL-sing Miguel CONJ CMPL-dance
Juán **a**.
Juan DEC
'Miguel didn't sing and Juan danced.'

b. ***Ni** c-achraa Miguél ne qui-ra'ánj
NEG CMPL-sing Miguel CONJ CMPL-dance
Juán **ma'**.
Juan DEC
'Miguel didn't sing and Juan danced.'

c. ?C-achráá Miguél ne **ni** qui-ra'anj
CMPL-sing Miguel CONJ NEG CMPL-dance
Juán **a**.
Juan DEC
'Miguel sang and Juan didn't dance.'

d. C-achráá Miguél ne **ni** qui-ra'anj
CMPL-sing Miguel CONJ NEG CMPL-dance
Juán **ma'**.
Juan DEC
'Miguel sang and Juan didn't dance.'

e. ?**Ni** c-achraa Miguél ne **ni** qui-ra'anj
NEG CMPL-sing Miguel CONJ NEG CMPL-dance
Juán **a**.
Juan DEC
'Miguel didn't sing and Juan didn't dance.'

f. **Ni** c-achraa Miguél ne **ni** qui-ra'anj
NEG CMPL-sing Miguel CONJ NEG CMPL-dance
Juán **ma'**.
Juan DEC
'Miguel didn't sing and Juan didn't dance.'

The declarative particle *ma'* cannot be used when only the first conjunct is negated, as seen in example (22 b) but can be used in all other examples where the second conjunct is negated, as seen in examples (22 d & f). Thus, the declarative particle *ma'* is diagnostic of scoping negation.

6 Coordination and Subordination at Different Levels of Grammar for CT's Syntactic Causative

This section examines whether CT's syntactic causative is subordinate or coordinate at different levels of grammar by comparing its properties to control verbs and canonical coordinate constructions with *ne̲* 'and'. This section demonstrates that while CT's syntactic causative is clearly diagnosable as subordinate at its s-structure, it does not display all of the properties of either subordination or coordination at both its c-structure and f-structure. This is contrary to the 'mismatch' account where different levels of grammar must be diagnosable as either subordinate or coordinate for a given construction.

6.1 C-structure

CT's syntactic causative does not exhibit the properties of control seen with control verbs, even when the arguments of the CAUSE clause are coreferential with the arguments of the EFFECT clause. Copy control is disallowed and a reflexive particle must be used, as seen in example (23 a-b).[11]

(23) a.

*Qui-na'nu'	**Juán**	rihaan	mesá	qui-'yaj	**Juán**.
CMPL-clean	Juan	PREP	table	CMPL-make	Juan

'Juan made (himself) clean the table.'

b.

Qui-na'nu'	**ma'an**	**Juán**	rihaan	mesá	qui-'yaj	Juán.
CMPL-clean	self.of	Juan	PREP	table	CMPL-make	Juan

'Juan made himself clean the table.'

Like control verbs and unlike canonical coordinate constructions with *ne̲* 'and', CT's syntactic causative has restrictions on the displacement of its second clause. The CAUSE clause can only occur sentence initially when the CAUSER is focused, as seen in example (24 a-b).

[11] Note that Hollenbach (1984) demonstrates that reflexives in Copala Triqui violate a number of Chomsky's (1981) binding principles and nothing else is implied about the c-structure of CT's syntactic causative from examples (23 a-b).

(24) a.

***Qui-'yaj**	**Juán**	qui-na'nu'	xnii	rihaan	mesá.
CMPL-make	Juan	CMPL-clean	child	PREP	table

‘Juan made the boy clean the table.’

b.

Juán	**qui-'yaj**	qui-na'nu'	xnii	rihaan	mesá.
Juan	CMPL-make	CMPL-clean	child	PREP	table

‘**Juan** made the boy clean the table.’

Restrictions on displacement seen in example (24 a-b) also demonstrate that CT’s syntactic causative does not have an OVS structure where the EFFECT clause and the CAUSE verb *'yaj* ‘do, make, cause’ form a constituent. CT’s syntactic causative is not subordinate in this sense. Instead, the CAUSE clause and EFFECT clause are unembedded sisters to each other.

Unlike both control verbs and canonical coordinate constructions with *ne* ‘and’ adverbs cannot occur in both clauses of CT’s syntactic causative. Adverbs are disallowed in the CAUSE clause, as seen in example (25 a-c).

(25) a.

A'yuj	qui-na'nu'	xnii	rihaan	mesá	qui-'yaj	Juán.
tomorrow	POT-clean	boy	PREP	table	POT-make	Juan

‘Juan will make the boy clean the table tomorrow.’

b.

Qui-na'nu'	xnii	**a'yuj**	rihaan	mesá	qui-'yaj	Juán.
POT-clean	boy	tomorrow	PREP	table	POT-make	Juan

‘Juan will make the boy clean the table tomorrow.’

c.

*Qui-na'nu'	xnii	rihaan	mesá	qui-'yaj	Juán	**a'yuj**.
POT-clean	boy	PREP	table	POT-make	Juan	tomorrow

‘Juan will make the boy clean the table tomorrow.’

Further, unlike canonical coordinate constructions with *ne* ‘and’, CT’s syntactic causative cannot take an overt coordinator, and thus cannot be interpreted as being covertly coordinate, as seen in example (26).

(26)

*Qui-xra'	xruj	**ne**	qui-'yaj	ra'a	chruun
CMPL-break	pot	CONJ	CMPL-break	branch	tree

‘The tree branch did it, and the pot broke.’

Example (9a) above also shows CT’s syntactic causative cannot take an overt complementizer that occurs with some complement taking verbs in CT.

6.2 F-structure

Like control verbs, CT’s syntactic causative requires a complement clause, the EFFECT clause. For example, searches of a corpus developed by Broadwell and the Albany Copala Triqui Working Group (n.d.) show that the CAUSE clause never occurs on its own. Thus, the EFFECT clause fulfills the grammatical function of being an argument of the CAUSE clause. However, like canonical coordinate clauses with *ne* ‘and’, the CSC applies to CT’s syntactic causative. The CAUSER cannot be focused without the CAUSE verb also occurring sentence initially, as seen in example (27).

(27) *Juán qui-na'nu' xnii rihaan mesá **qui-'yaj**.
Juan CMPL-clean child PREP table CMPL-make
‘**Juan** made the boy clean the table.’

6.3 S-structure

CT’s syntactic causative is clearly s-subordinate. CT’s syntactic causative can take two temporal adverbs of noncontradictory meaning when the adverbs occur in the EFFECT clause, as seen in example (28).[12]

(28) **A'yuj** **taxrej** qui-na'nu' xnii rihaan mesá
tomorrow early.morning POT-clean boy PREP table
qui-'yaj Juán.
POT-make Juan
‘Juan will make the boy clean the table tomorrow in the early morning.’

However, CT’s syntactic causative cannot take two temporal adverbs of contradictory meaning when in the EFFECT clause, as seen in example (29).

(29) *Aga' **vij** qui-na'nu' xnii rihaan mesá **aga'**
o’clock two CMPL-clean boy PREP table o’clock
'un' qui-'yaj Juán.
five CMPL-make Juan
‘At two, Juan made the boy clean the table at five.’

[12] This is true regardless of the placement of the adverbs in each clause for examples (28-29).

CT’s syntactic causative thus references one event that takes place at a distinct place and time, has the MEP, and is s-subordinate.

CT’s syntactic causative also exhibits operator scope and dependency, like control verbs and cosubordinate structures in RRG, and s-subordinate structures in LFG. When negated, it is implied that the EFFECT event still occurred, even if the specified CAUSER was not the agent of the action, making the EFFECT clause presupposed. Independent negation of each clause, and thus two instances of negation, are disallowed, as seen in example (30 a-c).

(30) a. **Ni** c-acaa ve' qui-'yaj Juán.
NEG CMPL-burn house CMPL-make Juan
‘Juan didn’t make the house burn.’

b. *C-acaa ve' **ni** qui-'yaj Juán.
CMPL-burn house NEG CMPL-make Juan
‘Juan didn’t make the house burn.’

c. ***Ni** c-otoj nij xnii **ni** qui-'yaj Juán
NEG CMPL-sleep PL boy NEG CMPL-make Juan.
‘Juan didn’t make the boys not sleep.’

The negative particle in example (30 a) is scoping given that only the use of the declarative particle *ma'* is grammatical for CT’s syntactic causative when it is negated, as seen in example (31 a-b).

(31) a. ***Ni** qui-na'nu xnii rihaan mesá qui-'yaj
NEG CMPL-clean child PREP table CMPL-make
Juán **a**.
Juan DEC
‘Juan didn’t make the boy clean the table’

b. **Ni** qui-na'nu' xnii rihaan mesá qui-'yaj
NEG CMPL-clean child PREP table CMPL-make
Juán **ma'**.
Juan DEC
‘Juan didn’t make the boy clean the table’

7 Modeling CT’s Syntactic Causative: Reconsidering Clause Linkage in LFG

This section provides a summary of the previous sections and an argument for the model of CT’s syntactic causative in LFG presented in Figure (1). Table (1) summarizes definitions from both ‘mismatch’ and cosubordinate approaches that the previous sections use to diagnose subordination and coordination at different levels of grammar.

	Subordination	**Coordination**
C-structure	Where a constituent occupies the complement, adjunct, or specifier positions of a maximally projecting dominating node (unembedded)	Where constituents are sister nodes of the same category and of the same category of immediately dominating node (unembedded)
F-structure	Where a constituent fulfills a grammatical function of another	Where constituents are members of a set
S-structure	Where a construction contains one speech act that links propositional contents via a predicate, and may also only reference one event	Where a construction contains two speech acts linked by a rhetorical relation and references more than one event

Table 1: Definitions of subordination & coordination at different levels of grammar

Table (2) summarizes the properties of canonical subordinate and canonical coordinate constructions in CT that the previous sections compare to CT's syntactic causative. The previous sections also use these properties to diagnose subordination and coordination at different levels of grammar.

	Canonical Subordinate Clauses (Control Verbs)	**Canonical Coordinate Clauses**	**CT's Syntactic Causative**
Copy Control	yes	no	no
Register Control	yes	no	no
Scoping Negation	yes	no	yes
Overt Coordinator	no	yes	no
Displacement of the second clause	no	yes	no
Temporal adverbs of contradictory meaning	no	yes	no

Table 2: Properties of subordination, coordination, & CT's syntactic causative

Sections (4-5) show that canonical subordinate and canonical coordinate structures in CT are subordinate and coordinate at different levels of grammar, respectively. In contrast, section (6) demonstrates that CT's syntactic causative displays mixed properties of subordination and coordination at both its c-structure and f-structure, despite being diagnosable as subordinate at its s-structure. This is contrary to the 'mismatch' account where different levels of grammar must be clearly diagnosable as either subordinate or coordinate.

A summary of these mixed properties is as follows: CT's syntactic causative is like true f-subordinate constructions in CT with one of its clauses being an argument of another. However, CT's syntactic causative is also like true f-coordinate structures in CT by exhibiting the CSC. CT's syntactic causative is like true c-subordinate constructions in CT because it disallows displacement of its second clause, the CAUSE clause, without focusing the CAUSER. Unlike true c-subordinate structures in CT, CT's syntactic causative disallows copy control, a property that can show one clause dominates another in CT. CT's syntactic causative also does not have a subordinate OVS structure where the CAUSE verb *'yaj* 'do, make, cause' dominates the EFFECT clause. Instead, the CAUSE clause and the EFFECT clause are distinct constituents, or sisters, and not embedded. CT's syntactic causative is also unlike both canonical subordinate and coordinate constructions in CT because adverbs cannot occur in both of its clauses, but only in the EFFECT clause. Finally, CT's syntactic causative is also unlike canonical coordinate constructions in CT by not being able to take an overt coordinator.

Given these properties of CT's syntactic causative, the EFFECT clause should be a non-projecting exocentric phrasal category S that can stand on its own, and not an IP that dominates the CAUSE clause. This is in contrast to control verbs, which this paper argues have an IP that dominates a complement clause of the category of S, similar to Broadwell's (2014) analysis of control verbs. The ungrammatically of adverbs in the CAUSE clause suggests it is of a different phrasal category than IP or S. This paper labels the CAUSE clause as a V' after Broadwell's (2014) analysis who argues there are no true VP's in CT. Thus, the EFFECT clause and CAUSE clause are not of the same phrasal category. CT's syntactic causative also does not exhibit other properties of true c-coordinate constructions in CT, so it cannot be diagnosed as being truly c-coordinate. CT's syntactic causative cannot be said to be meet the definition of c-subordination either, since V' is not a maximally projecting node.

Is CT's syntactic causative cosubordinate? Its c-structure and f-structure cannot be clearly diagnosed as either subordinate or coordinate but should be in the 'mismatch' account of atypical clause linkage types. CT's syntactic causative does meet the definition of cosubordination where a clause is non-embedded, yet grammatically dependent, and exhibits operator scope and dependency. However, asserting that CT's syntactic causative is cosubordinate may expand clause linkage typology when there is no agreed upon cross-linguistic syntactic criteria to justify this (Belyaev 2014:6; Bickel 2010). More phenomena of clause linkage types from different languages need to be modeled in LFG to see if cosubordination should be considered a genuine third type of clause linkage. This paper contributed to this aim by examining a construction with an atypical clause linkage type that does not take an overt coordinator or subordinator, as has been done in LFG previously (Belyaev 2014). At the least, this paper shows that clause linkage types that meet the definition of cosubordination can be successfully modeled in LFG.

8 Conclusion

CT's syntactic causative displays mixed properties of canonical subordinate and canonical coordinate constructions in CT. A 'mismatch' approach cannot account for all of the features of CT's syntactic causative because there is not a clearly diagnosable mismatch between subordination and coordination at different levels of grammar. CT's syntactic causative was clearly diagnosable as s-subordinate, but not clearly diagnosable as subordinate or coordinate at its c-structure and f-structure. However, CT's syntactic causative does meet the definition of cosubordination where an unembedded clause is grammatically dependent on another, as diagnosed through operator scope and dependency. More research on other languages is needed to determine if this is a genuine third kind of clause linkage. At the least, this paper expands the range of atypical clause linkage types that can be modeled in LFG.

References

Belyaev, Oleg. 2014. Systematic Mismatches: Coordination and Subordination at Three Levels of Grammar. *Journal of Linguistics* 51(2), 267–326.

Bickel, Balthasar. 2010. Capturing Particulars and Universals in Clause Linkage: A Multivariate Analysis. In Isabelle Bril (ed.), *Clause Linking and Clause Hierarchy: Syntax and Pragmatics*, pages 27–50, Philadelphia: John Benjamins Publishing Company.

Bohnemeyer, Jürgen and Van Valin Jr., Robert. 2017. The Macro-Event Property and the Layered Structure of the Clause. *Studies in Language* 41(1), 142–197.

Broadwell, George Aaron. 2019. Documenting Control and Pseudo-Control in Copala Triqui. Unpublished Manuscript. https://www.academia.edu/40163333/DOCUMENTING_CONTROL_AND_PSEUDO_CONTROL_IN_COPALA_TRIQUI.

Broadwell, George Aaron. 2014. Syntax from the Bottom Up: Elicitation, Corpus Data, and Thick Descriptions. In Mariam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG'14 Conference*, pages 131–157, Stanford: CSLI Publications.

Broadwell, George Aaron. 2012. Causative constructions in Copala Triqui. In *Working linguist blog*. http://workinglinguist.blogspot.com/search?q=causative.

Comrie, Bernard. 1996. The Syntax of Causative Constructions: Cross-Language Similarities and Divergencies. In Masayoshi Shibatani (ed.), *The Grammar of Causative Constructions*, *Syntax and Semantics* vol. 6, pages 261–312, New York: Academic Press.

Culicover, Peter W. and Jackendoff, Ray. 1997. Semantic Subordination Despite Syntactic Coordination. *Linguistic Inquiry* 28(2), 195–217.

Dixon, Robert M. W. 2000. A Typology of Causatives: Form, Syntax and Meaning. In Robert M. W. Dixon and Alexandra Y. Aikhenvald (eds.), *Changing Valency: Case Studies in Transitivity*, 1st ed., pages 30–83, Cambridge: Cambridge University Press.

Erickson de Hollenbach, Elena. 2005. *Gramática Popular del Trique de Copala*, 1st ed., *Serie de Gramáticas de Lenguas Indígenas de México* 11, Tlalpan Distrito Federal de México: Instituto Lingüístico de Verano.

Haspelmath, Martin. 2004. Coordinating Constructions: An Overview. In Martin Haspelmath (ed.), *Typological Studies in Language* vol. 58, pages 3–39, Amsterdam: John Benjamins Publishing Company.

Hollenbach, Barbara E. 1992. A Syntactic Sketch of Copala Trique. In C. Henry Bradley and Barbara Hollenbach (eds.), *Studies in the Syntax of Mixtecan Languages* vol. 4, *Publications in Linguistics Series* 111, pages 173–431, Dallas: Summer Institute of Linguistics & the University of Texas at Arlington.

Hollenbach, Barbara E. 1984a. Copala Trique Tone and Universal Tone Features. In Stuart Davis (ed.), *Studies on Native American Languages, Japanese and Spanish*, *Coyote Papers: Working Papers in Linguistics A-Z* vol. 5, pages 96–119, Tucson: University of Arizona Linguistics Circle.

Hollenbach, Barbara E. 1984b. Reflexives and Reciprocals in Copala Trique. *International Journal of American Linguistics* 50(3), 272–291.

Olson, Michael Leon. 1981. *Barai Clause Junctures: Towards a Functional Theory of Interclause Relations*. Ph.D. thesis, Australian National University.

Polinsky, Maria and Potsdam, Eric. 2006. Expanding the Scope of Control and Raising. *Syntax* 9(2), 171–192.

Przepiórkowski, Adam and Patejuk, Agnieszka. 2021. Coordinate Structures without Syntactic Categories. In Mariam Butt, Ida Toivonen, and Jamie Findlay (eds.), *Proceedings of the LFG'21 Conference*, this volume, Stanford: CSLI Publications.

Ross, John Robert. 1967. *Constraints on Variables in syntax*. Ph.D. thesis, Massachusetts Institute of Technology.

Van Valin Jr., Robert D. and LaPolla, Randy J. 1997. *Syntax: Structure, Meaning and Function*. Cambridge: Cambridge University Press.

Yuasa, Etsuyo and Sadock, Jerry M. 2002. Pseudo-Subordination: A Mismatch between Syntax and Semantics. *Journal of Linguistics* 38, 87–111.

Obligatory and arbitrary anaphoric control in adjuncts

James Donaldson
The University of Edinburgh

Proceedings of the LFG'21 Conference

On-Line

Miriam Butt, Jamie Y. Findlay, Ida Toivonen (Editors)

2021

CSLI Publications

pages 93–103

http://csli-publications.stanford.edu/LFG/2021

Keywords: adjunct control, dangling modifier, small clause, gerund, participle, adverbial, converb, free adjunct

Donaldson, James. 2021. Obligatory and arbitrary anaphoric control in adjuncts. In Butt, Miriam, Findlay, Jamie Y., & Toivonen, Ida (Eds.), *Proceedings of the LFG'21 Conference, On-Line*, 93–103. Stanford, CA: CSLI Publications.

Abstract

In this paper, I present an approach to the control of some nonfinite and verbless adjuncts in English. This involves a modification of the approach in my dissertation (Donaldson 2021), which used functional and arbitrary anaphoric control. Here, I propose that these adjuncts are instead controlled anaphorically in all cases, with both obligatory and arbitrary control available in principle.

1 Introduction

Many English participial adjunct clauses seem to be missing a subject. These are generally controlled by the subject of the matrix clause to which they are attached (1a), but not always (1b).

(1) a. Watching him, Thrasher realized that something in his appearance didn't ring true. (Green 1956: *The Last Angry Man*)

b. Watching him, it seemed as if a fibre, very thin but pure, of the enormous energy of the world had been thrust into his frail and diminutive body. (Woolf 1942: *The Death of the Moth*)

In Donaldson (2021), I argued that this was the result of a dual control pattern in which functional and anaphoric control readings were both in principle available. Other studies with similar conclusions include Green (2018) and Landau (2021).

I now claim that this duality instead involves obligatory and arbitrary anaphoric control. Functional control between the adjunct and matrix clauses is not involved at all. This new approach can be more consistently applied across the variety of adjuncts that display these control patterns. It also has the advantage of using the same f-structure for both readings: while these two types of control are resolved differently, they do not involve syntactic ambiguity.

2 Two types of control

It is clear that at least some adjunct control must be arbitrary anaphoric control. Otherwise, extrasentential controllers in sentences like (1b) could not be explained (Bresnan 1982: 396f., Butt et al. 1999: 39f.).

[†]I would like to thank the participants at LFG21 for a warm welcome and several generous discussions. In Donaldson (2021: 187), I expressed the hope that functional control could eventually be dispensed with altogether. But I would not have returned straight away to pursue a purely anaphoric approach had Péter Szűcs not also brought up this possibility in a question. I would also like to thank Mary Dalrymple, Geoff Pullum, and the anonymous reviewers for their helpful comments related to this paper. The remaining inadequacies are completely mine.

In the generative literature, these sentences are frequently analysed as involving logophoric control (Williams 1992; Landau 2017; Green 2019). Landau in particular argues that a null projection of the matrix clause provides a human experiencer to serve as controller.[1] But while human experiencers do indeed frequently control adjuncts that are not controlled by the matrix subject, extrasentential anaphoric controllers that are inanimate are common enough to argue against this approach (Donaldson 2021: 123-139):

(2) Being made of stainless steel, rust won't be an issue. (after Davies 2018)

The item that is made of stainless steel does not appear in the matrix clause, but whatever it is cannot be sentient and so cannot be classified as an experiencer.

In any case, arbitrary anaphoric control is not enough to explain all of the control patterns that we see. When these adjuncts are controlled by the matrix subject, they are more strictly associated with it (3a) than explicit pronouns would be (3b).

(3) a. While preparing himself/*herself, Harry phoned Sally.
    b. While she prepared herself, Harry phoned Sally.

Traditionally, this sort of strict association with the matrix subject has been used to argue in favour of functional control for adjuncts (Mohanan 1983). Aside from the two exceptions mentioned earlier (Bresnan 1982, Butt et al. 1999), most LFG analyses take a functional approach (Dalrymple 2001: 149, Kroeger 2004: 112, Bresnan et al. 2016 [2001]: 99, Dalrymple, Lowe & Mycock 2019: 589ff., Börjars, Nordlinger & Sadler 2019: 123ff., *inter alia*). As we have seen, this cannot be the whole story because extrasentential control cannot be functional.

But is functional control even part of the story? That is, could what appears to be functional control actually be obligatory anaphoric control? We can find support for this approach in the fact that we can rule out functional control in other adjuncts that appear to have the same control patterns. I will turn to these other adjuncts next.

3 Gerunds and participles

As we have seen, some participial adjuncts are introduced by prepositions like *while* (4a,b), *when*, *once*, and *if*. There are other adjuncts that superficially seem to belong to this category, but I will argue that they are actually

[1]See Landau (2021: 122-135) for indications of a shift towards a wider vision of non-obligatory control that marginally includes topicality.

gerundive adjuncts instead. They are introduced by prepositions like *after* (5a,b), *before*, *despite*, and *without*. Both groups involve the same control patterns: control by the subject of the matrix clause is strongly preferred (4a, 5a) (a preference that goes beyond what we see with explicit pronouns (3)) and extrasentential control is possible (4b, 5b).

(4) a. While enjoying himself/*herself at the park, Harry phoned Sally.
    b. While eating lunch by myself in the park, a seagull landed nearby.

(5) a. After composing himself/*herself, Harry phoned Sally.
    b. After eating lunch by myself in the park, the weather took a turn for the worse.

At first glance, these groups seem similar enough for there to be no reason to divide them. But as Stump (1981: 10f.) points out, distinctions emerge when we consider the environments created by these prepositions. *While* can make a variety of phrases predicative, such as NPs (6a), AdjPs (6b), and PPs (6c). In contrast, *after* might be able to select an *-ing* complement, but it cannot make an NP (7a), AdjP (7b), or PP (7c) predicative.

(6) a. While a teacher, he enjoyed talking to students.
    b. While still young, he started to worry about several things.
    c. While in jail, he repented.

(7) a. *After a teacher, he enjoyed his retirement.
    b. *After young, he started to worry about several things.
    c. *After in jail, he repented.

So it seems that we have to account for how prepositions like *while* create environments that are inherently predicational for the complement, whether that complement is headed by a participle, noun, adjective, or preposition. We will also have to make sure that our account can show why prepositions like *after* do not create inherently predicational environments.

One possible explanation is that an apparently identical *-ing* complement is participial with *while* and gerundive with *after* (*while admiring his efforts* and *after admiring his efforts*). There are several reasons to believe this is true. First, an *-en* complement must be participial and so will be compatible with *while* but incompatible with *after* (*while admired by many* but **after admired by many*). Next, explicit genitive subjects can be found with *after*-adjuncts but not with *while*-adjuncts (*after his leaving* but **while his leaving*)

(De Smet 2010: 1159f.). This is as expected if the former are gerundive and the latter are participial.[2]

The critical point for the current analysis is that these same control patterns are found even when the adjunct merely contains a gerund. That is, the gerund can be embedded within a non-gerundive NP in the adjunct, a position functional control cannot reach, and yet it shows the same strong preference for control by the matrix subject (8a) in addition to the potential to involve arbitrary anaphoric control (8b). These examples are particularly difficult to explain for generative accounts that use the Movement Theory of Control, such as Green (2019).

(8) a. After three days of preparing himself/*herself, Harry spoke to Sally about his concerns.

b. After three days of packing up, there was nothing left in the house.

Prepositions like *while* cannot be found with non-predicative NPs; they must make their complements predicative in their entirety. And so, *while* cannot select a non-predicative complement that has a predicative element embedded within it (**while three days of packing up* but *while busy with three days of packing up*).

Bare free adjuncts without any introductory prepositions[3] (9) pattern together with *while*-adjuncts as they similarly introduce predicative environments for phrases that might not normally involve predication (9c).

(9) a. Eating a sandwich in the park, John enjoyed his day off.

b. In trouble with his boss, John decided to call in sick.

c. A teacher at the local school, John had some insight into the situation.

I will therefore refer to adjuncts introduced by *after* as gerundive adjuncts and those introduced by *while* (or nothing) as empty absolute clauses, which can be compared with complete absolute clauses with explicit subjects (e.g., *His hands shaking, he attempted to operate the machine*). My reason for not calling them "participial adjuncts" is that the same patterns are found with verbless adjuncts (e.g., *while in love*, *when ready*). And "free adjunct" is insufficient as a cover term because it demands a prosodic gap; free adjuncts are a subset of the adjuncts which should be treated.

[2]For more on the importance of the genitive subject as a diagnostic, see Seiss (2008).

[3](9b) begins with a preposition, but it is part of the predicative element: John is described as in-trouble-with-his-boss.

4 An analysis of adjunct control in LFG

There are several points that need to be accounted for. First, the adjuncts we have examined all involve anaphoric control of a null pronoun PRO. Functional control from the matrix clause has been ruled out for some (namely, adjuncts with embedded gerunds like *after a year of complaining*) and so cannot be justified in the others because there are no substantive differences in control patterns (although see Donaldson (2021: 210-212) for an attempt to find differences). Next, empty absolute clauses introduce a predicative environment. In the case of free adjuncts like those in (9), the predicative environment appears without being selected by a preposition. Finally, gerundive adjuncts must admit explicit genitive subjects, in which case anaphoric control from outside the adjunct is not possible.

I will start by positing that the predicative environment in empty absolute clauses results from the introduction of a small clause. The small clause involves functional control between a null subject and the complement,[4] but the null subject itself is controlled anaphorically. This approach calls for f-structure without any corresponding overt elements in c-structure, and so the PRED value for the small clause will have to be constructionally specified (Dalrymple, Dyvik & Holloway King 2004).

This will allow me to propose f-structures for empty absolute clauses (both bare free adjuncts (10) and *while*-adjuncts (11)):

(10) a. Eating lunch, Roger talked.

b.
$$\left[\begin{array}{ll} \text{PRED} & \text{`TALK}\langle\text{SUBJ}\rangle\text{'} \\ \text{SUBJ} & \left[\begin{array}{ll}\text{PRED} & \text{`Roger'}\end{array}\right] \\ \text{ADJ} & \left\{ \left[\begin{array}{ll} \text{PRED} & \text{`BE}\langle\text{XCOMP}\rangle\text{SUBJ'} \\ \text{SUBJ} & \left[\begin{array}{ll}\text{PRED} & \text{`PRO'}\end{array}\right] \\ \text{XCOMP} & \left[\begin{array}{ll} \text{PRED} & \text{`EAT}\langle\text{SUBJ,OBJ}\rangle\text{'} \\ \text{SUBJ} & \\ \text{OBJ} & \left[\begin{array}{ll}\text{PRED} & \text{`LUNCH'}\end{array}\right] \end{array}\right] \end{array}\right] \right\} \end{array}\right]$$

[4] In this paper, I have assumed that the complement is open and therefore functionally controlled by the null subject, but a PREDLINK analysis after Butt et al. (1999) could also work.

(11) a. While eating lunch, Roger talked.

b. $$\left[\begin{array}{ll}
\textsc{pred} & \text{`}\textsc{talk}\langle\textsc{subj}\rangle\text{'} \\
\textsc{subj} & \left[\begin{array}{ll}\textsc{pred} & \text{`Roger'}\end{array}\right] \\
\textsc{adj} & \left\{\left[\begin{array}{ll}
\textsc{pred} & \text{`}\textsc{be}\langle\textsc{xcomp}\rangle\textsc{subj}\text{'} \\
\textsc{compform} & \textsc{while} \\
\textsc{subj} & \left[\begin{array}{ll}\textsc{pred} & \text{`}\textsc{pro}\text{'}\end{array}\right] \\
\textsc{xcomp} & \left[\begin{array}{ll}
\textsc{pred} & \text{`}\textsc{eat}\langle\textsc{subj,obj}\rangle\text{'} \\
\textsc{subj} & \\
\textsc{obj} & \left[\begin{array}{ll}\textsc{pred} & \text{`}\textsc{lunch}\text{'}\end{array}\right]
\end{array}\right]
\end{array}\right]\right\}
\end{array}\right]$$

Next, we can represent gerundive complements as having a structure in which the null pronoun could potentially be replaced by an explicit subject. I have assumed that this NP functionally controls the subject after Bresnan et al. (2016 [2001]: 316f.).[5]

(12) a. After eating lunch, Roger talked.

b. $$\left[\begin{array}{ll}
\textsc{pred} & \text{`}\textsc{talk}\langle\textsc{subj}\rangle\text{'} \\
\textsc{subj} & \left[\begin{array}{ll}\textsc{pred} & \text{`Roger'}\end{array}\right] \\
\textsc{adj} & \left\{\left[\begin{array}{ll}
\textsc{pred} & \text{`}\textsc{after}\langle\textsc{comp}\rangle\text{'} \\
\textsc{comp} & \left[\begin{array}{ll}
\textsc{pred} & \text{`}\textsc{eat}\langle\textsc{subj,obj}\rangle\text{'} \\
\textsc{poss} & \left[\begin{array}{ll}\textsc{pred} & \text{`}\textsc{pro}\text{'}\end{array}\right] \\
\textsc{subj} & \\
\textsc{obj} & \left[\begin{array}{ll}\textsc{pred} & \text{`}\textsc{lunch}\text{'}\end{array}\right]
\end{array}\right]
\end{array}\right]\right\}
\end{array}\right]$$

These representations allow us to capture all of the relevant information. Gerundive adjuncts (12) allow non-controlled alternatives with genitive subjects and do not have the necessary f-structure to enforce predication. Empty absolute clauses (10, 11) do not allow genitive subjects but do have the necessary XCOMP in f-structure to enforce predication. And although functional control is involved within the adjuncts, all control from the matrix clause

[5] Compare the treatment of *after* in (12) to how *while* is handled as a marker in (11). It would be perfectly reasonable to propose that *while*, too, has a PRED value and involves a similarly nested f-structure, but that would suggest that a null element should accomplish the same thing in (10). The nested approach may eventually prove to be the correct one.

(or elsewhere) is purely anaphoric.[6]

5 The incremental processing of adjuncts

What remains to be discussed is why anaphoric control should be obligatory in some circumstances and arbitrary in others. I believe that the best way to account for the facts is to assume that language users guess at a controller as soon as it becomes apparent that one is required (Donaldson 2021). This is in line with the accepted psycholinguistic position that anaphoric elements are interpreted immediately (Garrod & Sanford 1985, Sanford & Garrod 1989).

The adjuncts we have been examining can occur in initial, medial and final positions. When they are initial, it is not immediately clear how they will function with respect to upcoming linguistic material. As Diessel (2005: 456) points out, some free adjuncts (13a) are temporarily indistinguishable from gerundive subjects (13b), which exhibit arbitrary anaphoric control.

(13) a. Turning a sharp corner, Bill saw a dog.
 b. Turning a sharp corner was much easier with Bill's new car.

These initial adjuncts are processed immediately with reference to the discourse model, which is perpetually being updated, and so they exhibit arbitrary anaphoric control by entities that are associated with the speech act or are otherwise present in the discourse. Obligatory anaphoric control, which looks to the matrix clause for a controller, can be employed only after the matrix clause arrives. When a plausible competitor for control is made available through obligatory control, the result is potentially a garden path. And so in (14a), the hearer assumes that Fred is the driver until the possibility of Ted driving the car arrives. A pleonastic subject, on the other hand, rules out the possibility of obligatory control by the subject and results in smooth processing because the arbitrary guess can be maintained (14b). It is interesting to note that while (14a) is the one that causes the reader to stumble and reread the passage, (14b) is the one that falls afoul of the traditional rule that stipulates coreference with the subject of the matrix clause and would therefore be labeled as involving a so-called "dangling modifier" (Donaldson 2021: 1ff.).

[6]None of these structures involves XADJ, which should nevertheless be retained. Even if we limit ourselves to discussing adjuncts in English, there are many types that do not allow for extrasentential control. This topic is treated in the fourth chapters of Green (2018) and Landau (2021).

(14) a. Fred$_f$ sighed and stared at the road. $_{f \to t}$Driving at night, Ted$_t$ often fell asleep.

b. Fred$_f$ sighed and stared at the road. $_f$Driving at night, it was easy to fall asleep.

This incremental processing could be modelled through the step-by-step construction of f-structure found in Asudeh (2013) and Jones (2019).

Where we see an interesting divide is in the control of final adjuncts. The presence of the matrix clause precludes nearly all arbitrary options (15a). Obligatory control is the default here (15b).

(15) a. *Rust won't be an issue, being made of stainless steel.

b. This knife resists rusting, being made of stainless steel.

But a subset of arbitrary controllers is still available in final position: the collection of 'egophoric' pronouns described in Dahl (2000). Pronouns like *I*, *you*, and *one* do not use the antecedents that pronouns typically demand, as they are either deictic or arbitrary, and so adjuncts that are controlled in an equivalent way can appear in any position (Donaldson 2021: 138f.):

(16) a. There were several problems while contacting them.

b. The table should be set while taking care not to make noise.

c. A plastic tab broke while assembling the shelving unit.

d. The specified account will be charged after placing your order.

e. The weather was great after arriving.

Other than obligatory control, egophoric control is the only possibility that is available for final adjuncts because regular anaphoric reference back to established entities is no longer an option. Incidentally, the fact that inanimate controllers are ruled out for final adjuncts while egophoric controllers are not is probably behind the illusion that all controllers that are not matrix subjects must be logophoric, a view that drives many of the generative approaches.

6 Conclusion

In this paper, I have presented an approach to adjunct control that can account for varying control patterns without having those patterns result from structural differences. The structural differences that I did propose instead distinguish between gerundive adjuncts and open absolute clauses, the latter of which necessarily introduce predicative environments for their complements. This approach has better coverage of the empirical facts than its alternatives: it does not have to marginalise non-subject control, it provides a reason for the abundance of experiencer control without incorrectly stipulating logophoricity, and it can account for the fact that the controlled element can be embedded within a non-predicative adjunct.

References

Asudeh, Ash. 2013. Directionality and the production of ungrammatical sentences. *Studies in Linguistics. Special Issue on Directionality of Phrase Structure Building* 6. 83–106.

Börjars, Kersti, Rachel Nordlinger & Louisa Sadler. 2019. *Lexical-Functional Grammar: An Introduction*. Cambridge: Cambridge University Press.

Bresnan, Joan. 1982. Control and complementation. *Linguistic Inquiry* 13(3). 343–434.

Bresnan, Joan, Ash Asudeh, Ida Toivonen & Stephen Wechsler. 2016 [2001]. *Lexical-Functional Syntax, Second Edition*. Oxford: Blackwell.

Butt, Miriam, Tracy Holloway King, María-Eugenia Niño & Frédérique Segond. 1999. *A Grammar Writer's Cookbook*. Stanford: CSLI Publications.

Dahl, Östen. 2000. Egophoricity in discourse and syntax. *Functions of Language* 7(1). 37–77.

Dalrymple, Mary. 2001. *Syntax and Semantics Volume 34: Lexical Functional Grammar*. Bingley: Emerald Group.

Dalrymple, Mary, Helge Dyvik & Tracy Holloway King. 2004. Copular Complements: Closed or Open? In *Proceedings of the LFG04 Conference*, 188–198. Stanford: CSLI Publications.

Dalrymple, Mary, John J. Lowe & Louise Mycock. 2019. *The Oxford Reference Guide to Lexical Functional Grammar*. Oxford: Oxford University Press.

Davies, Mark. 2018. *The iWeb Corpus*. `https://www.english-corpora.org/iweb/`.

De Smet, Hendrik. 2010. English *-ing*-clauses and their problems: The structure of grammatical categories. *Linguistics* 48(6). 1153–1193.

Diessel, Holger. 2005. Competing motivations for the ordering of main and adverbial clauses. *Linguistics* 43(3). 449–470.

Donaldson, James. 2021. *Control in Free Adjuncts: The "Dangling Modifier" in English*. University of Edinburgh dissertation.

Garrod, Simon & Anthony J. Sanford. 1985. On the real-time character of interpretation during reading. *Language and Cognitive Processes* 1(1). 43–59.

Green, Jeffrey Jack. 2018. *Adjunct Control: Syntax and Processing*. University of Maryland dissertation.

Green, Jeffrey Jack. 2019. Non-subject control of temporal adjuncts. In E. Ronai, L. Stigliano & Y. Sun (eds.), *Proceedings of the Fifty-Fourth Annual Meeting of the Chicago Linguistic Society*, 137–148. Chicago: Chicago Linguistic Society.

Jones, Stephen M. 2019. *Modelling an incremental theory of Lexical Functional Grammar*. University of Oxford dissertation.

Kroeger, Paul R. 2004. *Analyzing Syntax: A Lexical-Functional Approach*. Cambridge: Cambridge University Press.

Landau, Idan. 2017. Adjunct control depends on voice. In Claire Halpert, Hadas Kotek & Coppe van Urk (eds.), *A Pesky Set: Papers for David Pesetsky*, 93–102. Cambridge, MA: MIT Working Papers in Linguistics.

Landau, Idan. 2021. *A Selectional Theory of Adjunct Control*. Cambridge, Mass.: MIT Press.

Mohanan, K. P. 1983. Functional and anaphoric control. *Linguistic Inquiry* 14(4). 641–674.

Sanford, A. J. & S. C. Garrod. 1989. What, when, and how?: Questions of immediacy in anaphoric reference resolution. *Language and Cognitive Processes* 4(3-4). SI235–SI262.

Seiss, Melanie. 2008. The English *-ing* Form. In *Proceedings of the LFG08 Conference*. Stanford: CSLI Publications.

Stump, Gregory Thomas. 1981. *The Formal Semantics and Pragmatics of Free Adjuncts and Absolutes in English*. Ohio State University dissertation.

Williams, Edwin. 1992. Adjunct control. In Richard K. Larson, Sabine Iatridou, Utpal Lahiri & James Higginbotham (eds.), *Control and Grammar*, 297–322. Dordrecht: Kluwer Academic Publishers.

Relative clauses in Wolof: An LFG account

Cheikh M. Bamba Dione
University of Bergen

Proceedings of the LFG'21 Conference

On-Line

Miriam Butt, Jamie Y. Findlay, Ida Toivonen (Editors)

2021

CSLI Publications

pages 104–124

http://csli-publications.stanford.edu/LFG/2021

Keywords: across-the-board, anaphoric control, LFG, parasitic gaps, relative clauses, resumptive pronouns, weak crossover, reconstruction, Wolof.

Dione, Cheikh M. Bamba. 2021. Relative clauses in Wolof: An LFG account. In Butt, Miriam, Findlay, Jamie Y., & Toivonen, Ida (Eds.), *Proceedings of the LFG'21 Conference, On-Line*, 104–124. Stanford, CA: CSLI Publications.

Abstract

This paper provides a description of the syntax of relative clauses in Wolof and presents a formal analysis of the facts described building on existing LFG work on relatives. The paper explores the distribution of the resumption and gap relativization strategies, providing a discussion of the status of the kinds of resumptive pronouns found in Wolof.

1 Introduction

This paper presents an analysis of Wolof relative clauses in the framework of Lexical Functional Grammar (LFG) (Bresnan 2001, Dalrymple 2001). Wolof is a West Atlantic language, an important branch of the Niger-Congo language family (Sapir 1971). Building on existing LFG work on relatives (Dalrymple 2001, Asudeh 2004, Camilleri and Sadler 2011a,b), I propose a uniform analysis for the types of relative clauses found in Wolof and show how such constructions can be accommodated in LFG quite straightforwardly. The analysis will also give a particular focus on the distribution of the resumption and gap relativization strategies. I will try to provide evidence for the status of the resumptive pronouns found in that language. I hope that this contribution will also lay the groundwork for a comparison to pronoun resumption in LFG.

This paper is organised as follows. Section 2 outlines and illustrates the basic morphosyntactic characteristics of relative clauses in Wolof. Section 3 provides a brief introduction to work on relative clauses in LFG which we build on. Section 4 presents a basic LFG analysis of the Wolof relative clauses. Section 5 discusses recent work on pronoun resumption in LFG, and section 6 presents the analysis proposed for Wolof resumptive pronouns. Section 7 concludes the discussion.

2 General properties of Wolof relative clauses

Similar to Bantu languages, Wolof has noun classes (McLaughlin 2010, Torrence 2013, Dione 2014b): 8 singular classes, and 2 plural classes. The indexes (or markers) for singular noun classes are: *b, g, j, k, l, m, s, w*, and for plural noun classes are: *y* and *ñ*.[1] Unlike Bantu languages, in Wolof, class membership is typically expressed by a class index on nominal dependents such as determiners and relative pronouns rather than on the noun itself.

Wolof has three types of determiners, as illustrated in (1). Morphologically, each determiner consists of a noun class index (CL) and a vowel, yielding the following patterns: *CL-i*, *CL-a*, and *CL-u*. Determiners with the CL-*i* pattern are interpreted as definite and proximal (DFP). Likewise, determiners with the CL-*a*

[1] Although the *k* and *ñ* classes are associated with humans, while *l* and *y* are typically non-human classes, the Wolof noun class system generally lacks semantic coherence (McLaughlin 2010).

pattern are interpreted as definite and distal (DFD). In contrast, determiners that exhibit the *a-CL* pattern are indefinite (IND). Definite determiners invariably follow the noun, while the indefinite determiner invariably precedes the noun. [2]

(1)	a. *jën w-i* fish CL-DFP "the fish here"	b. *jën w-a* fish CL-DFD "the fish there"	c. *a-w jën* IND-CL fish "a/some fish"

Furthermore, Wolof has three basic types of relative clauses (Torrence 2005, 2013), as illustrated in (2). These are distinguished by their 'relative markers' (as underlined). The relative markers are identified by their form which is strikingly similar to the determiners. The interpretation of the antecedent varies according to the form of the relative marker. In relative clauses like (2a) where the relative marker has the *CL-i* pattern, the antecedent is interpreted as definite and proximal (spatially, temporally, or in the discourse). On the other hand, the antecedent of relative markers that have the *CL-a* pattern, as in (2b), is interpreted as definite and distal. In contrast, when the relative marker occurs in the *CL-u* pattern, the antecedent is interpreted as indefinite, as in (2c). Also note the difference between the indefinite relative marker and the indefinite determiner. The former has the *CL-u* pattern, while the latter exhibits the *a-CL* pattern.

(2) a. *jën* ***w-i*** *janq b-i lekk* i-*Relative Clause*
fish CL-i girl CL-DFP eat
"the fish here that the girl ate"

b. *jën* ***w-a*** *janq b-i lekk* a-*Relative Clause*
fish CL-a girl CL-DFP eat
"the fish there that the girl ate"

c. *jën* ***w-u*** *janq b-i lekk* u-*Relative Clause*
fish CL-u girl CL-DFP eat
"some fish that the girl ate"

As examples (3) show, determiners may co-occur with the relative markers. The optionality brackets in (3) denote the fact that the determiner should be interpreted as true optionality. As Torrence (2013) pointed out, this is presumably because it is possible to recover the content of the determiner from the form of the relative marker. Moreover, relative markers and determiners (if present) obligatorily agree both with the relativized NP in noun class (otherwise the clause becomes ungrammatical). In (3), the relative markers and the determiners agree with the noun *jën* 'fish' in the *w* class.

[2] Abbreviations in the glosses: APPL: applicative: CL: noun class; COP: copula; DFP: definite proximal; DFD: definite distal; +F: finite; IPFV: imperfective; NDF: indefinite; NSFOC: non-subject focus; O: object; PFV: perfective; PL: plural; POSS: possessive; REL: relative; SUBJ: subject; SG: singular; 1, 2, 3: first, second, third person.

(3) a. *jën* ***w-i*** *janq b-i lekk (w-i)* i-*Relative Clause*
fish CL-i girl CL-DFP eat CL-DFP
"the fish here that the girl ate"

b. *jën* ***w-a*** *janq b-i lekk (w-a)* a-*Relative Clause*
fish CL-a girl CL-DFP eat CL-DFD
"the fish there that the girl ate"

c. *(a-w) jën* ***w-u*** *janq b-i lekk* u-*Relative Clause*
IND-CL fish CL-u girl CL-DFP eat
"some fish that the girl ate"

In contrast to English (4) and many other languages, in Wolof, the relative markers must be overt. Dropping them would make the clause ungrammatical (5).

(4) the fish __ the girl ate

(5) **jën __ janq b-i lekk*
fish __ girl CL-DFP eat
FOR: "some/the fish the girl ate"

The three types of relative markers may be used both in short or immediate distance dependencies (IDD) and in long distance dependencies (LDD) for relativization on all clause internal grammatical functions (GF). Thus, for Wolof, there seems to be no accessibility hierarchy effects (Keenan and Comrie 1977). It is possible to relativize subject, direct and indirect objects, possessors, as well as obliques and adjuncts. Examples (6b-6e) illustrate cases involving relativization of SUBJ. Example (6a) gives the base sentence.[3] In (6b), because the subject is in highest position, only a gap is allowed; a resumptive pronouns (RP) is not available. In contrast, in cases involving LDDs (6c-6e), pronoun resumption (e.g. using *mu-a*, *na* or *mu*) is obligatory. Thus, Wolof seems to be subject to the familiar Highest Subject Restriction (HSR)(Borer 1984, McCloskey 1990). This principle prohibits RPs in the highest subject position in unbounded dependencies. Many languages have been reported to be subject to this constraint, including Irish (McCloskey 1990), Hebrew (Shlonsky 1992) and Maltese (Camilleri and Sadler 2011a).

(6) a. *Janq b-i jox na góor g-i caabi j-i.*
girl CL-DFP give 3SG man CL-DFP key CL-DFP
"The girl gave the key to the man."

b. *janq b-i jox (*na) góor g-i caabi ji*
girl CL-REL give (*3SG) man CL-DFP key CL-DFP
"'the girl that gave the man the key'"

[3] The examples in (6) involve relative markers with the *CL-i* patterns, but relativization based on the *CL-a* or *CL-u* pattern would give similar constructions (thus, examples with the other types of relative markers may be omitted for lack of space). Also, to avoid confusion between the determiners and the relative markers, I will use the gloss (CL-REL) for relative markers in the rest of the paper.

c. *janq b-i ñu wax ni *(mu-a) jox góor g-i caabi*
girl CL-REL 3PL say that *(3SG-SFOC) give man CL-DFP key
j-i
CL-DFP
"the girl that they said that it's her who gave the man the key"

d. *janq b-i ñu wax *(mu) jox góor g-i caabi j-i*
girl CL-REL 3PL say *(3SG) give man CL-DFP key CL-DFP
"the girl that they said that she gives the man the key"

e. *janq b-i ñu wax ni jox *(na) góor g-i caabi j-i*
girl CL-REL 3PL say that give (3SG) man CL-DFP key CL-DFP
"'the girl that they said that she gave the man the key'"

In Wolof, relativization from embedded clauses typically involves embedded clefts (6c). There are three types of clefts in the language (Robert 1991, Torrence 2005, Dione 2012): subject, non-subject, and verb clefts. Clefting can be used to put the subject (6c), the predicate, or any constituent which is neither subject nor main verb into focus (non-subject cleft). For instance, the embedded clause in (6c) is a subject cleft, as indicated by the focus marker *mu-a*, which expresses 3SG subject (*mu*) and the subject focus (SFOC) copula *a*. As examples (6d-6e) show, it is also possible to have other embedded complement clause types such as narrative clauses (6d) and neutral perfective clauses (6e).[4]

The examples in (7) illustrate the relativization of primary objects (OBJ).[5] Here also, in short distance dependencies, only a gap is permitted, excluding an RP from the highest OBJ positions (7a). However, when extracting from the object position in long paths, there are two possibilities. If the embedded clause is a non-subject cleft non-subject cleft (NSC) (7b) or a non-finite complement clause (7d), then a gap and RP are freely interchangeable. Otherwise, in all other embedded clauses (including the other types of clefts), pronoun resumption is compulsory, as in (7c). In (7b), the embedded clause is a non-subject cleft, as indicated by *la*, which consists of the non-subject focus (NSFOC) copula *la* and an empty 3SG morph. Also, note that resumptive pronouns in non-subject clefts are typically strong pronouns (e.g. *moom*). These are very similar to French emphatic pronouns (e.g. *moi*, *toi*, *lui*,..) in the sense that they are only used in isolation, in emphatic positions, as objects of preposition, in dislocated positions, and cleft sentences, but otherwise never as direct or indirect objects (Zribi-Hertz and Diagne 2002). Object clitics, e.g. *ko* (3SG.O = third singular object) as in (7d), appear in object positions instead. Relativization on secondary objects (OBJ-TH) and applied objects (OBJ-APPL)[6] occur in a similar way to relativization of primary objects.

[4]For a detailed discussion of Wolof clause types, see Torrence (2005), Dione (2020).

[5]Wolof is a symmetrical language. The status of primary vs. secondary object is determined by word order (see Dione (2014a) for more details).

[6]For a more detailed discussion of applicative structures in Wolof, see e.g. Dione (2013), Harris (2015). Dione (2013) provided an LFG-based analysis of these constructions.

(7) a. *góor g-i janq b-i jox (*ko) caabi j-i*
man CL-REL girl CL-DFP give (*3SG.O) key CL-DFP
"the man that the girl gave the key"

b. *góor g-i Awa foog ni (moom) la janq b-i*
man CL-REL Awa think that (him) NSFOC.3SG girl CL-DFP
jox caabi j-i
give key CL-DFP
"the man that Awa thinks that the girl gave the key"

c. *góor g-i Awa foog ni janq b-i jox na *(ko)*
man CL-REL Awa think that girl CL-DFP give 3SG *(3SG.O)
caabi j-i
key CL-DFP
"the man that Awa thinks that the girl gave the fish"

d. *góor g-i xale y-i bëgg janq b-i jox (ko) caabi*
man CL-REL child CL-DFP want girl CL-DFP give (3SG.O) key
j-i
CL-DFP
"the man that the children want that the girl give him the key"

Relativization of obliques (OBL) and adjuncts (ADJ) is quite complex. In contrast to term functions (i.e. SUBJ, OBJ, OBJ-TH), relativization on OBL and ADJ typically requires valency change in terms of an applicative construction. This requirement holds for short distance dependencies (8b), but also for LDDs where the domain of extraction is a non-subject cleft (8c). For the other LDD cases, the valency change seems to be compulsory for the extraction of ADJ only (not OBL).

For instance, relativization of the oblique argument (i.e. *góor gi* 'the man') in (8a) triggers applicative derivation (8b-8c) with the suffix *-al*, by virtue of which an OBL argument is typically promoted to an applied object (OBJ-APPL) with the semantic role of beneficiary, recipient, or comitative (8). Here too, the RP is excluded in IDD (8b), but may alternate with a gap in LDDs that involve non-subject clefts (8c). As (8d) shows, if the domain of extraction in long paths is a clause other than an NSC, then the applicative derivation is prohibited (i.e. there is no valency change) and the presence of a (strong) resumptive pronoun is required.

(8) a. *Janq b-i wax na ak góor g-i ci kër g-i.*
girl CL-DFP talk 3SG to man CL-DFP in house CL-DFP
"The girl talked to the man in the house."

b. Oblique → Applied Object (IDD)
góor g-i janq b-i wax-(**al**) (*ko) ci kër g-i*
man CL-REL girl CL-DFP talk-APPL (*him) in house CL-DFP
"the man that the girl talked to in the house"

c. Oblique → Applied Object (LDD, non-subject cleft)

góor g-i ñu foog ni (moom) la janq b-i
man CL-REL 3PL think that (him) NSFOC.3 girl CL-DFP
wax-(**al**) ci kër g-i*
talk-APPL in house CL-DFP

"the man that they think that the girl talked to in the house"

d. Oblique → Oblique (LDD, neutral)

*góor g-i ñu foog ni janq b-i wax-(*al) na ak*
man CL-REL 3PL think that girl CL-DFP talk-(*APPL) 3SG to
**(moom) ci kër g-i*
*(him) in house CL-DFP

"the man that they think that the girl talked to in the house"

Likewise, relativization of a locative adjunct, as in (8a), triggers applicative derivation by which the adjunct is promoted to a special kind of oblique (9a-9b), i.e. OBL-LOC (for locative oblique). Here, the applicative derivation is compulsory (both in IDD and LDD) and occurs by means of the suffix *-e*, which introduces participants with an instrumental (10b), locative (9a), or manner role. The distribution of gap and RP is similar to what we observed for relativization of OBL.

(9) a. Locative adjunct → OBL-LOC (IDD)

kër g-i janq b-i wax-(**e**) (*fa) ak góor g-i*
house CL-REL girl CL-DFP talk-APPL (*there) with man CL-DFP

"the house where the girl talked to the man"

b. Locative adjunct → OBL-LOC (LDD, non-subject cleft)

*kër g-i ñu foog ni *(fa) la janq b-i*
house CL-REL 3PL think that *(there) NSFOC.3 girl CL-DFP
wax-(**e**) ak góor g-i*
talk-APPL with man CL-DFP

"the house where they think that the girl talked to the man"

As with locative adjuncts, relativization of instrumental adjuncts, as in (10), also triggers an obligatory applicative process (with the *-e* form). However, in the latter case, the instrumental becomes an applied object rather than an oblique (10b-10d). Here again, RP is prohibited in IDD (leaving a gap) as in (10b), but required in LDD if the domain of extraction is not a non-subject cleft, e.g. as in (10d), which is a perfective affirmative clause; otherwise, the RP may alternate with a gap (10c).

(10) a. *Janq b-i ubbi na bunt b-i ak caabi j-i.*
girl CL-DFP open 3SG door CL-DFP with key CL-DFP

"The girl opened the door with the key."

b. Instrumental adjunct → Object (IDD)

caabi j-i janq b-i ubb-(**e**) (*ko) bunt b-i*
key CL-REL girl CL-DFP open-APPL (*it) door CL-DFP

"the key that the girl opened the door with"

c. Instrumental adjunct → Object (LDD, non-subject cleft)

caabi j-i ñu foog ni (moom) la janq b-i
key CL-REL 3PL think that (it) NSFOC.3 girl CL-DFP
ubb-(**e**) bunt b-i*
open-APPL door CL-DFP

"the key that they think that the girl opened the door with"

d. Instrumental adjunct → Object (LDD, neutral perfective)

caabi j-i ñu foog ni janq b-i ubb-(**e**) na *(ko)*
key CL-REL 3PL think that girl CL-DFP open-APPL 3SG *(it)
bunt b-i
door CL-DFP

"the key that they think that the girl opened the door with"

Finally, a gap is not licensed as POSS (11).

(11) a. *xale b-i ma xam yaay-*(am)*
child CL-REL 1SG know mother-POSS.3SG

"the child whose mother I know"

b. *xale b-i ñu foog ni xam naa yaay-*(am)*
child CL-REL 3PL think that know 1SG mother-POSS.3SG

"the child that they think I know his mother"

Table (1) summarises the distribution pattern for the Wolof relative clauses in both IDD and LDDs. For IDDs, only gap is allowed, except for relativization of POSS (which always requires pronoun resumption). For LDDs, gap is typically permitted only if the domain of extraction is a non-subject cleft (NSC) or a non-finite complement clause; otherwise only RPs are allowed. Furthermore, relativization of OBL in IDD requires applicative derivation. In contrast, extraction of OBL from a long path triggers applicative if the domain of extraction is a non-subject cleft; otherwise the OBL remains in situ and applicative derivation is not permitted. Relativization of locative or instrumental ADJ triggers both non-subject clefting and valency change in terms of applicative derivation. This distribution raises some interesting issues that will be discussed below.

GF	IDD	LDD	Restriction	GF	IDD	LDD	Restriction
SUBJ	Gap	RP		OBL	Gap	Gap/RP	+APPL in IDD/LDD with NSC, otherwise -APPL
OBJ	Gap	Gap/RP		Loc. ADJ	Gap	Gap/RP	+APPL
OBJ-TH	Gap	Gap/RP		Ins. ADJ	Gap	Gap/RP	+APPL
OBJ-APPL	Gap	Gap/RP	+APPL	POSS	RP	RP	

Table 1: Summary for Wolof Relatives.

3 Analysis of relative clauses in LFG

In LFG, relative clauses, like topicalization and wh-questions, are instances of long-distance dependencies (LDD) (Dalrymple 2001, Bresnan 2001). LDDs are constructions where "a displaced constituent bears a syntactic function usually associated with some other position in the sentence" (Dalrymple 2001, p. 389).

Unlike constructions such as topicalization, relative clauses involve two long-distance dependencies. The first dependency holds between the displaced (or fronted) constituent (also called filler), e.g. the NP *kër* 'house' in (12a-12c), and the within-clause grammatical (GF) it fills (e.g. OBJ). The filler plays two roles simultaneously: it bears the syntacticized TOPIC function (Bresnan and Mchombo 1987) and the within-clause GF it fills. The relation between the two positions must be controlled according to the Extended Coherence Condition (Dalrymple 2001, p. 390), which basically states that, in order for the f-structure to be coherent, the TOPIC must be linked to a GF within the clause. The second dependency holds between the relative pronoun and its position within the fronted phrase. Following previous works (Butt et al. 1999, Dalrymple 2001, Falk 2001), the relative pronoun is analyzed at the f-structure level as contributing to the RELPRO feature within the relative clause.

As examples (12a-12c) show, in relative clauses, the distance between the fronted material and the within-clause GF can be local (12a) but also potentially unlimited (12b-12c), hence the name *long-distance dependencies*. In Wolof, similar to English (Dalrymple 2001) and many other languages, the path can pass through any number of COMP (12b) or XCOMP (12c) clauses with some restrictions.

(12) a. *kër g-i jigéen j-i tabax (*ko)*
house CL-REL woman CL-DFP build (*it)
"The house that the woman built"

b. *kër g-i ñu wax ni Awa foog na ni jigéen j-i*
house CL-REL 3PL say that Awa think 3SG that woman CL-DFP
*tabax na *(ko)*
build 3SG *(it)
"The house that they said that Awa thinks that the woman have built"

c. *kër g-i ñu wax ni Awa foog na ni jigéen j-i*
house CL-REL 3PL say that Awa think 3SG that woman CL-DFP
*bëgg na *(ko) tabax*
want 3SG *(it) build
"The house that they said that Awa think that the woman wants to build"

Furthermore, while in (12a-12c), the TOPIC also bears the OBJ function, it might be the SUBJ or OBL, and so on in other examples. In LFG, this situation is accounted for in terms of "functional uncertainty" (Dalrymple 2001, Austin 2001) about the grammatical function of the TOPIC. This is typically expressed in terms of equations like (13) which links the TOPIC to a grammatical function as specified by the symbol GF which represents a disjunction of all relevant grammatical functions (i.e. SUBJ, OBJ, OBJ-TH, OBL, and so on).

(13) $(\uparrow \text{TOPIC}) = (\uparrow \{\text{COMP} \mid \text{XCOMP}\}^* \text{ GF})$

There are typically restrictions on the relation between the filler and the within-clause GF in long-distance dependency constructions. These restrictions are defined in terms of island constraints (Falk 2001), including complex noun phrase constraints (CNPC), adjunct constraints and *wh*-island constraints. To satisfy such constraints, resumptive pronouns might provide the possibility (not always as discussed below) to fill the gaps in the domain of extraction. The analysis of resumptive pronouns in LFG in general and in Wolof in particular will be addressed in sections 5 and 6, respectively. Before that, section 4 presents the basic analysis of Wolof relative clauses I propose within the LFG framework.

4 Basic Analysis of Wolof Relative Clauses in LFG

To account for the relative clauses in Wolof, I will draw on the analysis of English restrictive relative clauses provided in Dalrymple (2001). This approach has inspired the analysis of relative clauses for languages like Modern Greek (Chatsiou 2010) and Maltese (Camilleri and Sadler 2011a,b). In the same spirit, I propose the following c-structure rules in (14-15) for the analysis of Wolof relative clauses. The rule in (14) states that a relativized noun phrase (NP) consists of nominal head (NOM)[7] and *CP* adjuncts. The f-structure of the *CP* is assumed to be a member of the set of modifiers of the noun phrase, i.e. $\downarrow \in (\uparrow \text{ADJ})$.

(14) $\text{NP} \rightarrow \underset{\uparrow=\downarrow}{\text{NOM}} \quad \underset{\downarrow \in (\uparrow \text{ADJ})}{\text{CP*}}$

The rule for the *CP* relative is given in (15), which states that the CP consists of an obligatory relative phrase constituent *RelP* and an *IP*.

(15) $\text{CP} \rightarrow \underset{\substack{(\uparrow \text{TOPIC})=\downarrow \\ (\uparrow \text{TOPIC})=(\uparrow \text{RTOPICPATH}) \\ (\uparrow \text{RELPRO PRON-TYPE}) =_c \text{rel} \\ @\text{REL-FEAT}}}{\text{RelP}} \quad \underset{\uparrow=\downarrow}{\text{IP}}$

For Wolof, *RelP* is the specifier of *CP* and consists just of a relative marker, which is analyzed as a relative pronoun. In previous work, the Wolof relative markers have received different analyses, including connectives (Voisin-Nouguier 2002)

[7] NOM includes a wide variety of nominals: common nouns, proper names, quantifiers (e.g. *ñépp* 'everybody') and strong pronouns (e.g. *moom* 'him' as in *moom mi Awa gis* 'him who Awa saw').

and complementizers (Torrence 2013). On my analysis, however, the relative markers are relative pronouns. This is because, in most of the Wolof relative clauses examples we discussed so far, there is clearly a gap. For instance, in (16), the gap in the relative clause shows the absence of relativized *meew* 'milk'. Here, *mi* is the only word which can reasonably contribute the f-structure required for the verb to find an OBJ-TH argument. A different analysis for (16), for example one wherein *mi* is some kind of complementizer which introduces topic and certain agreement features, but does not contribute a semantic predicate on its own, would create a real problem in terms of the LFG wellformedness principles (Bresnan 2001). Furthermore, there do not appear to be cases that clearly rule out the pronominal nature of *mi* (and similar relative markers). It seems like the relative marker must be a relative pronoun. From the perspective of LFG they contribute a PRED 'pro'.

(16) *meew m-i janq b-i jënd-al góor g-i* __
milk CL-REL girl CL-DFP buy-APPL man CL-DFP
"the milk which the girl bought for the man"

The fact that *RelP* consists just of a relative pronoun contrasts with the situation in languages like English where several phrases (e.g. NPs, PPs, APs, and AdvP) can instantiate *RelP* (Dalrymple 2001, p. 404). This is because, in English, the relative pronoun lures some additional material (e.g. *whose book*; *whose brother's book*; *a friend of whose brother*; *in which*;...) along with it when moving to the front of the sentence. This phenomenon, known as pied piping (Ross 1967), does not seem to occur in Wolof relative clauses. Thus, the possible instantiations of *RelP* are basically relative pronouns.

The first equation (↑ TOPIC)=↓ in (15) constrains the f-structure associated with *RelP* to bear the TOPIC role in the f-structure. Subsequently, the second equation (↑ TOPIC)=(↑ RTOPICPATH) ensures that the TOPIC function also fills a within-clause GF, as required by the Extended Coherence Condition. RTOPICPATH represents the long-distance path relating these two positions and is defined for Wolof as given in (18). The third constraint (↑ RELPRO PRON-TYPE) $=_c$ rel requires the value of the RELPRO attribute to be a relative pronoun.

The definition of @REL-FEAT is given in (17). This contains constraints that enforce agreement between the head noun and the relative pronoun. These constraints unify all class, number, and person information. In other words, the annotations (↑ RELPRO NUM)=(↑ NUM) and (↑ RELPRO PERS)=(↑ PERS) state that RELPRO must have a relative pronoun, and its NUM and PERS must match the NUM and PERS of the relativized NP. The annotation (↑ RELPRO CLASS) = (↑ NOUN-CLASS) puts similar constraints regarding noun class agreement. The symbol DIRGF (19) encodes the direct (nominal) grammatical functions.

(17) REL-FEAT ≡ (↑ RELPRO NUM) = (↑ NUM)
(↑ RELPRO PERS) = (↑ PERS)
(↑ RELPRO CLASS) = (↑ CLASS)

(18) RTOPICPATH $\equiv$ {COMP | XCOMP}* DIRGF | OBL-LOC
@APPL-FEAT

(19) DIRGF $\equiv$ SUBJ | OBJ | OBJ-APPL | OBJ-TH

As we saw in section 2, extraction of a locative adjunct triggers applicative derivation with the argument being promoted to OBL-LOC. This requirement is encoded in @APPL-FEAT, which is defined as shown in (20). This additional condition ensures that the f-structure of the domain of extraction contains the attribute APPLICATIVE with value '+', but also that the morphological form of the derivation suffix be *-e* to avoid ambiguity with other types of applicatives.

(20) APPL-FEAT $\equiv$ ($\uparrow$ APPLICATIVE) $=_c$ +
($\uparrow$ APPL-FORM) $=_c$ e

The c- and f-structure representations associated with example (21) are given in Figure 1 (some minor morphosyntactic features are omitted for lack of space). As the f-structure shows, the TOPIC function is coindexed with OBJ expressing the dependency between the filler and the grammatical function from which it has been extracted. The other dependency, which involves the relative pronoun and its position is also made visible through co-indexation of TOPIC with RELPRO. Agreement (in number, person, and noun class) between the relativized NP and the relative pronoun is ensured by the constraints given in (17). Otherwise, the resulting f-structure would be deemed ungrammatical.

(21) *kër g-i xale y-i tabax*
house CL-REL child CL-DFP build
"the house that the children built"

The lexical entry for the relative pronoun *gi* is shown in (22). The relative pronoun specifies number, person, noun class and deixis features of the fronted material. It also indicates the type of pronoun (here relative). A different pronoun such as *ba* would have almost identical features, except for the DEIXIS attribute, which would have the value *distal*. In contrast, the relative pronoun *bu* would lack the DEIXIS attribute.

(22) *gi* PRON ($\uparrow$ PRED)='pro'
($\uparrow$ NUM)=sg
($\uparrow$ PERS)=3
($\uparrow$ CLASS G)=+
($\uparrow$ DEIXIS)=prox
($\uparrow$ PRON-TYPE)=rel
@ANTPROAGR

Another important constraint that needs to be handled is agreement between the antecedent, the relative pronoun and the determiner (if present). As mentioned above, all these three elements must agree in number, person, definiteness, and noun class. For instance, the c-structure and f-structure of the determiner phrase *DP house gii* 'the house' are given in Figure 2. The determiner introduces a *DET*

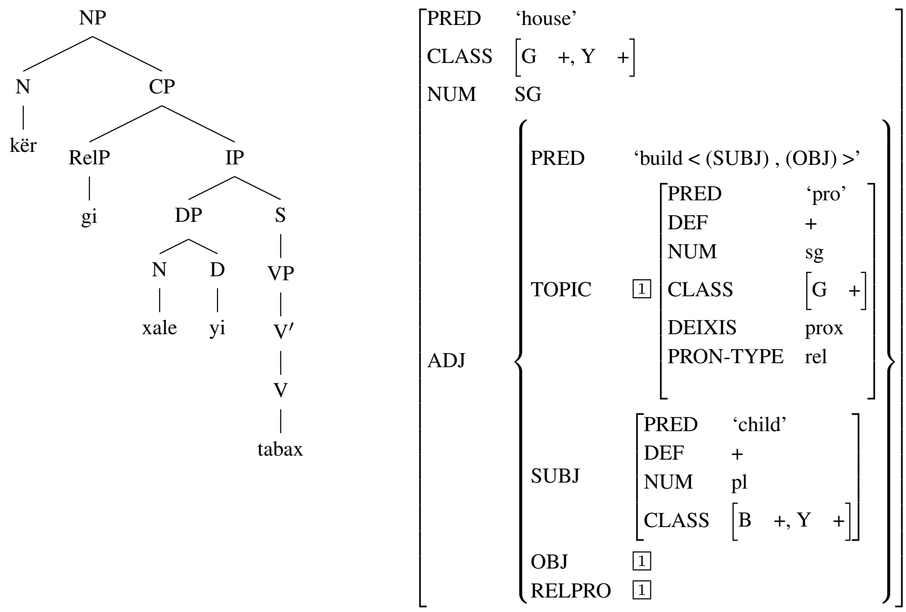


Figure 1: C- and f-structure of example (21)

feature under *SPEC* that indicates the semantic predicate *gi*, the deixis (proximal) and the type of the determiner (e.g. definite). It also specifies the person and number of the structure. Agreement between the determiner and the noun is controlled via a constraining equation — not displayed here — which, for instance, makes sure that the determiner *gi* agrees with the noun *kër* in the *G* class, i.e. a noun with the f-structure [G +].

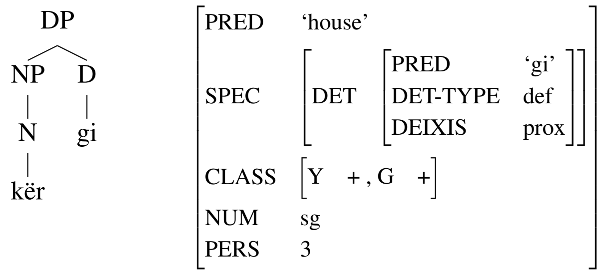


Figure 2: C-structure and f-structure of the DP *kër gi* 'the house'

The constraints defined in ANTPROAGR as shown in (23) enforce agreement between the antecedent and the relative pronoun (and indirectly agreement between the antecedent and the determiner).

(23) ANTPROAGR $\equiv$ ((ADJ $\in$ PATH* $\uparrow$) SPEC DET DET-TYPE) = def
((ADJ $\in$ PATH* $\uparrow$) SPEC DET DEIXIS) = prox
((ADJ $\in$ PATH* $\uparrow$) CLASS G) = +
((ADJ $\in$ PATH* $\uparrow$) NUM) = sg
((ADJ $\in$ PATH* $\uparrow$) PERS) = 3

(24) PATH = {COMP|XCOMP}

5 Resumptive Pronouns in LFG

As one of the earliest work on pronoun resumption in LFG, Falk (2002) considers resumptive pronouns as elements that are not licensed in the normal way by functional uncertainty equations, but rather by establishing a referential (anaphoric) identity between the two positions. He considered that this analysis is able to account for the similarities and differences between gaps and resumptive pronouns. Other subsequent works in LFG, including (Asudeh 2011, Camilleri and Sadler 2011a), make a key distinction between (i) true resumptive pronouns (TRP), (ii) gaps and (iii) 'false' resumptive (or intrusive) pronouns (FRP). TRPs are bound pronouns whereas gaps are bound variables: both are bound elements. TRPs are grammatically licensed bound pronouns, while FRPs are not grammatically licensed (but rather a processing or performance phenomenon). These two types of pronouns display different properties that can be summarized as follows. True resumptives permit binding by a quantifier resisting an e-type interpretation (*every, each, no*) as in (25a), support a list answer (25b), and support functional answers to questions. In contrast, intrusive pronouns do not support any of the aforementioned properties. These examples are taken from Camilleri and Sadler (2011a).

(25) a. I'd like to review every book that Mary couldn't remember if she'd read TRP/*FRP before.
b. Which of the linguists do you think if Mary hires TRP/*FRP everyone will be happy? (—— Chris, Daniel or Bill).

Asudeh (2011) (building on McCloskey (1990)) made a distinction between two types of true resumptive pronouns: syntactically active resumptives (SAR) and syntactically inactive resumptives (SIR). SARs do not behave like gaps and are instances of anaphorically bound pronouns in the syntax. These are the types of RPs found in languages like Irish and Hebrew (Asudeh 2011). On the other hand, SIRs display gap-like properties, meaning that they are functionally controlled. The resumptive pronoun is treated as the bottom of a filler-gap dependency by restricting out the pronominal PRED value. According to Asudeh (2011), Swedish and Vata exemplify languages with resumptive pronouns of this type. He proposed five main syntactic diagnostics to distinguish SARs from SIRs: syntactic islands, weak crossover, across-the-board (ATB), parasitic gaps (PG) and reconstruction. The most robust diagnostics are syntactic islands and weak crossover (WCO). SIRs but not SARs are island sensitive, subject to WCO, reconstruction licensed, allow ATB extraction and license PG. SARs are anaphorically bound, but SIRs are syntactically gap-like (i.e. absent in f-structure) and hence not anaphorically bound. This next section explores the status of resumptive pronouns in Wolof.

6 Wolof resumptive pronouns

Building on previous works in LFG (Asudeh 2011, Camilleri and Sadler 2011a), I will addresses two fundamental questions regarding resumptive elements in Wolof relative clauses. The first question is whether these elements are true resumptive pronouns or not, according to the diagnostics discussed in section (5). The second question is whether they are syntactically active (SAR) or syntactically inactive (SIR) pronouns. My investigation will closely mirror the methods used by Camilleri and Sadler (2011a,b) for Maltese, as that language shows striking similarities to Wolof in some extent.

To answer the first question, I provide data for the comparison between Wolof and English. The patterns in (26) are strikingly similar to the English examples in (25), suggesting that these elements are indeed true resumptives and not intrusive pronouns. (26a) shows that a resumptive pronoun may be bound by a quantifier (e.g. *bépp* 'every') resisting an e-type interpretation. (26b) shows that the pronoun in question supports a list answer (and so is a resumptive), and (26c) demonstrates that it supports a functional answer to a *wh*-question. Together, these examples seem to provide evidence that Wolof has true resumptives rather than intrusive pronouns in these contexts.

(26) a. *bépp téeré b-u Samba fàtte ni jàng na ko/*FRP démb*
every book CL-u Samba forget that read 3SG it/*__ yesterday
"every book that Samba forgot that he has read it yesterday"

b. *Ban jàngalekat nga foog ni su ko/*FRP Awa jëlee*
which teacher 2SG think that if 3SG.O/*__ Awa employ-PFV
ñépp di-na-ñu bég?
everyone IPFV-+F-3pl be.happy
"Which teacher do you think that if Mary succeeds in employing (him), everyone will be happy?"
"Omar, Faatu wala Birane" (= Omar, Faatu or Birane)

c. *K-an mu-a-y jigéen j-i bépp góor xam*
CL-an 3SG-COP-IPFV woman CL-i every man know
yaay-(am)*
mother-3SG.POSS
"Which is the woman$_i$ whom every man knows her$_i$ mother?"
– "Awa" (=Awa)
– "jabaram" (=his wife)
– *Samba, Awa ak Omar Faatu (= Samba, Awa and Omar, Faatu)

Now to answer the second question, we will use the five diagnostics as proposed by Asudeh (2011), starting with syntactic islands. Example (27) illustrates the Complex Noun Phrase Constraint (CNPC), with a (second) relative dependency into a CNP created by relativisation. Although the relativised position is one which is normally accessible to the gap strategy, the resumptive is obligatory here as a gap would cause a syntactic constraint violation.

(27) *kër g-i ma xam góor g-i *(**ko**) tabax*
house CL-i 1SG know man CL-DFP *(it) build
"the house that I know the man who built it"

Relativization out of adjuncts (e.g. the bracketed constituent in (28a)) leaves a gap. Crucially, as (28b-28c) show, it appears that, for Wolof, both gaps and resumptive pronouns obey the adjunct island contraints.[8]

(28) Adjunct island

a. *Samba xam na Awa [laata góor g-i tabax kër g-i]*
Samba know 3SG Awa before man CL-DFP build house CL-DFP
"Samba knew Awa before the man built the house."

b. **góor g-i Samba xam Awa [laata (mu) tabax kër g-i]*
man CL-i Samba know Awa before 3SG build house CL-DFP
"the man that Samba knew Awa before he built the house"

c. **kër g-i Samba xam Awa [laata góor g-i tabax (ko)]*
house CL-i Samba know Awa before man CL-DFP build (it)
"the house that Samba knew Awa before the man built (it)"

Example (29) illustrates a *wh*-island where a *wh*-expression, *k-an* 'who', has been clefted into an embedded CP. As can be seen, with the RP, the construction is not subject to the *wh*-island constraint. However, without the resumptive pronoun, the long-distance dependency would be subject to island constraints.

(29) *wh*-Island

a. *Samba xam na [k-an mu-a tabax kër g-i]*
Samba know 3SG CL-an 3SG-SFOC build house CL-DFP
"Samba knows who built the house."

b. *kër g-i Samba xam [k-an mu-a *(ko) tabax]*
house CL-REL Samba know CL-an 3SG-SFOC *(it) build
"the house that Samba knows who it was that built it"

The examples about complex noun phrase constraints (CNPC) and *wh*-island constructions seem to provide evidence that TRPs (unlike gaps) are felicitous within

[8]Palauan (Georgopoulos 1991) shows similary to Wolof in that extraction from an adjunct is ungrammatical, even with a resumptive pronoun.

these kinds of syntactic islands. However, both TRP and gap seem to be subject to the Adjunct Island Constraint, which appears to be too strong in Wolof.

Besides syntactic islands, weak crossover is the most robust SAR/SIR diagnostic. Let us consider (30), which is an instance of relativisation on the OBJ. The dependency between the antecedent *góor* (or the TOPIC) and the TRP 'crosses over' the possessive in *jabar-am*, but the sentence is perfectly well-formed. By contrast, and although both gap and TRP are generally available for relativisation on the OBJ, employing a version of (30) with a gap rather than a TRP is ungrammatical.

(30) *góor$_i$ g-i ma xam ni jabar-*(am$_i$) bàyyi na ko*
man CL-REL 1SG know that wife-3SG.POSS leave 3SG 3SG.O
"the man that I know that his wife left him"

The Wolof data seem to provide support that the RPs found in that language are SARs (i.e. they should be treated as anaphoric pronouns at f-structure). On the basis of this evidence, the basic analysis of Wolof relative clauses given above can be extended by substituting (15) with (31). The only change is the addition of an anaphoric dependency (↑ TOPIC) = ((↑ RRPPATHσ) ANTECEDENT) to allow for the use of a resumptive), and adding the resumptive path definition in (32).

(31) CP → RelP IP

(31) CP	→	RelP	IP
		(↑ TOPIC)=↓ { (↑ TOPIC)=(↑ RTOPICPATH) | (↑ TOPIC) = ((↑ RRPPATHσ) ANTECEDENT) } (↑ RELPRO PRON-TYPE) $=_c$ rel @REL-FEAT	↑=↓

(32) RRPPATH ≡ { ARGF }* GF
GF ≡ { SUBJ, OBJ, OBJ-APPL, POSS }
ARGF ≡ { SUBJ, OBJ, OBL, XCOMP, COMP }

As in Maltese (Camilleri and Sadler 2011a), the general impossibility of using a resumptive in the highest subject position may be captured by an anti-locality condition, as proposed in Asudeh (2004).

(33) Anti-Locality Condition:
(↑ σ ANTECEDENT) $\neq$ ((↑ SUBJ) TOPIC)σ

The SAR/SIR diagnostic based on parasitic gaps is somewhat difficult to verify, as Wolof does not seem to have parasitic gap-like constructions (Torrence 2013). An example of parasitic gaps is illustrated for English in (34) where the second "gap" (marked with a p-subscript) appears to be dependent on the first "gap". The second gap is "parasitic" in the sense that it can appear only by virtue of the appearance of the first gap. As can be seen in (36), there are two possible scenarios that would give a grammatical sentence for these kinds of constructions: (i) a gap licenses an RP or an RP licences another RP. In any case, a second gap is impossible (as a TRP is required instead). Thus, a gap cannot license another gap. Likewise, a resumptive pronoun cannot license a gap. If the embedded clause in (36) were a non-subject cleft instead, the first gap would be prohibited. We conclude, then, that the parasitic gap diagnostic is not exactly applicable in Wolof.

(34) Awa saw the car you bought __ in order to fix $\_\_{p}$ up.

(35) *Awa gis na nga jënd woto b-i ngir defar ko*
Awa see 3SG 2SG buy car CL-DFP in.order.to fix.up 3SG.O
"Awa saw the car you bought in order to fix it up."

(36) *woto b-i Awa gis nga jënd (ko) ngir defar *(ko)*
car CL-REL Awa see 2SG buy 3SG.O in.order.to fix.up 3SG.O
"the car Awa saw you bought in order to fix it up"

Furthermore, let us consider the distribution of gaps and resumptive pronouns in across-the-board (ATB) constructions. According to this diagnostic, SARs should not mix with gaps in ATB constructions. Example (37) involves an instance of ATB constructions in form of coordination of IPs (i.e. the TOPIC is outside the coordination). This example shows coordination under the relative pronoun with a gap in the first conjunct and an obligatory RP in the second conjunct. A gap would not be possible in the second conjunct. Conversely, a RP is not permitted in the first conjunct. The ATB data suggest that gap and TRP not only mix up, but that configuration is the only possible one. Following the approach developed in Asudeh (2011), the ATB data might suggest that Wolof also has SIRs (i.e. functionally controlled RPs or audible gaps). The result of this diagnostic would then be inconsistent with the results of the previous diagnostics. Camilleri and Sadler (2011b) faced a similar issue for Maltese and could not drawn any conclusion about the interaction of TRPs with ATB phenomena in relation to the SIR/SAR status of Maltese resumptives. Here too, we will leave this analytic issue for further work.

(37) *téeré b-i Awa jënd te Samba jàng *(ko)*
book CL-REL Awa buy and Samba read 3SG.O
"the book that Awa bought and Samba read"

Our final SAR/SIR diagnostic concerns the distribution of gaps and TRP in reconstruction contexts. As Camilleri and Sadler (2011b) pointed out, in such contexts the fronted material shows a range of (interpretive) behaviours appropriate for its *in situ* position or function. In the standard LFG's approach to LDDs (with gaps), the unbounded dependency between the filler and the gap is captured via functional control. This allows a prediction of the "reconstruction" properties by associating the filler with both the discourse function (e.g. TOPIC) and the within-clause function (e.g. SUBJ, OBJ, OBJ-TH, ..). The Wolof examples in (38) show instances of *scope reconstruction*: a gap is under the scope of a quantifier. As (38a) shows, the TRP is required in long paths if the clause is not a non-subject cleft (the gap being prohibited). If otherwise, the filler is extracted from a non-subject cleft, as in (38b), the TRP is not allowed (only gap is permitted).

(38) a. *kër g-u ñu wax ni xale b-u nekk bëgg na *(ko)*
house CL-REL 3PL say that child CL-REL exist love 3SG (3SG.O)
"a house which they said that every child loves"

b. *kër g-u ñu wax ni la xale b-u nekk bëgg*
house CL-REL 3PL say that NSFOC.3 child CL-REL exist love
*(*ko)*
(3SG.O)
"a house which they said that every child loves"

Examples (39) illustrate binding reconstruction (e.g. of reflexive pronouns). The patterns are similar to what we found for scope reconstruction with respect to the distribution of gaps and RPs and the impact of the clause type.

(39) a. *nataalu-u yaay-am b-u ñu wax ni [doom j-u*
picture-of mother-3SG.POSS CL-REL 3PL say that child CL-u
*nekk] bëgg na *(ko)*
exist love +F-3SG (3SG.O)
"a picture of his mother which they said that every child loves"

b. *nataalu-u yaay-am b-u ñu wax ni la*
picture-of mother-3SG.POSS CL-REL 3PL say that NSFOC.3
*doom j-u nekk bëgg (*ko)*
woman CL-u exist love (3SG.O)
"a picture of his mother which they said that every child loves"

According to Asudeh (2011), reconstruction would provide an evidence for SIR status. This is because reconstruction itself is a phenomenon that distinguishes gaps from pronouns. These Wolof data seem to suggest it is not possible to reconstruct into a resumptive in Wolof when extracting from the immediate position or from a long path where the in situ position was located in a non-subject cleft construction. Otherwise, it seems to be possible to reconstruct into a resumptive in Wolof. If reconstruction is indicative of SIR status, then this data set show (in part) inconsistencies with the results of other diagnostics, which support SAR status for Wolof resumptives. However, as Camilleri and Sadler (2011b) indicated, the status of the reconstruction diagnostic itself may be open to question.

7 Conclusion

This paper has provided a description of the syntax of relative clauses, which constitute a major source of linguistically interesting constructions in Wolof. I have provided an analysis of these structures in LFG, building on previous LFG work on relatives. The discussion raised a number of issues on how to account for the status of Wolof resumptive pronouns at the functional level. On the basis of two major diagnostics (concerning islandhood and weak crossover) developed in Asudeh (2011), I have argued that the resumptive pronouns found in Wolof are syntactically active pronouns. The Wolof data show striking similarities to the observations made in other languages such as Maltese (Camilleri and Sadler 2011a,b). In Wolof too, islandhood and weak crossover seem to be quite robust, while the remaining diagnostics (ATB extraction, parasitic gaps, reconstruction) seem to be less robust because it is less clear that the relevant property is entirely syntactic.

References

Asudeh, Ash. 2004. *Resumption as resource management*. Ph.D. thesis, Stanford University.

Asudeh, Ash. 2011. Toward a Unified Theory of Resumption. In A. Rouveret, ed., *Resumptive pronouns at the interfaces*. John Benjamins Publishing.

Austin, Peter K. 2001. Lexical functional grammar. *International Encyclopedia of the Social and Behavioural Sciences* pages 8748–8754.

Borer, Hagit. 1984. Restrictive relatives in modern hebrew. *Natural Language & Linguistic Theory* 2(2):219–260.

Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Oxford: Blackwell.

Bresnan, Joan and Sam A. Mchombo. 1987. Topic, pronoun, and agreement in Chicheŵa. *Language* 63(4):741–782.

Butt, Miriam, Tracy Holloway King, María-Eugenia Niño, and Frédérique Segond. 1999. *A Grammar Writer's Cookbook*. Stanford, CA: CSLI.

Camilleri, Maris and Louisa Sadler. 2011a. An lfg approach to non-restrictive relative clauses in maltese. *Essex Research Reports in Linguistics* 60(6).

Camilleri, Maris and Louisa Sadler. 2011b. Restrictive relative clauses in Maltese. In M. Butt and T. H. King, eds., *The Proceedings of the LFG '11 Conference*. CLSI Publications.

Chatsiou, Kakia. 2010. *A lexical functional grammar approach to Modern Greek relative clauses*. Ph.D. thesis, University of Essex.

Dalrymple, Mary. 2001. *Lexical-Functional Grammar (Syntax and Semantics, Volume 34) (Syntax and Semantics)*. Academic Press.

Dione, Cheikh Bamba. 2020. Clause structure, pro-drop and control in Wolof: an LFG/XLE perspective. *Nordic journal of African studies* 28(3).

Dione, Cheikh M. Bamba. 2012. An LFG Approach to Wolof Cleft Constructions. In M. Butt and T. H. King, eds., *The Proceedings of the LFG '12 Conference*. Stanford, CA: CSLI Publications.

Dione, Cheikh M. Bamba. 2013. Valency Change and Complex Predicates in Wolof: An LFG Account. In M. Butt and T. H. King, eds., *The Proceedings of the LFG '13 Conference*. Stanford, CA: CSLI Publications.

Dione, Cheikh M. Bamba. 2014a. *Formal and Computational Aspects of Wolof Morphosyntax in Lexical Functional Grammar*. Ph.D. thesis, University of Bergen, Norway.

Dione, Cheikh M Bamba. 2014b. LFG parse disambiguation for Wolof. *Journal of Language Modelling* 2(1):105–165.

Falk, Yehuda. 2001. *Lexical Functional Grammar: an Introduction to Parallel Constraint-Based Syntax*. Stanford, CA: CSLI Publications.

Falk, Yehuda N. 2002. Resumptive Pronouns in LFG. In M. Butt and T. H. King, eds., *The Proceedings of the LFG '02 Conference*. National Technical University of Athens.

Georgopoulos, Carol. 1991. *Syntactic variables: Resumptive pronouns and A′ binding in Palauan*, vol. 24. Springer Science & Business Media.

Harris, Christen. 2015. *Applicative Structure in Wolof*. Ph.D. thesis, The University of Western Ontario.

Keenan, Edward L and Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic inquiry* 8(1):63–99.

McCloskey, James. 1990. Resumptive pronouns, ā-binding, and levels of representation in irish. In *The syntax of the modern Celtic languages*, pages 199–248. Brill.

McLaughlin, Fiona. 2010. Noun classification in Wolof: When affixes are not renewed. *Studies in African Linguistics* 26(1).

Robert, Stéphane. 1991. Approche énonciative du système verbal: le cas du wolof. *Editions du CNRS* .

Ross, J.R. 1967. Constraints on variables in syntax. *Doctoral dissertation, Massachusetts Institute of Technology* .

Sapir, David J. 1971. West Atlantic: an inventory of the languages, their noun class systems and consonant alternation. *Current Trends in Linguistics* 7(1):43–112.

Shlonsky, Ur. 1992. Resumptive pronouns as a last resort. *Linguistic Inquiry 1992* 23(3):443–468.

Torrence, Harold. 2013. *The clause structure of Wolof: insights into the left periphery*, vol. 198. John Benjamins Publishing.

Torrence, William Harold. 2005. *On the Distribution of Complementizers in Wolof*. Ph.D. thesis, University of California, Los Angeles, CA.

Voisin-Nouguier, Sylvie. 2002. *Relations entre fonctions syntaxiques et fonctions sémantiques en wolof*. Ph.D. thesis, Université Lumière, Lyon 2. France.

Zribi-Hertz, Anne and Lamine Diagne. 2002. Clitic placement after syntax: evidence from Wolof person and locative markers. *Natural Language & Linguistic Theory* 20(4):823–884.

Beyond c-structure and f-structure: On the argument-adjunct distinction in O'dam

Michael Everdell
The University of Texas at Austin

Paul B. Melchin
Carleton University

Ash Asudeh
University of Rochester/Carleton University

Daniel Siddiqi
Carleton University

Proceedings of the LFG'21 Conference

On-Line

Miriam Butt, Jamie Y. Findlay, Ida Toivonen (Editors)

2021

CSLI Publications

pages 125–145

http://csli-publications.stanford.edu/LFG/2021

Keywords: complement clauses, control constructions, relative clauses, LRFG, argumenthood, quantifiers, O'dam

Everdell, Michael, Melchin, Paul B., Asudeh, Ash, & Siddiqi, Daniel. 2021. Beyond c-structure and f-structure: On the argument-adjunct distinction in O'dam. In Butt, Miriam, Findlay, Jamie Y., & Toivonen, Ida (Eds.), *Proceedings of the LFG'21 Conference, On-Line*, 125–145. Stanford, CA: CSLI Publications.

Abstract

In this paper we examine the argumenthood properties of Controlled Complement Clauses and Non-Complement Subordinate Clauses in O'dam. We show that in O'dam only controlled COMPs are arguments, while other putative complement clauses are adjunct relative clauses that elaborate on a pronominal OBJ incorporated in the matrix verb. We use the L_RFG framework to capture both the argumenthood properties of the two types of clauses in O'dam as well as the patterns of object marking on the matrix verb by taking advantage of mismatches between c-structure (phrase structure and f-descriptions) and v-structure (the vocabulary items realizing this structure).

1 Introduction

In this paper we discuss the distinction between arguments and adjuncts in the Uto-Aztecan language O'dam. We focus on two types of subordinate clauses that previous literature has grouped together as complement clauses (Willett 1991, García Salido 2014). We call these two subordinate clause types Controlled Clausal Complements (CCC), shown in (1), and Non-Complement Subordinate Clauses (NCS), shown in (2).[1][2]

(1) Tɨmu-ñi-ch [na=ñi-ch mɨɨ]$_{CCC}$
finish-1SG.SBJ-PFV SUB=1SG.SBJ-PFV run.SG.PFV
'I finished running.' (García Salido 2014: 283)

(2) Sap jup Ø-kaich-'am [na=Ø ba-tu-m-maki-a'
REP.UI IT 3SG.PO-say-3PL.SBJ SUB=3SG.SBJ CMP-DUR-MID-give-IRR
gu tumiñ]$_{NCS}$
DET money
'According to them, they said that money will be received.' (García Salido 2014: 281)

We will argue that only CCCs have the grammatical function COMP, while NCSs have the grammatical function ADJ. We will additionally argue that NCSs, as in (2), are headless relative clauses and that the object marking on the verb is an incorporated pronoun that takes the NCS as its referent.

[†]This project was in part funded by NSF-DDRIG BCS-1946625. The data here comes from our consultants Eli Soto Gurrola, Yamileth Gurrola, Wendy Gurrola, and Mauro Aguilar, who continue to help us understand the O'dam language. Thank you for the invaluable feedback from Gabriela García Salido, Luke Adamson, Kristin Denlinger, John Beavers, Stephen Wechsler, and the audiences of LFG21, WCCFL, NaCC, and the Syntax and Semantics Research Seminar at UT.

[1]García Salido (2014) terms CCCs Type 3 complement clauses and NCSs Type 1 complement clauses.

[2]Most of the glossing we use is taken from Leipzig abbreviations, here we show the abbreviations which do not have their standard Leipzig values: ∼ = reduplication ADVR = Adverbializer; EST = Stative; PO = primary object REP.UI = Reportative Unknown Information

This paper proceeds as follows: In §2 we overview basic background on the O'dam speaking community. Then in §3 we discuss previous work on the argument-adjunct distinction in O'dam and the preverbal quantifier test in §3.2. We then discuss the c-structural shape of O'dam subordinate clauses in §4 and the features of CCCs that distinguish them from other subordinate clauses in §4.1. In §4.2s we show that NCSs are distinct from CCCs in both their coreference on the matrix verb and their argumenthood properties. We propose that NCSs are not complements of their matrix verb, but that the verb selects for an OBJ with a referent associated with the NCS, which we back up with c-structural (in §4.2.1) and interpretational (in §4.2.2) properties of NCSs. In §5 we show that the LFG account leads to mismatches between argumenthood diagnostics, and thus must rely on stipulations of argumenthood. Finally, in §6 we show how the framework of Lexical-Realizational Functional Grammar ($\mathrm{L_R}$FG) accounts for the distinction between clausal complements, in terms of object marking, while maintaining a principled definition of argumenthood.

2 The O'dam

O'dam (glottocode: sout2976) is a Uto-Aztecan language in the Tepiman subgroup, shown in Figure 1. O'dam is spoken primarily in the southern region of Durango and Nayarit, Mexico, in the part of the Sierra Madre known as the Gran Nayar. In Figure 2 we see Southern Tepehuan towns, with the O'dam speaking communities being those loosely centered around Santa María de Ocotán.

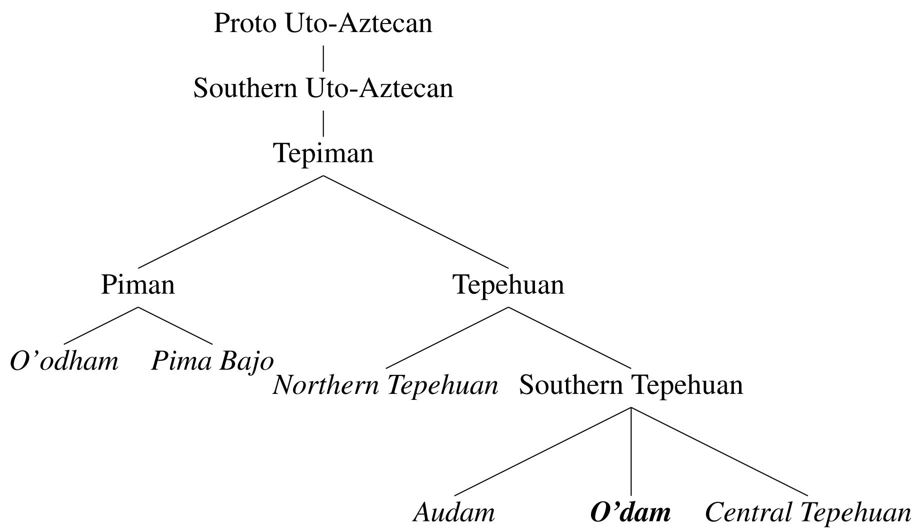


Figure 1: Uto-Aztecan family tree

Official censuses count O'dam as part of Southern Tepehuan, which includes O'dam's sister languages Audam (Southwestern Tepehuan) and Central Tepehuan. Southern Tepehuan has 36,543 speakers (INEGI 2015), of which O'dam is the most widely spoken and the best studied of the three varieties. Currently children are learning O'dam as their L1, although increasing economic pressure is pushing O'dam speakers into communities where Spanish is more dominant; see García Salido and Everdell (2020). Geographically, the Southern Tepehuan languages are surrounded by other Southern Uto-Aztecan languages: Cora, Huichol and Mexicanero, although O'dam generally live in towns consisting of just O'dam or O'dam and mestizos.[3] The speakers we work with are fluent in Spanish and

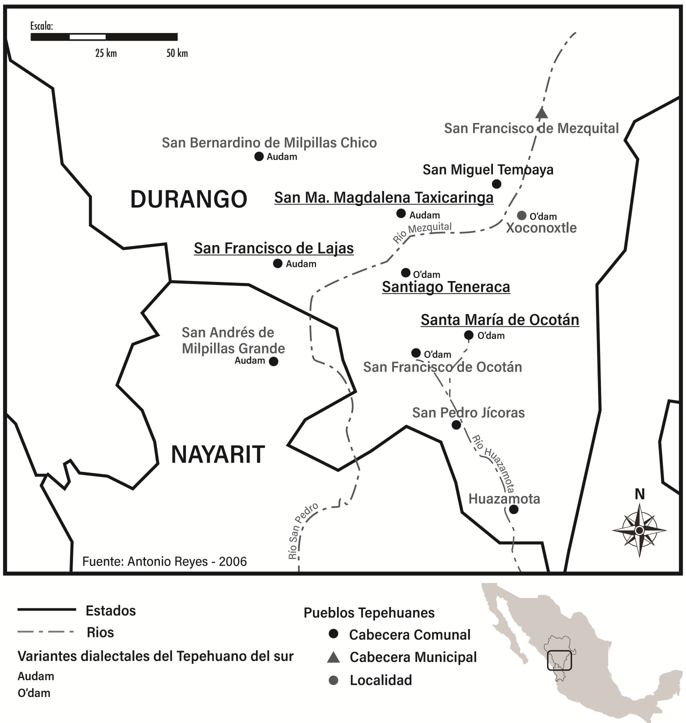


Figure 2: Map of Southern Tepehuan towns in Mexico (García Salido and Everdell 2020: 90)

O'dam and split their time between Durango City and their respective communities of Juktɨr (Santa María de Ocotán), Koba'ram (La Candelaria) and Suusbhaikam (Los Charcos).

3 Argumenthood in O'dam

3.1 Diagnosing argumenthood

The distinction between arguments and adjuncts in O'dam is not a clear one. Nominals lack case marking, the only element required for a clause is a verb, and verbal

[3] *Mestizo* is the majority ethnic group in Mexico, consisting of people who have mixed European and indigenous heritage. O'dam generally do not recognize *mestizos* as Indigenous.

dependents (arguments and adjuncts) can occur in any order following the verb, although typically no more than two XPs of any grammatical function appear in a given clause (Willett 1991, García Salido 2014, Everdell in progress).

Previous work on O'dam relies on two diagnostics for argumenthood (Willett 1991, Willett and Willett 2013, García Salido 2014). Subjects and primary objects are diagnosed based on their coreference with verbal affixes, shown in Tables 1 and 2. In (3) we see that the the 1SG subject suffix *-(i)ñ* and the 3PL primary object prefix *ja-* coreference the experiencer and stimulus of the seeing event. O'dam only permits a single object to be coreferenced on the verb, such non-coreferenced objects of ditransitives are called secondary objects. Because secondary objects lack any verbal coreference or obligatory exponent in the clause, previous work on O'dam has generally assumed their existence through entailment (e.g. 'give' entails a theme and recipient).[4]

To illustrate the difference between objects, the applicative in (3) licenses a beneficiary when combined with the verb *nɨɨya'* 'see', as in (3) (Everdell and García Salido 2021).[5] The primary object OBJ) is the 3SG stimulus, while we see in the gloss that the beneficiary is 1SG, making it the secondary object (OBJ$_\theta$). However, the beneficiary lacks an exponent in the clause, it is optional, and in another discourse context could be any person-number combination. secondary objects optionally receive XP exponents along with primary objects, as in (4) where *gu tatoxkolh* '(the) pigs' is the primary object (OBJ) and *gu koi'* '(the) food' is the secondary object (OBJ$_\theta$). secondary objects licensed by applicatives also often receive primary object status, as in (5a) where the *-dha* applicative combines with *ɨxcho'* 'hide' to license a person the patient is hidden from, compare (5b).

	Subject	Primary Object
1SG	*-(i)ñ*	*jiñ-*
2SG	*-(a)p*	*(ju)m-*
3SG	*-Ø*	*Ø-*
1PL	*-(i)ch*	*(ji)ch-*
2PL	*-(a)pim*	*jam-*
3PL	*-(a)m*	*ja-*

Table 1: Non-topic subject and primary object markers

	SG	PL
1	*(ji)ñ-*	*(ji)ch-*
2/3	*(ju)m-*	*(ju)m-*

Table 2: Middle primary object markers

[4]Recent work exploring these secondary objects has found them to pattern with primary objects and subjects in a number of ways, see Everdell (in progress).

[5]The tilde (∼) indicates reduplication, according to the Leipzig Glossing Rules.

(3) Añ gu=x bu∼pui-ch-ɨk ji na=**ñ**
1SG.SBJ DET=COP IT∼eye-CAUS-PNCT FOC SUB=1SG.SBJ
bha=**ja**-nɨ'ñ-dha' ma'n
DIR =3PL.PO-see-APPL one
'I only was looking at the ugly ones **for me**.' (García Salido 2014: 80)

(4) Añ tu-ja-maa gu ta∼toxkolh gu koi'
1SG.SBJ DUR-3PL.PO-give.PFV DET PL∼pig DET food
'As for me, I gave food to the pigs.' (García Salido 2014: 49)

(5) a. Ja-ɨxchoi-dha-'-iñ [gu biiñ]$_{\text{OBJ}_\theta}$ [gu=ñ
3PL.PO-hide-APPL-IRR-1SG.SBJ DET mezcal DET=1SG.POSS
jikkulh]$_{\text{OBJ}}$ na=pai'dhuk koxi-a'
uncle.PL SUB=when sleep-IRR
'I'm going to hide the mezcal from my uncle when he goes to sleep' (adapted from (Willett and Willett 2013: 73))

b. Ka-xi-Ø-ɨxcho-'-ap dhi kiis na=m
PERF-IMP-3SG.PO-hide-IRR-2SG.SBJ DEM cheese SUB=3PL.SBJ
cham jich-jugii'ñ-dha-' gu ja'tkam
NEG 1PL.PO-finish-APPL-IRR DET people
'Hide this cheese so the people don't finish ours! [Esconde el queso para que no se lo acabe la gente]' (Willett and Willett 2013: 73)

Even though standard argumenthood tests fail for secondary objects, previous work has assumed that they are arguments. This assumption arises from the fact that they are entailed by the verb (e.g., Everdell and García Salido 2021). This characteristic has been shown to be a (somewhat mixed) indicator of argumenthood (Cappelen and Lepore 2005, Needham and Toivonen 2011, Barbu 2015, Barbu and Toivonen 2016a,b, Moura and Miliorini 2018). The factors determining primary and secondary objecthood are currently not well understood although in texts the primary object is most often the one with the highest animacy and number (García Salido 2014: 46ff). Everdell (2021) however finds that primary and secondary objects are symmetrical with respect to argumenthood tests other than verbal coreference, for example the preverbal quantifier test we use here. We treat primary objects as OBJ and secondary objects as OBJ_θ. In §3.2 we return to the properties of preverbal quantifiers that make them a useful argumenthood test, before turning to CCCs and NCSs.

3.2 Preverbal quantifiers

Quantifiers in O'dam are a distributionally defined class of elements that immediately precede determiners in DPs, what we call the constituent position, or precede the verb, what we call the preverbal position. Although many O'dam quantifiers have quantifier semantics we have not checked whether all of them do and at issue here is their f-structural properties and ability to associate with a discontinuous XP, not their s-structures or compositional meanings.

In the constituent position quantifiers quantify whatever XP they are a constituent with, as in (6). In this position the grammatical function of the XP in the larger clause is not relevant. In the preverbal position quantifiers act as an argumenthood diagnostic, they quantify arguments of the associated verb and not adjuncts. We see this in (7) where *ma'n* 'one/a'[6] can quantify the object *gu bhan* 'the/a coyote' but not the locative *mu pue'mlo* 'the town down there', which are systematically treated as adjuncts in O'dam, see also Everdell (2021, in progress) for further evidence of preverbal quantifiers as an argumenthood test.

(6) a. Ø-tɨɨ-ñi-ch [**ma'n** gu bhan]$_{DP}$ mu
3SG.PO-see.PFV-1SG.SBJ-PFV one DET coyote DIST.LOWER
pue'mlo
town
'I saw one/a coyote in that town'

b. Ø-tɨɨ-ñi-ch gu bhan [**ma'n** mu
3SG.PO-see.PFV-1SG.SBJ-PFV DET coyote one DIST.LOWER
pue'mlo]$_{Loc}$
town
'I saw the/a coyote in a town (down there)'

(7) **Ma'n** Ø-tɨɨ-ñi-ch [gu bhan]$_{Argument}$ [mu
one 3SG.PO-see.PFV-1SG.SBJ-PFV DET coyote DIST.LOWER
pue'mlo]$_{Adjunct}$
town
'I saw one/a coyote in that town'
*I saw the/a coyote in one/a town (down there)

In (8) we see the quantifier *bɨx* 'all' preceding the verb *nɨiya'* 'see' with a 1SG experiencer subject and a 3PL stimulus object. The subject is not quantifiable here because the 1SG subject is not compatible with the plural interpretation forced by *bɨx*. In (8a) we see that *bɨx* can quantify the OBJ of *nɨiya'* 'their teachers', but not the embedded possessor 'my friends' in (8b). Thus, we see that preverbal quantifiers in O'dam show DP island effects. We also see in (8c) that properties of the verb itself can be quantified in addition to more standard parts of the argument set of a verb's f-structure (i.e. the SUBJ and OBJ).

(8) **Bɨx** ja-nɨi'-iñ [gu ja-mamtuxi'ñ-dham [gu=ñ
all 3PL.PO-see-1SG.SBJ DET 3PL.POSS-teach-NMLZ DET=1SG.POSS
a'∼mi']$_{DPpossessor}$]$_{DPpossessum}$
PL∼friend

[6] Definiteness in O'dam is pragmatic (Willett 1991: 206-7) so *ma'n* can be interpreted as a true numeral or as an indefinite marker.

a. 'I see **all of the teachers** of my friends'

b. *I see the teachers of **all of my friends**

c. 'I see all of the teachers of my friends (e.g. if the teachers are trying to hide)'

The correlation between argumenthood and preverbal quantification suggests that quantification is mediated by f-structure, where grammatical functions and argumenthood are encoded, rather than at c-structure (see Al Khalaf 2019 and referenced therein). The functional equation for *bix* 'all' is given in (9). Here, AF is a variable over the argumental grammatical functions. The '*' notation indicates that the feature can be assigned to the current f-structure, including OBJ_θ in (26), or to any that can be reached via a path of AF functions, which will be discussed below. The f-structure feature QUANT, and values like ALL, are a simplifying substitute for an account in Glue Semantics (see, e.g., Dalrymple et al. 2019: chap. 8), which would involve the relevant portion of the path specification.

(9) $(\uparrow \text{AF}^* \text{ QUANT}) = \text{ALL}$

We have shown that quantifiers in the preverbal position quantify members of a verb's AF list and do not quantify those of the ADJ set. We have additionally shown that the verb itself is treated as a member of the set of Argument Functions by preverbal quantifiers. We now use this information to discuss the argumenthood properties of CCCs and NCSs in §4.

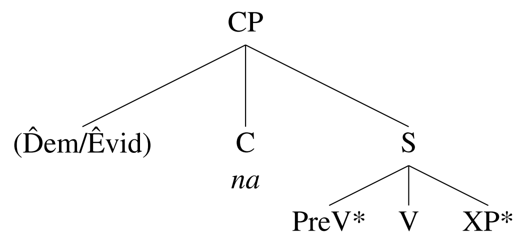
4 Subordinate clauses

The basic structure for all subordinate clauses in O'dam, complement or otherwise, involves projecting a CP over an S, which is a basic non-subordinate clause (Everdell and Melchin 2021, Everdell in progress).[7]

Within a basic clause, the V is the verb complex and can be understood as comprising the verbal word (Everdell in progress). The PreV is where preverbal quantifiers occur and consists of various scopally ordered non-projecting functional particles that roughly align with the clausal spine (Ramchand and Svenonius 2014) The only projecting heads attested in the PreV are topic XPs. The XP position following the V consists of all non-topic phrasal dependents of the verb regardless of grammatical function, see Everdell (in progress) for a fuller discussion of O'dam constituency.

[7]There are various subordinators (see García Salido 2014), however the general subordinator *na* is the only relevant one for our purposes.

(10)



While all subordinate clauses in O’dam share the same basic c-structural form, previous work grouped CCCs and NCSs as complement clauses because they are associated with special marking on the verb, which we discuss in §4.1 and §4.2. We will show that CCCs are true complement clauses, while NCSs are headless relative clauses with the ADJ grammatical function.

4.1 Controlled clausal complements

CCCs, shown in (11), are finite and fully saturated for their arguments, what Stiebels (2007) calls “inherent control.” Previous work primarily diagnoses CCCs by the obligatory coreference marking and interpretation of the controller and controllee (Willett 1991, García Salido 2014). This is shown in (11), where the subject of *poderu’* ‘be able to’ controls the subject of *manteneru’* ‘support’ and both must be 3PL and be interpreted as consisting of the same set of individuals. While both subject and object controllers are attested, shown in Table 3,[8] we have only identified controlled subjects in O’dam. Nonetheless, our analysis here would apply the same to a CCC with a controlled object because the AF path would be the same; this is currently an unconfirmed prediction. We only find exhaustive control in the language; to our knowledge partial control constructions à la Landau (2000) do not exist.

(11) Na=**m**$_i$-gu’ ba-poder [na=**m**$_{i/*j}$ jich-mantener-ka’]$_{CCC}$
SUB=3PL.SBJ-ADVR CMP-can SUB=3PL.SBJ 1PL.PO-support-EST
ja’p sap jum-aa’
DIR REP.UI MID-think.PFV
‘Because they could support us, he thought so.’ (lit. Because they$_i$ could they$_{i/*j}$ support us, he thought so) (adapted from García Salido 2014: 283)

Verb	Meaning	Controller	Verb	Meaning	Controller
poderu’	‘be able to’	Subject	*chia’*	‘send/causative’	Object
tɨɨmo’	‘finish’	Subject			

Table 3: Attested control verbs in O’dam

We analyze control verbs as taking the CCC as a COMP argument. They also take another core argument, the controller, that must be coreferenced with the sub-

[8]Citation forms for verbs in O’dam are always given with the *-(a)’* irrealis suffix.

ject of the embedded clause; this is the direct object if the matrix verb has a thematic subject present, and subject otherwise. The control relationship is specified as in (12), for instances where the controller is the object, and (13) for subject controllers, adopted from Asudeh (2005).[9]

(12) $(\uparrow\mathrm{OBJ})_{\sigma} = ((\uparrow\mathrm{COMP\ SUBJ})_{\sigma}\ \mathrm{ANTECEDENT})$

(13) $(\uparrow\mathrm{SUBJ})_{\sigma} = ((\uparrow\mathrm{COMP\ SUBJ})_{\sigma}\ \mathrm{ANTECEDENT})$

When a quantifier sits in the preverbal position of a control verb, we find that it can quantify the arguments of both the control verb and the controlled verb. This is shown in (14) using the the analytical causative *chia'* 'send' and the quantifier *dɨlh* 'only' in the matrix preverbal position. In (14a) and (14b) we see that *dɨlh* can quantify the SUBJ and OBJ of *chia'*, where the OBJ is also the controller of the CCC subject. In (14c) we see that *dɨlh* can quantify the CCC clause as a whole, as we saw was possible for preverbal quantifiers in simple clauses in §3.2.

(14) Dɨlh jam-chia-mɨ-t na=pim bopooy-a' jix=io'm
only 2PL.PO-send-3PL.SBJ-PFV SUB=2PL.SBJ run.PL-IRR COP=very

a. '**Only they** told you.PL to run faster'
b. 'They told **only you all** (as opposed to anyone else) to run faster'
c. 'They told you all **to only run** faster (as opposed to do anything else faster)'

We also see in (15) that a non-controlled object of a CCC is quantifiable from the preverbal position of the control verb. Because quantifiers must be compatible with the elements they quantify (i.e. a quantifier with a plural interpretation cannot quantify a singular DP) the sentence in (15) would be ungrammatical if the non-controlled OBJ of the CCC was not available, because all other participants in the control construction are singular.

(15) Gok jiñ-chia-pi-ch na=ñ jup duñi-a' gu tacos
two 1SG.PO-send-2SG.SBJ-PFV SUB=1SG.SBJ IT do-IRR DET tacos
'You wanted me to make two tacos'

Since the arguments of the CCC are arguments of a COMP function, they fall within the scope of preverbal quantifiers as specified in (9) in §3.2. The f-structure for the object control construction in (15) is given in Figure 3, while the f-structure for the subject control construction seen above in (1) is given in Figure 4. Note that in this analysis verbs with an object controller are ditransitive. They pattern with other ditransitives in that they only show agreement with one of the objects/complements. See §5 for further discussion of ditransitives in O'dam.

We have shown that preverbal quantifiers can quantify through all argumenthood functions of their associated verb. For control constructions, treating CCCs

[9] If it turns out that O'dam does have controlled objects then there would be another set of equations equivalent to (12,13) but with the specification $((\uparrow\mathrm{COMP\ OBJ})_{\sigma}\ \mathrm{ANTECEDENT})$.

$$\begin{bmatrix}
\text{PRED} & \text{`cause'} \\
\text{SUBJ} & \begin{bmatrix}\text{PRED} & \text{`pro'} \\ \text{PERS} & 2 \\ \text{NUM} & \text{SG}\end{bmatrix} \\
\text{OBJ} & \begin{bmatrix}\text{PRED} & \text{`pro'} \\ \text{PERS} & 1 \\ \text{NUM} & \text{SG}\end{bmatrix}_i \\
\text{COMP} & \begin{bmatrix}\text{PRED} & \text{`make'} \\ \text{SUBJ} & \begin{bmatrix}\text{PRED} & \text{`pro'} \\ \text{PERS} & 1 \\ \text{NUM} & \text{SG}\end{bmatrix}_i \\ \text{OBJ} & \begin{bmatrix}\text{PRED} & \text{`tacos'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{SG} \\ \text{QUANT} & \text{TWO}\end{bmatrix}\end{bmatrix}
\end{bmatrix}$$

Figure 3: F-structure for CCC with object antecedent

as having the COMP function captures the ability for quantifiers in the preverbal position of a control verb to also quantify arguments of the controlled verb. We now move to §4.2 where we will see that same is not true of NCSs.

4.2 Non-complement subordinate clauses

NCSs are diagnosed by 3SG OBJ coreference on a transitive verb, as shown in (16).[10] A list of attested verbs that permit an NCS object is shown in Table 4

(16) Jix=bhai' jix=Ø-maat [na cham ji'xkat jugio-ka' gu
COP=good COP=3SG.PO-know SUB NEG never eat-EST DET
tu']$_{NCS}$
something
'Because it is good for him to know that he could not eat it.

Most verbs that select for NCSs also permit nominal objects that receive a DP exponent, as seen in (17), where the 3PL OBJ prefix is coreferenced with the DP 'the men who live in Teneraca', which is not an NCS. However, when the antecedent is an NCS the coreferring verbal object prefix is 3SG even when it has a plural referent, as seen in (18) where the quantifier *bɨx* 'all' enforces a plural interpretation of the referent of the NCS (i.e. the places where my family members live).

[10] NCSs in O'dam must be selected for by the verb and we have no verbs that select for a clausal subject.

$$\begin{bmatrix} \text{PRED} & \text{`finish'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`pro'} \\ \text{PERS} & 1 \\ \text{NUM} & \text{SG}\end{bmatrix}_i \\ \text{COMP} & \begin{bmatrix} \text{PRED} & \text{`run'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`pro'} \\ \text{PERS} & 1 \\ \text{NUM} & \text{SG}\end{bmatrix}_i \end{bmatrix} \end{bmatrix}$$

Figure 4: F-structure for CCC with subject antecedent

Verb	Gloss	Verb	Gloss
aa'	'want'	*maat*	'know'
choodo'n	'be afraid of'	*niiya'*	'see'
iibhi'ñ	'fear'	*taata'*	'feel'
ilhdha'	'believe'	*titda'*	tell
kaaya'	'hear'	*ti'ñcho'*	'remember'
kaich	'say'		

Table 4: Attested verbs that permit NCS object

(17) Pix cham **ja-ñii'ñ-ap** [gu chi∼chio'ñ na=m kio
MIR NEG 3PL.PO-see-2SG.SBJ DET PL∼man SUB=3PL.SBJ live
mummu Chianarkam]$_{DP}$
DIST.LOWER Teneraca
'You have not ever seen the men who live in Teneraca'

(18) Añ joidham **ti-Ø-nii** [bix na=m pai' kio
1SG.SBJ enjoy DUR-3SG.PO-see all SUB=3PL.SBJ where live
gu=ñ pamil]$_{NCS}$
DET=1SG.POSS family
'I like all of the (various) places where my family lives'

When *maat* 'know' takes a NCS, as in (19), we see that the quantifier *bix* in the matrix preverbal position can quantify the NCS as a clause, in (19a), but not the dependents of the NCS, in (19b) and (19c) respectively.

(19) Bix jix=**Ø**-mat-iñ na=m jaroi' mii-'ñ
all COP=3SG.PO-know-1SG.SBJ SUB=3PL.SBJ someone burn-APPL
gu ku'a'
DET firewood

a. I know who.PL **completely burned the firewood** NCS

b. *I know who.PL burned **all of the firewood.** OBJECT

c. *I know **all of them** who burned the firewood. SUBJECT

In contrast, when *maat* 'know' takes a pronominal complement referring to an individual, as in (20), we see that it can quantify the ones who burned the firewood, in (20c), who are now the object of *maat*. However, in (20a) we see that now *bɨx* cannot quantify the BURN NCS like it could in (19a) when *maat* took an NCS object.

(20) Bɨx jix=**ja-**mat-iñ na=m jaroi' mɨɨ-'ñ gu
all COP=3PL.PO-know-1SG.SBJ SUB=3PL.SBJ someone burn-APPL DET
ku'a'
firewood

a. *I know who.PL **completely burned the firewood** NCS

b. *I know who.PL burned **all of the firewood.** OBJECT

c. I know **all of them** who burned the firewood. SUBJECT

We analyze these verbs as taking a function that is a pronominal OBJ that is coindexed with the NCS ADJ, rather than COMP as with CCCs. This OBJ is specified as being pronominal, and may be coreferenced with a CP realizing the clause. However, the CP appears in the f-structure with the grammatical function ADJ, rather than as an argument of the clause

The lack of preverbal quantification for arguments of the CP is now explained: The actual argument of the verb is a pronoun, referring to the NCS itself. However, the arguments of the NCS are only specified in f-structure (if at all) in an ADJ structure. Thus they fall outside the path specified by (↑ AF*) in our quantifier equation in (9).

The f-structure for (21) is shown in Figure 5.

(21) Bɨx **jix=Ø-mat-iñ** na=m jaroi' mɨɨ-'ñ
all COP=3SG.PO-know-1SG.SBJ SUB=3PL.SBJ someone burn-APPL
gu ku'a'
DET firewood
'I know who.PL completely burned the firewood' (Lit. I know that people completely burned the firewood)

In this section we have explained that giving the NCSs the ADJ grammatical function correctly captures the behavior of preverbal quantifiers. In §4.2.1 and §4.2.2 we will give evidence that verbs that previous work assumed selected for a NCS actually select for a pronominal OBJ with an clausal referent.

4.2.1 CP exponents of NCSs are headless relative clauses

When the referent of the NCS is not the eventuality, as in (22), we find that there is always a *wh*-word, in this case *pai'* 'where'.

$$\left[\begin{array}{ll} \text{PRED} & \text{`know'} \\ \text{SUBJ} & \left[\begin{array}{ll}\text{PRED} & \text{`pro'}\\ \text{PERS} & 1\\ \text{NUM} & \text{SG}\end{array}\right] \\ \text{OBJ} & \left[\begin{array}{ll}\text{PRED} & \text{`pro'}\\ \text{PERS} & 3\\ \text{NUM} & \text{SG}\\ \text{QUANT} & \text{ALL}\end{array}\right]_i \\ \text{ADJ} & \left\{ \left[\begin{array}{ll} \text{PRED} & \text{`burn'} \\ \text{SUBJ} & \left[\begin{array}{ll}\text{PRED} & \text{`someone'}\\ \text{PERS} & 3\\ \text{NUM} & \text{PL}\end{array}\right] \\ \text{OBJ} & \left[\begin{array}{ll}\text{PRED} & \text{`firewood'}\\ \text{PERS} & 3\\ \text{NUM} & \text{SG}\end{array}\right]\end{array}\right]_i \right\} \end{array}\right]$$

Figure 5: F-structure for NCS

(22) Añ joidham tɨ-Ø-nɨi [bɨx na=m **pai'** kio
1SG.SBJ enjoy DUR-3SG.PO-see all SUB=3PL.SBJ **where** live
gu=ñ pamil]$_{NCS}$
DET=1SG.POSS family
'I like all of the (various) places where my family lives'

We see in (23) that there is no *wh*-word in the NCS.

(23) Jix=bhai' jix=Ø-maat [na cham ji'xkat jugio-ka' gu
COP=good COP=3SG.PO-know SUB NEG never eat-EST DET
tu']$_{NCS}$
something
'Because it is good for him to know that he could not eat it.

García Salido (2021) finds such *wh*-words a diagnostic feature of headless relative clauses, which are always adjuncts, as in (24).

(24) Añ jix=io'm tu-jua [na **gu'** ap jix=io'm
1SG.SBJ COP=hard DUR-work.PFV SUB **why** 2SG.SBJ COP=hard
tu-jua]$_{headlessRC}$
DUR-work.PFV
''I worked hard because you worked hard.' (García Salido 2021: 70)

The syntactic shape of NCSs matches that of headless relative clauses when the referent is not an eventuality (i.e. they require a *wh*-word). This suggests that the

OBJ of the matrix verb is a pronominal that refers to the NCS, rather than the OBJ being the full clause itself, as with CCCs.

4.2.2 The 'personal' distinction

The difference in interpretation of verbs selecting for a typical DP object versus an object associated with NCS also suggests that for the latter, the NCS is treated as a relative clause of the elaborating on the OBJ of the matrix verb. In (25) we see two minimally contrastive sentences using the verb *maat* 'know'. Both sentences express that the speaker knows something about the multiple people who burned all of the firewood her friend had collected. In (25a) the object of *maat* 'know' is a 3PL pronoun refering to the individuals, which the headless relative clause modifies. This structure expresses that the speaker personally knows the people who burned the firewood. In (25b) the object of *maat* is a 3SG pronoun referring to the NCS, which the headless relative clause modifies. This structure expresses that the speaker did see who burned the firewood but does not know those people personally.

(25) a. Bɨx **jix=ja-mat-iñ** [na=m jaroi' mɨi-'ñ
all COP=3PL.PO-know-1SG.SBJ SUB=3PL.SBJ who burn-APPL
gu ku'a']$_{headlessRC}$
DET firewood
'I know who all burned the firewood' (Lit. I know all of them, who burned the firewood)

b. Bɨx **jix=Ø-mat-iñ** [na=m jaroi' mɨi-'ñ
all COP=3SG.PO-know-1SG.SBJ SUB=3PL.SBJ someone burn-APPL
gu ku'a']$_{headlessRC}$
DET firewood
'I know who.PL completely burned the firewood' (Lit. I know that people completely burned the firewood)

5 Interim summary: The LFG account

The analysis proposed so far accounts for which constituents can or can't receive preverbal quantification. Quantifiers assign a QUANT feature to any f-structure accessible via a path consisting only of argument functions. The arguments of a CCC are found in a COMP, so they can be quantified. The arguments of an NCS are in an ADJ and cannot be quantified.

However, the set of constituents that can be quantified is wider than the set diagnosed by verbal coreference. In ditransitives, only one object argument is coreferenced by verbal morphology, while both may be quantified, as shown in (26), where either the *recipient* or the *theme* may be quantified, while only the recipient is head-marked. We know that the OBJ in (26) is the recipient because

(26) is only acceptable with a 3PL recipient, coreferenced with *ja-*. If the recipient was the OBJ_θ, then it could be any person-number combination.

(26) Gok ja-maa-ñi-ch gu tɨ∼tbi-chuk
two 3PL.PO-give.PFV-1SG.SBJ-PFV DET PL∼play-POSSD
'I gave them two toys.'
'I gave toys to two (people).'

We thus have two mismatches between verbal coreference and preverbal quantification: OBJ_θ (non-coreferenced objects of a ditransitive) and properties of the event itself (often the scale). While this is a perfectly satisfactory syntactic analysis, it is incomplete, because it lacks a deeper explanation for the form of these grammatical functions. We turn to L_RFG to provide exactly that explanation.

6 An L_RFG analysis

Lexical-Realizational Functional Grammar (L_RFG) is a synthesis of Distributed Morphology (DM) as a theory of morphological realization and LFG as a theory of grammatical architecture (Melchin et al. 2020). As a descendant of LFG, it is a declarative, representational and constraint-based theory ideally suited to modelling nonconfigurationality, like in O'dam. As a descendant of DM, it provides a realizational, morpheme-based view of word-formation and is good at modelling complex morphological structures, including those found in highly agglutinative languages such as O'dam (Tallman et al. 2018). In L_RFG, as in DM, the terminal nodes of the c-structure are not words or morphemes (i.e., they contain no phonological material), but are instead bundles of features which are realized by Vocabulary Items (VIs) at v(ocabulary)-structure.

Our L_RFG account takes advantage of this distinction between c-structure and v-structure to account for the argumenthood mismatches. In L_RFG, features of all arguments are present in the c-structure nodes that map to the verb's f-structure. However, the VIs that realize these nodes are systematically specified only for certain grammatical functions. In this way, argumenthood and c-structure features are strictly correlated, while the features of the relevant nodes that get overt exponence are dependent on the VIs available to realize them.

We assume that features of all arguments, including both theme and recipient, are introduced by a node in the c-structure associated with the verb (i.e., in the V). A schematic L_RFG c-structure for the O'dam verb is shown in Figure 6.

In the c-structure of (26), the node hosting object agreement features, Agr_O, is specified for features of both the primary object (OBJ, the *recipient*) and the secondary object (OBJ_θ, the *theme*), as in (27).

(27) $\text{Agr}_\text{O}{}' \rightarrow \ldots \quad \underset{\text{@ObjAgree}}{\text{Agr}_\text{O}} \quad \ldots$

We use a template for object agreement (Dalrymple et al. 2004), where the optional material allows us to capture transitives and ditransitives in a single template:

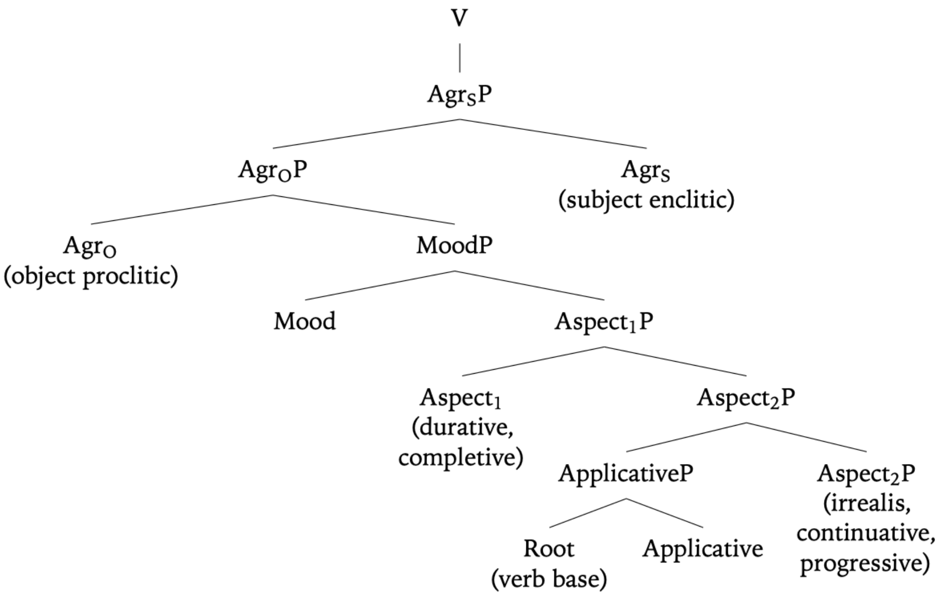


Figure 6: Schematic $\mathrm{L_RFG}$ c-structure for the O'dam verb

(28) ObjAgree:= $(\uparrow \textsc{obj pers}) = \{1|2|3\}$
$(\uparrow \textsc{obj num}) = \{\textsc{sg}|\textsc{pl}\}$
$\left(\begin{array}{l}(\uparrow \textsc{obj}_\theta \ \textsc{pers}) = \{1|2|3\}\\ (\uparrow \textsc{obj}_\theta \ \textsc{num}) = \{\textsc{sg}|\textsc{pl}\}\end{array}\right)$

However, the Vocabulary Items that realize $\mathrm{Agr_O}$ are only specified for features of one object, as in (29). The full list of subject and primary object markers is shown in Tables 1 and 2 in Section 4 above.

(29) $\langle\ [\mathrm{Agr_O}],\quad \Phi\left\{\begin{array}{l}(\uparrow \textsc{pluso}) = \texttt{\%gf}\\ (\uparrow \texttt{\%gf}\ \textsc{pers}) = 3\\ (\uparrow \texttt{\%gf}\ \textsc{num}) = \textsc{pl}\end{array}\right\}\rangle \quad \Rightarrow \quad$ *ja-*

The label PLUSO is a variable over OBJ and $\textsc{obj}_\theta$, as in Findlay (2016, 2020). The arbitrary local name `%gf` ensures that PERSON and NUMBER values are for the same argument. The choice of which of the two PLUSO arguments is expressed is due to a complex interaction between the available VIs and certain pragmatic factors (see for example García Salido 2014: 48ff). However, in either case there will be only one agreement morpheme available in the set of O'dam's VIs for the two object functions.[11]

[11]A reviewer raised the question of what advantage this analysis has over a traditional LFG analysis in which a single node in the c-structure can host either OBJ or $\textsc{obj}_\theta$. However, this raises the question of how the features of both object arguments can be in the f-structure, if only one of them appears in a given c-structure. This analysis allows us to have the features of both arguments present in c-structure, and therefore in f-structure, while only one is ever realized overtly.

We assume that the QUANT features are assigned by the f-description, such as that in (9) for *bɨx* 'all', in the c-structure node of the preverbal quantifier, regardless of whether there is overt agreement on the verb. In other words, the assignment of QUANT features is not dependent on overt morphology, though both are determined by the grammatical function (and thus, the argumenthood) of the relevant participants. When there is no surface morphology, we take this as evidence that the O'dam Vocabulary lacks such an exponent. This is cross-linguistically typical with so-called "unmarked" or high-frequency feature combinations; see for example the work of Haspelmath, whose viewpoint is summarized in Haspelmath and Sims (2010: ch. 12).

In $\mathrm{L_RFG}$, in cases like this where there is no VI that is dedicated to the expression of the relevant features, a linearly adjacent VI in the v-structure *spans* the unexpressed features. This allows the mapping between c-structure and v-structure to maximally satisfy the *MostInformative* constraint (Melchin et al. 2020: 273), which resolves the competition between forms by ensuring the v-structure realizes the largest subset of f-descriptions present in the c-structure using the smallest number of VIs. Thus, the relationship between terminal nodes and VIs is many-to-one in $\mathrm{L_RFG}$, using the mechanism of Spanning (Haugen and Siddiqi 2016, Merchant 2015, Ramchand 2008, Svenonius 2016) that was developed for DM and similar models; that is, one VI may realize features of multiple terminal nodes.

For this reason, the framework is similar to the Lexical Sharing model proposed for LFG by Wescoat (2002, 2005, 2007), but maintains the complex internal structures of words as part of syntax. One difference between $\mathrm{L_RFG}$ and Lexical Sharing is the notion which $\mathrm{L_RFG}$ calls *Pac-Man Spanning* (Haugen and Siddiqi 2016, Melchin et al. 2020: 284). According to Pac-Man Spanning, VIs can span any number of adjacent preterminal nodes, as long as the spanning doesn't obscure a meaning (including semantic/conventionalized presuppositions) that could otherwise be realized via an overt exponent. This is the $\mathrm{L_RFG}$ alternative to so-called "null morphemes" in most morpheme-based realizational models: lacking any dedicated exponent of its own, functional material is absorbed into the expression of a neighboring terminal.

This spanning of unmarked feature combinations can be seen in the O'dam agreement system when the object has 3rd-person singular features. While this is marked as Ø in the list of agreement markers in Table 1, and in examples such as (21), we assume that the $\mathrm{Agr_O}$ node hosting the features in these contexts is actually realized by the VI for the neighboring verb root. That is, we assume there is no VI of category $\mathrm{Agr_O}$ that realizes 3rd-person singular features of PLUSO arguments, rather than assuming the existence of a dedicated null morpheme specified for these features. Therefore, in the examples above, the symbol Ø in glosses should be taken to indicate this kind of spanning, rather than the presence of a null morpheme.

In this analysis, there are thus two reasons for mismatches between verbal coreference and argumenthood (and therefore preverbal quantification, which is dependent on argumenthood), both made available by the $\mathrm{L_RFG}$ framework. The first occurs when there is a VI available to realize some, but not all, of the features of the $\mathrm{Agr_O}$ terminal node. This occurs in ditransitives, where $\mathrm{Agr_O}$ has features of both OBJ and OBJ_θ, but the VIs realizing this category systematically only contain features of one PLUSO argument. Thus, the $\mathrm{Agr_O}$ VI realizes only a subset of the node's features. The second mismatch occurs when the object is 3rd-person singular, for which there is no $\mathrm{Agr_O}$ VI at all, in which case the node is realized by a neighboring node in an instance of Pac-Man Spanning.

7 Conclusion

Following Everdell's (2021) overview of O'dam argumenthood tests, we have shown that CCCs and NCSs pattern differently with regards to their argumenthood status, contra previous work that assumed they were both clausal complements. While CCCs as clausal complements of their control verb, NCSs pattern with adjuncts of their matrix verb, with the exception of the NCS as a whole. Combined with an analysis of CCCs as COMP and NCS as ADJ, this explains the differences in preverbal quantification of the arguments of the different types of clauses. Our analysis of the OBJ of an apparent NCS selecting verb as only having a pronominal OBJ with the NCS acting as a relative clause of that OBJ, explains the argumenthood status of that clause, as well as the varying shape of NCSs and the impersonal interpretation of verbs with an OBJ associated with an NCS. However, in LFG this account leaves unexplained the mismatches between preverbal quantification and the other main argumenthood diagnostic in O'dam, coreference by verbal affixes. In particular, a potential problem for standard LFG is that coreference only captures a subset of the arguments identified by preverbal quantification. These mismatches can be explained in $\mathrm{L_RFG}$ as mismatches between c-structure terminal nodes and their v-structure exponents, allowing arguments to be consistently present in c-structure.

References

Al Khalaf, Eman. 2019. Floating quantifiers are autonomous phrases: A movement analysis. *Glossa: a journal of general linguistics* 4(1): 89.

Asudeh, Ash. 2005. Control and semantic resource sensitivity. *Journal of Linguistics* 41: 465–511.

Barbu, Roxana-Maria. 2015. Verbs and Participants: Nonlinguists' Intuitions. Master's thesis, Carleton University.

Barbu, Roxana-Maria, and Ida Toivonen. 2016a. Arguments and Adjuncts: at the Syntax-Semantics Interface. In Ellen Thompson, ed., *Proceedings of the Florida Linguistics Yearly Meeting (FLYM) 3*.

—. 2016b. Event participants and linguistic arguments. In A. Papafragou, D. Grodner, D. Mirman, and J.C. Trueswell, eds., *Proceedings of the 38th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Butt, Miriam, and Ida Toivonen, eds. 2020. *Proceedings of the LFG20 Conference*. Stanford, CA: CSLI Publications.

Cappelen, Herman, and Ernest Lepore. 2005. *Insensitive semantics: A defense of semantic minimalism and speech act pluralism*. Oxford: Blackwell Publishing.

Dalrymple, Mary, Ronald M. Kaplan, and Tracy Holloway King. 2004. Linguistic Generalizations over Descriptions. In Miriam Butt and Tracy Holloway King, eds., *Proceedings of the LFG04 Conference*, 199–208. Stanford, CA: CSLI Publications.

Dalrymple, Mary, John J. Lowe, and Louise Mycock. 2019. *The Oxford Reference Guide to Lexical Functional Grammar*. Oxford: Oxford University Press.

Everdell, Michael. 2021. From Whence Arise Arguments? Argumenthood in O'dam. Poster presented at the *2021 SSILA Annual Meeting*. University of California, Berkeley. January 7-10, 2021.

—. in progress. Argument Realization in O'dam. Ph.D. thesis, The University of Texas at Austin.

—. in progress. Constituency in O'dam. In Adam J.R. Tallman, Sandra Auderset, and Hiroto Uchihara, eds., *Constituency and Convergence in the Americas*. Language Science Press.

Everdell, Michael, and Gabriela García Salido. 2021. Los aplicativos en o'dam. Paper presented at *The 2021 Friends of Uto-Aztecan Conference/Taller de los amigos de las lenguas yutoaztecas (FUAC/TALY)*. The University of Texas at Austin, June 3-5.

Everdell, Michael, and Paul B. Melchin. 2021. On the argument-adjunct distinction: Evidence from O'dam preverbal quantifiers. Ms., The University of Texas at Austin.

Findlay, Jamie Y. 2016. Mapping theory without argument structure. *Journal of Language Modelling* 4(2): 293–338.

—. 2020. Lexical Mapping Theory and the anatomy of a lexical entry. In Butt and Toivonen 2020, 127–147.

García Salido, Gabriela. 2014. Clause Linkage in Southeastern Tepehuan, a Uto-Aztecan Language of Northern Mexico. Ph.D. thesis, The University of Texas, Austin.

—. 2021. Headless Relative Clauses in Southeastern Tepehuan (O'dam). In Ivano Caponigro, Harold Torrence, and Roberto Zavala Maldonado, eds., *Headless Relative Clauses in Mesoamerica*, 58–78. New York: Oxford University Press.

García Salido, Gabriela, and Michael Everdell. 2020. Southern Tepehuan (Durango and Narayit, Mexico) - Language Snapshot. *Language Documentation and Description* 19: 89–98.

Haspelmath, Martin, and Andrea D. Sims. 2010. *Understanding Morphology*. New York: Routledge, 2nd edn.

Haugen, Jason D., and Daniel Siddiqi. 2016. Towards a restricted realizational theory: Multimorphemic monolistemicity, portmanteaux, and post-linearization spanning. In *Morphological Metatheory*, 343–386. Amsterdam: John Benjamins.

INEGI. 2015. Instituto Nacional de Estadística y Geografía. http://www.inegi.org.mx/default.aspx.Mexico.

Landau, Idan. 2000. *Elements of control: Structure and meaning in infinitival constructions*, vol. 51. Springer Science & Business Media.

Melchin, Paul B., Ash Asudeh, and Dan Siddiqi. 2020. Ojibwe Agreement in Lexical-Realizational Functional Grammar. In Butt and Toivonen 2020, 268–288.

Merchant, Jason. 2015. How much context is enough? Two cases of span-conditioned stem allomorphy. *Linguistic Inquiry* 46(2): 273–303.

Moura, Heronides, and Rafaela Miliorini. 2018. Toward a comprehension of an intuition: criteria for distinguishing verbal arguments and adjuncts/Para compreender uma intuicao: criterios para distinguir argumentos de adjuntos verbais. *Alfa: Revista de Lingüística* 62(3): 575–594.

Needham, Stephanie, and Ida Toivonen. 2011. Derived arguments. In Miriam Butt and Tracy Holloway King, eds., *Proceedings of the LFG11 Conference*, 401–421. Stanford: CSLI.

Ramchand, Gillian. 2008. *Verb Meaning and the Lexicon*. Cambridge MA: MIT Press.

Ramchand, Gillian, and Peter Svenonius. 2014. Deriving the functional hierarchy. *Language sciences* 46: 152–174.

Stiebels, Barbara. 2007. Towards a typology of complement control. *ZAS papers in linguistics* 47(8).

Svenonius, Peter. 2016. Spans and words. In Daniel Siddiqi and Heidi Harley, eds., *Morphological Metatheory*, 201–222. John Benjamins.

Tallman, Adam J.R., Dennis Wylie, Eric Adell, Natalia Bermudez, Gladys Camacho, Patience Epps, Michael Everdell, Ambrocio Gutierrez, Cristian Juarez, and Anthony C. Woodbury. 2018. Constituency and the morphology-syntax divide in the languages of the Americas: towards a distributional typology. In *21st Annual Workshop on American Indigenous Languages*. University of California Santa Barbara.

Wescoat, Michael T. 2002. *On lexical sharing*. Stanford University.

—. 2005. English nonsyllabic auxiliary contractions: An analysis in LFG with lexical sharing. In Miriam Butt and Tracy Holloway King, eds., *Proceedings of the LFG05 Conference*, 468–486.

—. 2007. Preposition-determiner contractions: an analysis in optimality-theoretic lexical-functional grammar with lexical sharing. In Miriam Butt and Tracy Holloway King, eds., *Proceedings of the LFG07 Conference*, 439–459.

Willett, Elizabeth R., and Thomas L. Willett. 2013. *Diccionario Tepehuano de Santa María de Ocotán, Durango*. Mexico: Summer Institute of Linguistics.

Willett, Thomas L. 1991. *A reference grammar of Southeastern Tepehuan*. Dallas, TX: Summer Institute of Linguistics.

Approaches to scope islands in LFG+Glue

Matthew Gotham
University of Oxford

Proceedings of the LFG'21 Conference

On-Line

Miriam Butt, Jamie Y. Findlay, Ida Toivonen (Editors)

2021

CSLI Publications

pages 146–166

http://csli-publications.stanford.edu/LFG/2021

Keywords: scope, scope islands, quantification, Glue semantics, type logic

Gotham, Matthew. 2021. Approaches to scope islands in LFG+Glue. In Butt, Miriam, Findlay, Jamie Y., & Toivonen, Ida (Eds.), *Proceedings of the LFG'21 Conference, On-Line*, 146–166. Stanford, CA: CSLI Publications.

Abstract

In this paper I examine two possible approaches to scope islands in LFG with Glue semantics: one in which constraints on scope level are imposed by means of constraints off the path of an inside-out functional uncertainty, and one in which they are imposed through the structural rules of the fragment of linear logic used for meaning composition, by making the fragment multi-modal. For each approach, I show how it could be made to account for novel empirical arguments made by Barker (2021), and go on to argue in favour of the multi-modal Glue approach.

1 Introduction

At the outset of theory design in formal linguistics, the theorist is faced with a fundamental choice. Do you start with something relatively constrained, and then find ways to loosen it as the evidence demands? Or do you start with something relatively *un*constrained, and then find ways to constrain it as required? As a theory of the syntax/semantics interface, Glue semantics is towards the unconstrained end of the spectrum. This paper is addressed at the need to constrain Glue with respect to the phenomenon of quantifier scope, in particular the (non-)ability of a quantified noun phrase to take scope outside of its minimal clause.

1.1 Background

It is a feature of the Glue approach to semantic composition that many instances of quantifier scope ambiguity are resolved purely at the level of linear logic proofs. For example, the two interpretations of (1) shown below, surface scope and inverse scope respectively, can both be derived from the same f-structure and the same associated meaning constructors, as shown in (2) and (3) respectively.[1]

(1) Someone sees everything.
$\Rightarrow$ $\mathsf{someone}(\lambda x.\mathsf{everything}(\lambda y.\mathsf{see}(x, y)))$
$\Rightarrow$ $\mathsf{everything}(\lambda y.\mathsf{someone}(\lambda x.\mathsf{see}(x, y)))$

[†]I thank the audience of LFG'21 for helpful and encouraging feedback. This research is funded by an Early Career Fellowship from the Leverhulme Trust.

[1]The subscripts e (entities) and p (propositions) represent types. Semantically, we can think of p as equivalent to $s \rightarrow t$. I use $\boxed{n}$ and $\mathbf{n}$ interchangeably for f-structure labels / linear logic formulae. We have the following logical constants on the meaning side:

$\mathsf{every}, \mathsf{some} :: ((e \rightarrow p) \times (e \rightarrow p)) \rightarrow p$
$\mathsf{everything}, \mathsf{something}, \mathsf{someone} :: (e \rightarrow p) \rightarrow p$
$\mathsf{not} :: p \rightarrow p$

$$
\dfrac{\begin{array}{c}\mathsf{someone} : \\ (\mathbf{1}_e \multimap \mathbf{0}_p) \multimap \mathbf{0}_p\end{array} \qquad \dfrac{\dfrac{\begin{array}{c}\mathsf{everything} : \\ (\mathbf{2}_e \multimap \mathbf{0}_p) \multimap \mathbf{0}_p\end{array} \qquad \dfrac{\dfrac{\dfrac{\begin{array}{c}\lambda v.\lambda u.\mathsf{see}(u,v) : \\ \mathbf{2}_e \multimap (\mathbf{1}_e \multimap \mathbf{0}_p)\end{array} \quad [y : \mathbf{2}_e]^1}{\lambda u.\mathsf{see}(u,y) : \mathbf{1}_e \multimap \mathbf{0}_p} \qquad [x : \mathbf{1}_e]^2}{\mathsf{see}(x,y) : \mathbf{0}_p}}{\lambda y.\mathsf{see}(x,y) : \mathbf{2}_e \multimap \mathbf{0}_p} \multimap \mathrm{I}^1}{\mathsf{everything}(\lambda y.\mathsf{see}(x,y)) : \mathbf{0}_p}}{\lambda x.\mathsf{everything}(\lambda y.\mathsf{see}(x,y)) : \mathbf{1}_e \multimap \mathbf{0}_p} \multimap \mathrm{I}^2}{\mathsf{someone}(\lambda x.\mathsf{everything}(\lambda y.\mathsf{see}(x,y))) : \mathbf{0}_p}
$$

Figure 1: Derivation of the surface scope interpretation of (1) from (3)

$$
\dfrac{\begin{array}{c}\mathsf{everything} : \\ (\mathbf{2}_e \multimap \mathbf{0}_p) \multimap \mathbf{0}_p\end{array} \qquad \dfrac{\dfrac{\begin{array}{c}\mathsf{someone} : \\ (\mathbf{1}_e \multimap \mathbf{0}_p) \multimap \mathbf{0}_p\end{array} \qquad \dfrac{\begin{array}{c}\lambda v.\lambda x.\mathsf{see}(x,v) : \\ \mathbf{2}_e \multimap (\mathbf{1}_e \multimap \mathbf{0}_p)\end{array} \quad [y : \mathbf{2}_e]^1}{\lambda x.\mathsf{see}(x,y) : \mathbf{1}_e \multimap \mathbf{0}_p}}{\mathsf{someone}(\lambda x.\mathsf{see}(x,y)) : \mathbf{0}_p}}{\lambda y.\mathsf{someone}(\lambda x.\mathsf{see}(x,y)) : \mathbf{2}_e \multimap \mathbf{0}_p} \multimap \mathrm{I}^1}{\mathsf{everything}(\lambda y.\mathsf{someone}(\lambda x.\mathsf{see}(x,y))) : \mathbf{0}_p}
$$

Figure 2: Derivation of the inverse scope interpretation of (1) from (3)

(2)
$$
\boxed{0}\left[\begin{array}{ll}
\text{PRED} & \text{`see}\langle\boxed{1},\boxed{2}\rangle\text{'} \\
\text{TENSE} & \text{PRES} \\
\text{SUBJ} & \boxed{1}\left[\begin{array}{ll}\text{PRED} & \text{`someone'}\end{array}\right] \\
\text{OBJ} & \boxed{2}\left[\begin{array}{ll}\text{PRED} & \text{`everything'}\end{array}\right]
\end{array}\right]
$$

(3)
$$
\begin{array}{r}
\mathsf{someone} : (\mathbf{1}_e \multimap \mathbf{0}_p) \multimap \mathbf{0}_p \\
\lambda y.\lambda x.\mathsf{see}(x,y) : \mathbf{2}_e \multimap (\mathbf{1}_e \multimap \mathbf{0}_p) \\
\mathsf{everything} : (\mathbf{2}_e \multimap \mathbf{0}_p) \multimap \mathbf{0}_p
\end{array}
$$

The proofs deriving these interpretations are shown in Figures 1 and 2, respectively.[2] However, not all instances of scope ambiguity can be handled quite as simply as this.

1.2 Scope level

Consider (4), which has the surface scope and inverse linking interpretations shown below, and has the (simplified) f-structure shown in (5).

(4) A member of every board resigned.

$\Rightarrow$ $\mathsf{some}(\lambda x.\mathsf{every}(\mathsf{board}, \lambda y.\mathsf{member\text{-}of}(x,y)), \mathsf{resign})$

[2]Throughout this paper, un-annotated steps of inference should be read as instances of $\multimap$ elimination, to save space.

'Someone who is a member of every board resigned.'
$\Rightarrow$ $\mathsf{every}(\mathsf{board}, \lambda y.\mathsf{some}(\lambda x.\mathsf{member\text{-}of}(x, y), \mathsf{resign}))$
'For every board, someone member of that board resigned.'

(5)
$$\boxed{0}\begin{bmatrix} \textsc{pred} & \text{'resign}\langle\boxed{1}\rangle\text{'} \\ \textsc{subj} & \boxed{1}\begin{bmatrix} \textsc{pred} & \text{'member}\langle\boxed{2}\rangle\text{'} \\ \textsc{spec} & \begin{bmatrix} \textsc{pred} & \text{'a'} \end{bmatrix} \\ \textsc{obj} & \boxed{2}\begin{bmatrix} \text{"every board"} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Unlike (1), the difference between the two interpretations of (4) *does* depend on a difference in meaning constructors. Specifically, the meaning constructor associated with *every board* in (5) is as shown schematically in (6), where $\Box$ is a placeholder.

(6) $\lambda P.\mathsf{every}(\mathsf{board}, P) : (\mathbf{2}_e \multimap \Box_p) \multimap \Box_p$

To derive the surface scope interpretation, the formula shown as $\Box$ in (6) has to be $\boxed{1}$, while to derive the inverse linking interpretation, it has to be $\boxed{0}$. I will refer to this choice as the choice of 'scope level' for a meaning constructor. Examples like (5) differ from those like (2) in that there *is* a choice of scope level for at least one quantifier—and this choice moreoever matters.

In the literature, there are essentially two approaches to resolving scope level. The first, adopted e.g. by Lev (2007), Andrews (2010) and Gotham (2019), is to treat this an an instance of functional uncertainty. The lexical entry for *every* could contain the description shown in (7), where we leave SCOPEPATH unspecified for now but note that it should at least include OBJ (for $\boxed{1}$) and SUBJ OBJ (for $\boxed{0}$).[3]

(7) $\%A = (\textsc{ScopePath}\ \uparrow)$
$\lambda P.\lambda Q.\mathsf{every}(P, Q) : (\uparrow_e \multimap \uparrow_p) \multimap ((\uparrow_e \multimap \%A_p) \multimap \%A_p)$

The second approach, which is more widely adopted (including in Dalrymple et al. (2019)), is to use quantification in the linear logic fragment to express the various possible scope levels. The meaning constructor for *every* would then look something like (8).

(8) $\lambda P.\lambda Q.\mathsf{every}(P, Q) : (\uparrow_e \multimap \uparrow_p) \multimap \forall X((\uparrow_e \multimap X_p) \multimap X_p)$

[3]In the fragment of second-order linear logic most often assumed for Glue, as described e.g. in (Dalrymple et al. 2019, Chapter 8), it would be strictly speaking incoherent to have both $\mathbf{2}_e$ and $\mathbf{2}_p$, since the subscripts are supposed to be sort labels and a formula cannot belong to more than one sort. Nevertheless, you *can* do this in the XLE+Glue implementation (Dalrymple et al. 2020). I am not able to speak to what is going on under the hood in the implementation, but one coherent way to interpret the notation would be to take the subscripts to be unary propositional functions, as in Gotham and Haug (2018). Another would be to switch from a second-order to a first-order system, and treat e and p as predicates to which the f-structure labels are arguments (Kokkonidis 2008).

In the proof, X can then be instantiated to either $\mathbf{0}$ or $\mathbf{1}$, deriving the respective interpretations, as shown below.

$$\dfrac{\begin{array}{c}\lambda Q.\mathsf{every}(\mathsf{board}, Q) : \\ \forall X.(\mathbf{2}_e \multimap X_p) \multimap X_p\end{array}}{\begin{array}{c}\lambda Q.\mathsf{every}(\mathsf{board}, Q) : \\ (\mathbf{2}_e \multimap \mathbf{0}_p) \multimap \mathbf{0}_p\end{array}}\ \forall_E \qquad \text{or} \qquad \dfrac{\begin{array}{c}\lambda Q.\mathsf{every}(\mathsf{board}, Q) : \\ \forall X.(\mathbf{2}_e \multimap X_p) \multimap X_p\end{array}}{\begin{array}{c}\lambda Q.\mathsf{every}(\mathsf{board}, Q) : \\ (\mathbf{2}_e \multimap \mathbf{1}_p) \multimap \mathbf{1}_p\end{array}}\ \forall_E$$

2 Scope islands

The point of departure for this paper is the fact that this choice of scope level is not entirely free. Consider, for example, (9), which has the surface scope interpretation, but *not* the inverse scope interpretation—it can only mean that there is a particular warden who thinks that every prisoner escaped, and not that for every prisoner, there is some warden or other who thinks that prisoner escaped.

(9) A warden thinks that every prisoner escaped.

$\Rightarrow$ $\mathsf{some}(\mathsf{warden}, \lambda x.\mathsf{think}(x, \mathsf{every}(\mathsf{prisoner}, \mathsf{escape})))$

$\not\Rightarrow$ $\mathsf{every}(\mathsf{prisoner}, \lambda y.\mathsf{some}(\mathsf{warden}, \lambda x.\mathsf{think}(x, \mathsf{escape}(y))))$

Given the (simplified) f-structure for (9) shown in (10), that would amount to saying that the meaning constructor associated with *every* (or *every prisoner*) can take $\boxed{2}$ as its scope level, but not $\boxed{0}$.

(10)
$$\boxed{0}\left[\begin{array}{ll}\text{PRED} & \text{`think}\langle\boxed{1},\boxed{2}\rangle\text{'} \\ \text{TENSE} & \text{PRES} \\ \text{SUBJ} & \boxed{1}\left[\text{“a warden”}\right] \\ \text{COMP} & \boxed{2}\left[\begin{array}{ll}\text{PRED} & \text{`escape}\langle\boxed{3}\rangle\text{'} \\ \text{TENSE} & \text{PAST} \\ \text{SUBJ} & \boxed{3}\left[\text{“every prisoner”}\right]\end{array}\right]\end{array}\right]$$

The received wisdom (May 1977) about examples like these is that the inverse scope interpretation is unavailable because finite clauses are ‘scope islands’, meaning that no quantifier inside of one can take scope out of it. The received wisdom seems to favour the functional uncertainty approach to scope level, as this constraint can be imposed by appropriately defining SCOPEPATH from (7), e.g. as shown in (11). By contrast, it is harder to see how such a constraint could be stated in the approach using quantification in the linear logic fragment to fix scope level.

(11)
$$\text{SCOPEPATH} \equiv \left(\begin{array}{cc}\text{GF}^* & \text{GF} \\ \neg(\rightarrow \text{TENSE}) & \end{array}\right)$$

The well-known fact that indefinites are not so constrained—that they can take 'exceptional scope' (Charlow 2014), as in (12)—can then be accounted for by allowing their scope level to be fixed by a less constrained path.[4]

(12) Every warden thinks that a prisoner escaped.

$\Rightarrow$ $\mathsf{every}(\mathsf{warden}, \lambda x.\mathsf{think}(x, \mathsf{some}(\mathsf{prisoner}, \mathsf{escape})))$

$\Rightarrow$ $\mathsf{some}(\mathsf{prisoner}, \lambda y.\mathsf{every}(\mathsf{warden}, \lambda x.\mathsf{think}(x, \mathsf{escape}(y))))$

2.1 Varieties of scope island

However, it is becoming increasingly clear that the received wisdom is too simplistic. As pointed out by Barker (2021), not all finite clauses are scope islands for all quantifiers. For example, (13) *does* have an interpretation where *every prisoner* takes widest scope, as shown—it *can* mean that for every prisoner, some accomplice or other ensured that that prisoner escaped.

(13) An accomplice ensured that every prisoner escaped.

$\Rightarrow$ $\mathsf{some}(\mathsf{accomplice}, \lambda x.\mathsf{ensure}(x, \mathsf{every}(\mathsf{prisoner}, \mathsf{escape})))$

$\Rightarrow$ $\mathsf{every}(\mathsf{prisoner}, \lambda y.\mathsf{some}(\mathsf{accomplice}, \lambda x.\mathsf{ensure}(x, \mathsf{escape}(y))))$

So, *every N* can take scope out of a finite clause, provided that clause is embedded by *ensured*. Note, however, that this does not mean that the clause embedded by *ensured* is not a scope island at all. As shown in (14), it *is* a scope island for *no N*. I have marked (14) as questionable because the one interpretation it does have conflicts with world knowledge about what it means to be an accomplice; it can only mean (implausibly) that there is a particular accomplice who ensured that no prisoner escaped, and not (more plausibly) that no prisoner is such that some accomplice or other ensured that that prisoner escaped.

(14) ?An accomplice ensured that no prisoner escaped.

$\Rightarrow$ $\mathsf{some}(\mathsf{accomplice}, \lambda x.\mathsf{ensure}(x, \mathsf{not}(\mathsf{some}(\mathsf{prisoner}, \mathsf{escape}))))$

$\not\Rightarrow$ $\mathsf{not}(\mathsf{some}(\mathsf{prisoner},$
$\lambda y.\mathsf{some}(\mathsf{accomplice}, \lambda x.\mathsf{ensure}(x, \mathsf{escape}(y)))))$

These observations invite the hypotheses that, in some sense, (i) *think* induces a stronger scope island than *ensure*, and (ii) *every N* is a stronger island-escaper than *no N*. The hypotheses are confirmed by filling in the gap in the paradigm: since the complement of *think* is a scope island for *every N*, if *every N* is a stronger island-escaper than *no N*, then we expect the complement of *think* to be a scope island for *no N* as well. This prediction is borne out, as shown in (15).

(15) A warden thinks that no prisoner esaped.

$\Rightarrow$ $\mathsf{some}(\mathsf{warden}, \lambda x.\mathsf{think}(x, \mathsf{not}(\mathsf{some}(\mathsf{prisoner}, \mathsf{escape}))))$

$\not\Rightarrow$ $\mathsf{not}(\mathsf{some}(\mathsf{prisoner}, \lambda y.\mathsf{some}(\mathsf{warden}, \lambda x.\mathsf{think}(x, \mathsf{escape}(y)))))$

[4] We ignore the possibility of treating indefinites as something other than quantifiers, semantically.

Meanwhile, being a strong enough island-escaper to take scope out of the complement of *think*, *an N* can certainly take scope out of the complement of *ensure*:

(16) Every accomplice ensured that a prisoner escaped.

$\Rightarrow$ $\textsf{every}(\textsf{accomplice}, \lambda x.\textsf{ensure}(x, \textsf{some}(\textsf{prisoner}, \textsf{escape})))$

$\Rightarrow$ $\textsf{some}(\textsf{prisoner}, \lambda y.\textsf{every}(\textsf{accomplice}, \lambda x.\textsf{ensure}(x, \textsf{escape}(y))))$

These data imply an implicational relationship, which Barker (2021) dubs the 'Scope Island Subset Constraint' (SISC):

SISC Given any two scope takers, the set of scope islands that trap one is a subset of the set of scope islands that trap the other.

So far we have only looked at three scope-takers and two clause-embedders, but a further piece of evidence in favour of the SISC comes from the behaviour of negative polarity items (NPIs). To be licensed, an NPI must be interpreted within the scope of an appropriate 'negative' licensor—Fry (1999) shows a method for ensuring this in LFG+Glue. However, as is acknowledged by Fry (1999), this method has the shortcoming that it does not ensure that an NPI be interpreted in the scope of its *closest* relevant licensor. For example, in (17) there are two potential licensors for the NPI *anyone*—*surprised* and *didn't*—but the NPI has to be interpreted as scoping under both of them, as shown.

(17) Martha is surprised that Mary didn't help anyone.

$\Rightarrow$ $\textsf{surprise}(\textsf{not}(\textsf{someone}(\lambda x.\textsf{help}(\textsf{mary}, x))), \textsf{martha})$

$\not\Rightarrow$ $\textsf{surprise}(\textsf{someone}(\lambda x.\textsf{not}(\textsf{help}(\textsf{mary}, x))), \textsf{martha})$

That is to say, (17) can mean that Martha is surprised that there's no-one that Mary helped, but not that Martha is surprised that there's someone that Mary didn't help (or equivalently, that Martha is surprised that Mary didn't help *everyone*). A natural explanation for this distinction would be that, in addition to being *licensors* for NPIs, at least some such expressions—such as overt negation—also induce *scope islands* for NPIs.

Meanwhile, like *a N* but unlike *every N* and *no N*, *any N* can take scope out of a clause embedded by *thinks*, as (18) shows.

(18) If Mary thinks anyone is to blame, that person is Bob.

$\Rightarrow$ $\textsf{if}(\textsf{someone}(\lambda x.\textsf{think}(\textsf{mary}, \textsf{blame}(x))), \textsf{think}(\textsf{mary}, \textsf{blame}(\textsf{bob})))$

Here, the antecedent of the conditional provides the relevant context for NPI licensing. The form of the consequent is chosen so as to privilege the interpretation of the antecedent according to which *there is someone* that Mary thinks is to blame, i.e. in which *anyone* takes scope over *thinks* (but under *if*, which licenses it).

So, *any N* seems to be a weaker island-escaper than *a N*, but a stronger island-escaper than *every N* and *no N*. The SISC therefore predicts, given the fact that negation induces a scope island for *any N*, that it also induces a scope island for

clause embedder	quantifier *an N*	*any N*	*every N*	*no N*	**island strength**
not		*	*	*	3
think			*	*	2
ensure				*	1
escaper strength	3	2	1	0	

Table 1: Relative strength of islands and escapers

every N and *no N*. Once again, the prediction is borne out, as shown in (19) and (20) respectively.[5]

(19) Jesus didn't heal everyone.

$\Rightarrow$ $\mathsf{not}(\mathsf{everyone}(\lambda x.\mathsf{heal}(\mathsf{jesus}, x)))$
$\equiv \mathsf{someone}(\lambda x.\mathsf{not}(\mathsf{heal}(\mathsf{jesus}, x)))$
$\not\Rightarrow$ $\mathsf{everyone}(\lambda x.\mathsf{not}(\mathsf{heal}(\mathsf{jesus}, x)))$
$\equiv \mathsf{not}(\mathsf{someone}(\lambda x.\mathsf{heal}(\mathsf{jesus}, x)))$

(20) Simon didn't receive nothing.

$\Rightarrow$ $\mathsf{not}(\mathsf{not}(\mathsf{something}(\lambda x.\mathsf{receive}(\mathsf{simon}, x))))$
$\equiv \mathsf{something}(\lambda x.\mathsf{receive}(\mathsf{simon}, x))$
$\not\Rightarrow$ $\mathsf{not}(\mathsf{something}(\lambda x.\mathsf{not}(\mathsf{receive}(\mathsf{simon}, x))))$
$\equiv \mathsf{everything}(\lambda x.\mathsf{receive}(\mathsf{simon}, x))$

We can summarise the empirical landscape in Table 1, adapted from (Barker 2021, Table 1). An asterisk in a cell means that the relevant scope taker is unable to take scope out of the island induced by the relevant clause embedder.[6] In the following two sections I will outline and evaluate two possible approaches to these data.

3 Blocking features and off-path constraints

It is still possible to impose some of the relevant constraints on scope level using the kind of inside-out functional uncertainty technique exemplified in (11). However, additional difficulties arise with the attempt. First of all, it is not clear how the scope island induced by negation can be accounted for, since the mainstream view of negation in LFG is that it is represented in f-structure either by the value of a NEG or POL feature, or as a member of the ADJ set, at the matrix level (Dalrymple et al. 2019, 67–69). The point is that in none of these accounts does negation embed the f-structure representing the negatum, and so the issue of scope level does not

[5]In some varieties of English, (20) has a negative concord interpretation, where *nothing* is interpreted as equivalent to the NPI *anything*. This is a separate issue which does not affect the discussion.

[6]The table in Barker (2021) is somewhat different, partly because he considers issues that there is not space to address here, for example the semantics of focus.

arise.

For example, take (21), representing a simplified version of the f-structure of (19) according to the INESS XLE-WEB (Rosén et al. 2012). Since the f-structure introduced by negation does not lie on the path between the f-structure for *everyone* and any possible scope level, it cannot be used to constrain quantifier scope—there is only one possible scope level for *everyone*: $\boxed{0}$.

(21)
$$\boxed{0}\left[\begin{array}{ll} \textsc{pred} & \text{`heal}\langle\boxed{1},\boxed{2}\rangle\text{'} \\ \textsc{subj} & \boxed{1}\left[\begin{array}{ll}\textsc{pred} & \text{`Jesus'}\end{array}\right] \\ \textsc{obj} & \boxed{2}\left[\begin{array}{ll}\textsc{pred} & \text{`everyone'}\end{array}\right] \\ \textsc{adj} & \left\{\boxed{3}\left[\begin{array}{ll}\textsc{pred} & \text{`not'}\end{array}\right]\right\} \end{array}\right]$$

In fact, we could view this issue with negation as an instance of a more general question, namely what the connection is between extra- and intra-clausal scope rigidity phenomena: ‘scope islands’ and ‘scope freezing’, respectively. If we use features in f-structure to impose constraints on scope level and thus account for scope islands, we need a completely different account of scope freezing. In many languages simple two-quantifier sentences like (1) are *not* ambiguous, for instance—an empirical fact that needs accounting for given Glue’s general unconstrainedness. In Gotham (2019) I proposed an account of scope freezing for examples like this, which could certainly be combined with a blocking features-based account of scope islands (as I suggested there), but perhaps it would be preferable to account for both kinds of scope rigidity within the same framework. I will return to this issue in Section 5.

With respect to the difference between *think* and *ensure*-type verbs, we *can* impose the relevant constraints by introducing different types of blocking feature. One way of doing so is exemplified in Figure 3, where we have sentence-embedding verbs projecting a SCOPEISLAND feature into their complement f-structure, and scope-takers sensitive to those features.

If we leave aside negation, the feature specifications in Figure 3 capture the facts in Table 1. But even so, problems remain. For one thing, it remains an open question as to whether or not the SCOPEISLAND feature can be independently motivated. Verbs of attitude and perception seem to pattern with *think*, so one could argue that there is a semantic generalization, but at the moment the only thing for this feature to do would be to enforce scope-islandhood.

Another potential problem relates to the SISC. Given the setup in Figure 3, there is nothing about the theory that prevents us from giving a lexical entry for a quantifier *nunone* that can take scope out of a clause embedded by *think* but *not* *thinks* V

$(\uparrow \textsc{comp}\ \textsc{ScopeIsland}) = 2$

ensured V

$(\uparrow \textsc{comp}\ \textsc{ScopeIsland}) = 1$

someone N

$\%X = (\textsc{gf}^{*}\ \textsc{gf}\uparrow)$

$\mathsf{someone} : (\uparrow_e \multimap \%X_p) \multimap \%X_p$

everyone N

$$\%Y = \left(\begin{array}{cc} \textsc{gf}^{*} & \textsc{gf}\uparrow \\ (\rightarrow \textsc{ScopeIsland}) \neq 2 & \end{array}\right)$$

$\mathsf{everyone} : (\uparrow_e \multimap \%Y_p) \multimap \%Y_p$

no-one N

$$\%Z = \left(\begin{array}{cc} \textsc{gf}^{*} & \textsc{gf}\uparrow \\ (\rightarrow \textsc{ScopeIsland}) \neq \{1 \mid 2\} & \end{array}\right)$$

$\lambda P.\mathsf{not}(\mathsf{someone}(P)) : (\uparrow_e \multimap \%Z_p) \multimap \%Z_p$

Figure 3: Possible lexical constraints on scope

out of a clause embedded by *ensure*, as shown in (22).

(22) *nunone* N

$$\%C = \left(\begin{array}{cc} \textsc{gf}^{*} & \textsc{gf}\uparrow \\ (\rightarrow \textsc{ScopeIsland}) \neq 1 & \end{array}\right)$$

$\mathsf{nunone} : (\uparrow_e \multimap \%C_p) \multimap \%C_p$

If we wanted to state the SISC as a grammar-wide constraint, then, we would have to do so by stating a constraint on the form of possible descriptions, to rule out lexical entries like (22). This is not impossible, but the relevant constraint would in all likelihood be quite messy and it is an open question exactly what form it would take. In view of these limitations, it is worth considering an alternative.

4 Multi-modal Glue semantics

An alternative approach is to impose the relevant constraints within Glue semantics itself. This requires some complication of the linear logic fragment used in Glue, but it can be argued on the basis of the data we have seen that the complication is linguistically motivated.

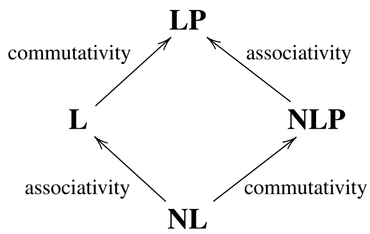


Figure 4: Substructural type logics (Moot and Retoré 2012, 111)

Within the family of substructural type logics, the base fragment[7] of linear logic used in Glue is equivalent to the Lambek calculus with permutation **LP**. It is thus both commutative and associative, and so relates to the others as shown in Figure 4. Commutativity means that the premises in a proof have no particular order (or, equivalently, can be freely reordered), as shown schematically in (23), and associativity means that the premises in a proof have no particular grouping (or, equivalently, can be freely regrouped), as shown schematically in (24).

(23) $$\frac{(\Gamma, \Delta) \vdash A}{(\Delta, \Gamma) \vdash A}$$

(24) $$\frac{((\Gamma, \Delta), \Sigma) \vdash A}{(\Gamma, (\Delta, \Sigma)) \vdash A}$$

So far, **LP** has been a good choice of logic for Glue: unlike in categorial grammar, the logic is not meant to account for word order and so it makes sense for it to be commutative. So far it has also made sense for the logic to be associative, but scope islands may actually give us a reason to care about how premises are grouped, and so restrict associativity. We can do so selectively by combining elements of **LP** (as before) and **NLP** (which is non-associative) in a multimodal system, where the modes correspond to the island/escaper strengths outlined in Table 1. An implementation of these ideas is given by the rules of inference shown in Figure 5, in combination with the lexical entries shown in Figure 6.[8] In the lexical entries, $\multimap$ (without a mode index) is shorthand for $\multimap_{\rceil 0}$, and $\multimap_i$ means the choice of index is free (so these can be seen as parameterized lexical entries).

The idea behind the rules in Figure 5 is that the $\rfloor$ modes show blocking strength, and the $\rceil$ modes show escaping strength. Note that now, because we no longer assume generalized associativity, there is bracketing on the left hand side of sequents. The mode indices on those brackets correspond to mode indices on occurrences of $\multimap$. Commutativity is ensured by the structural rule P (for *permutation*), and we have restricted associativity thanks to the rule MA (*mixed associativity*). The modes

[7]By 'base fragment' I mean without quantification such as was discussed in connection with (8).

[8]The meaning constructor given for *not* is of type $(e \to p) \to (e \to p)$ rather than $p \to p$ in order to allow (and in fact, require) a quantifier in subject position to take scope over negation. An alternative way of achieving this aim will be explored in Section 5. Note that we are now able to fix scope level with linear logic quantification, although the IOFU method is also still available.

$$\frac{}{x : A \vdash x : A}\ \text{axiom}$$

For modes $i, j \in \{\upharpoonleft 0, \upharpoonleft 1, \upharpoonleft 2, \upharpoonleft 3, \downharpoonleft 1, \downharpoonleft 2, \downharpoonleft 3\}$:

$$\frac{\Gamma \vdash x : A \quad \Delta \vdash f : A \multimap_i B}{(\Gamma, \Delta)^i \vdash f(x) : B} \multimap_i \mathrm{E} \qquad \frac{(x : A, \Gamma)^i \vdash y : B}{\Gamma \vdash \lambda x.y : A \multimap_i B} \multimap_i \mathrm{I}$$

$$\frac{(\Gamma, \Delta)^i \vdash x : A}{(\Delta, \Gamma)^i \vdash x : A}\ \mathrm{P} \qquad \frac{((\Gamma, \Delta)^i, \Sigma)^j \vdash x : A}{(\Gamma, (\Delta, \Sigma)^j)^i \vdash x : A}\ \mathrm{MA}$$

$\leftarrow$ provided that j does not block i

j blocks $i \Leftrightarrow j = \downharpoonleft m, i = \upharpoonleft n$ and $m > n$

Figure 5: Proposed rules of inference for multi-modal Glue

not Adv
$\%A = (\textsc{adj} \in \uparrow)$
$\lambda P.\lambda x.\mathsf{not}(P(x)) : ((\%A\ \textsc{subj})_e \multimap_i \%A_p) \multimap_{\downharpoonleft 3} ((\%A\ \textsc{subj})_e \multimap_i \%A_p)$

thinks V
$\lambda p.\lambda x.\mathsf{think}(x, p) : (\uparrow \textsc{comp})_p \multimap_{\downharpoonleft 2} ((\uparrow \textsc{subj})_e \multimap_i \uparrow_p)$

ensured V
$\lambda p.\lambda x.\mathsf{ensure}(x, p) : (\uparrow \textsc{comp})_p \multimap_{\downharpoonleft 1} ((\uparrow \textsc{subj})_e \multimap_i \uparrow_p)$

a D
$\lambda P.\lambda Q.\mathsf{some}(P, Q) : (\uparrow_e \multimap \uparrow_p) \multimap \forall X((\uparrow_e \multimap_{\upharpoonleft 3} X_p) \multimap X_p)$

any D
$\lambda P.\lambda Q.\mathsf{some}(P, Q) : (\uparrow_e \multimap \uparrow_p) \multimap \forall X((\uparrow_e \multimap_{\upharpoonleft 2} X_p) \multimap X_p)$

every D
$\lambda P.\lambda Q.\mathsf{every}(P, Q) : (\uparrow_e \multimap \uparrow_p) \multimap \forall X((\uparrow_e \multimap_{\upharpoonleft 1} X_p) \multimap X_p)$

no D
$\lambda P.\lambda Q.\mathsf{not}(\mathsf{some}(P, Q)) : (\uparrow_e \multimap \uparrow_p) \multimap \forall X((\uparrow_e \multimap X_p) \multimap X_p)$

Figure 6: Some partial lexical entries for the fragment

$$
\dfrac{
  \dfrac{
    \dfrac{
      \begin{array}{c}[\text{escaped}] \vdash \\ \mathsf{escape} : \\ \mathbf{3}_e \multimap_{\upharpoonright 1} \mathbf{2}_p\end{array}
      \qquad
      \begin{array}{c}\vdots \\ [\text{every prisoner}] \vdash \\ \lambda P.\mathsf{every}(\mathsf{prisoner}, P) : \\ (\mathbf{3}_e \multimap_{\upharpoonright 1} \mathbf{2}_p) \multimap \mathbf{2}_p\end{array}
    }{
      \begin{array}{c}([\text{escaped}], [\text{every prisoner}]) \vdash \\ \mathsf{every}(\mathsf{prisoner}, \mathsf{escape}) : \mathbf{2}_p\end{array}
    }
    \qquad
    \begin{array}{c}[\text{ensured}] \vdash \\ \lambda p.\lambda x.\mathsf{ensure}(x, p) : \\ \mathbf{2}_p \multimap_{\downharpoonright 1} (\mathbf{1}_e \multimap_{\upharpoonright 3} \mathbf{0}_p)\end{array}
  }{
    \begin{array}{c}(([\text{escaped}], [\text{every prisoner}]), [\text{ensured}])^{\downharpoonright 1} \vdash \\ \lambda x.\mathsf{ensure}(x, \mathsf{every}(\mathsf{prisoner}, \mathsf{escape})) : \mathbf{1}_e \multimap_{\upharpoonright 3} \mathbf{0}_p\end{array}
  }
  \qquad
  \begin{array}{c}\vdots \\ [\text{an accomplice}] \vdash \\ \lambda P.\mathsf{some}(\mathsf{accomplice}, \\ P) : \\ (\mathbf{1}_e \multimap_{\upharpoonright 3} \mathbf{0}_p) \multimap \mathbf{0}_p\end{array}
}{
  \begin{array}{c}((([\text{escaped}], [\text{every prisoner}]), [\text{ensured}])^{\downharpoonright 1}, [\text{an accomplice}]) \vdash \\ \mathsf{some}(\mathsf{accomplice}, \lambda x.\mathsf{ensure}(x, \mathsf{every}(\mathsf{prisoner}, \mathsf{escape}))) : \mathbf{0}_p\end{array}
}
$$

Figure 7: Surface scope interpretation of (13)

interact in the MA rule in such a way that, in combination with the lexicon shown in Figure 6, just the right scope takers are able to escape from just the right islands.

4.1 Multi-modal Glue in action

The time has come to look at some examples. We have the space to go through two: (13), which permits an inverse scope interpretation, and (9), which does not. Given the (simplified) f-structure of (13) shown in (25) and the appropriately instantiated meaning constructors shown in (26), both the surface scope and inverse scope interpretations are available, as shown in Figures 7 and 8 respectively.

(13) An accomplice ensured that every prisoner escaped.

(25)
$$
\boxed{0}\begin{bmatrix}
\text{PRED} & \text{`ensure}\langle\boxed{1},\boxed{2}\rangle\text{'} \\
\text{SUBJ} & \boxed{1}\begin{bmatrix}\text{“an accomplice”}\end{bmatrix} \\
\text{COMP} & \boxed{2}\begin{bmatrix}\text{PRED} & \text{`escape}\langle\boxed{3}\rangle\text{'} \\ \text{SUBJ} & \boxed{3}\begin{bmatrix}\text{“every prisoner”}\end{bmatrix}\end{bmatrix}
\end{bmatrix}
$$

(26) [an accomplice] $:= \lambda P.\mathsf{some}(\mathsf{accomplice}, P) : (\mathbf{1}_e \multimap_{\upharpoonright 3} \mathbf{0}_p) \multimap \mathbf{0}_p$
[ensured] $:= \lambda p.\lambda x.\mathsf{think}(x, p) : \mathbf{2}_p \multimap_{\downharpoonright 2} (\mathbf{1}_e \multimap_{\upharpoonright 3} \mathbf{0}_p)$
[every prisoner] $:= \lambda P.\mathsf{every}(\mathsf{prisoner}, P) : \forall X((\mathbf{3}_e \multimap_{\upharpoonright 1} X_p) \multimap X_p)$
[escaped] $:= \mathsf{escape} : \mathbf{3}_e \multimap_{\upharpoonright 1} \mathbf{2}_p$

The proof in Figure 8 depends on two instances of mixed associativity in order to ‘move’ the variable y to the outside of the premise structure so that it can be abstracted at the step of $\multimap_{\upharpoonright 1}$ introduction. The crucial MA step is the first one, having the schematic form shown in (27).

(27)
$$
\frac{((\Gamma, \Delta)^{\upharpoonright 1}, \Sigma)^{\downharpoonright 1} \vdash \ldots}{(\Gamma, (\Delta, \Sigma)^{\downharpoonright 1})^{\upharpoonright 1} \vdash \ldots}\ \text{MA}
$$

$$
\dfrac{
\dfrac{
\dfrac{
\dfrac{
\dfrac{
\dfrac{y : \mathbf{3}_e \vdash y : \mathbf{3}_e \qquad [\text{escaped}] \vdash \mathsf{escape} : \mathbf{3}_e \multimap_{\uparrow 1} \mathbf{2}_p}
{(y : \mathbf{3}_e, [\text{escaped}])^{\uparrow 1} \vdash \mathsf{escape}(y) : \mathbf{2}_p}
\qquad
[\text{ensured}] \vdash \lambda p.\lambda x.\mathsf{ensure}(x,p) : \mathbf{2}_p \multimap_{\downarrow 1} (\mathbf{1}_e \multimap_{\uparrow 3} \mathbf{0}_p)
}
{((y : \mathbf{3}_e, [\text{escaped}])^{\uparrow 1}, [\text{ensured}])^{\downarrow 1} \vdash \lambda x.\mathsf{ensure}(x, \mathsf{escape}(y)) : \mathbf{1}_e \multimap_{\uparrow 3} \mathbf{0}_p}
}
{(y : \mathbf{3}_e, ([\text{escaped}], [\text{ensured}])^{\downarrow 1})^{\uparrow 1} \vdash \lambda x.\mathsf{ensure}(x, \mathsf{escape}(y)) : \mathbf{1}_e \multimap_{\uparrow 3} \mathbf{0}_p}
\,\text{MA}
\qquad
\begin{array}{c}\vdots\\ [\text{an accomplice}] \vdash \lambda P.\mathsf{some}(\mathsf{accomplice}, P) : (\mathbf{1}_e \multimap_{\uparrow 3} \mathbf{0}_p) \multimap \mathbf{0}_p\end{array}
}
{((y : \mathbf{3}_e, ([\text{escaped}], [\text{ensured}])^{\downarrow 1})^{\uparrow 1}, [\text{an accomplice}]) \vdash \mathsf{some}(\mathsf{accomplice}, \lambda x.\mathsf{ensure}(x, \mathsf{escape}(y))) : \mathbf{0}_p}
}
{(y : \mathbf{3}_e, (([\text{escaped}], [\text{ensured}])^{\downarrow 1}, [\text{an accomplice}]))^{\uparrow 1} \vdash \mathsf{some}(\mathsf{accomplice}, \lambda x.\mathsf{ensure}(x, \mathsf{escape}(y))) : \mathbf{0}_p}
\,\text{MA}
}
{\vdots}
\multimap_{\uparrow 1} \text{I}
$$

$$
\dfrac{
(([\text{escaped}], [\text{ensured}])^{\downarrow 1}, [\text{an accomplice}]) \vdash \lambda y.\mathsf{some}(\mathsf{accomplice}, \lambda x.\mathsf{ensure}(x, \mathsf{escape}(y))) : \mathbf{3}_e \multimap_{\uparrow 1} \mathbf{0}_p
\qquad
\begin{array}{c}\vdots\\ [\text{every prisoner}] \vdash \lambda P.\mathsf{every}(\mathsf{prisoner}, P) : (\mathbf{3}_e \multimap_{\uparrow 1} \mathbf{0}_p) \multimap \mathbf{0}_p\end{array}
}
{((([\text{escaped}], [\text{ensured}])^{\downarrow 1}, [\text{an accomplice}]), [\text{every prisoner}]) \vdash \mathsf{every}(\mathsf{prisoner}, \lambda y.\mathsf{some}(\mathsf{accomplice}, \lambda x.\mathsf{ensure}(x, \mathsf{escape}(y)))) : \mathbf{0}_p}
$$

Figure 8: Inverse scope interpretation of (13)

$$
\dfrac{
  \dfrac{
    \dfrac{y : \mathbf{3}_e \vdash y : \mathbf{3}_e \qquad \begin{array}{c}[\text{escaped}] \vdash \\ \mathsf{escape} : \\ \mathbf{3}_e \multimap_{\uparrow 1} \mathbf{2}_p\end{array}}
    {\begin{array}{c}(y : \mathbf{3}_e, [\text{escaped}])^{\uparrow 1} \vdash \\ \mathsf{escape}(y) : \mathbf{2}_p\end{array}}
    \qquad
    \begin{array}{c}[\text{thinks}] \vdash \\ \lambda p.\lambda x.\mathsf{think}(x, p) : \\ \mathbf{2}_p \multimap_{\downarrow 2} (\mathbf{1}_e \multimap_3 \mathbf{0}_p)\end{array}
  }
  {\begin{array}{c}((y : \mathbf{3}_e, [\text{escaped}])^{\uparrow 1}, [\text{thinks}])^{\downarrow 2} \vdash \\ \lambda x.\mathsf{think}(x, \mathsf{escape}(y)) : \mathbf{1}_e \multimap_3 \mathbf{0}_p\end{array}}
  \qquad
  \begin{array}{c}[\text{a warden}] \vdash \\ \lambda P.\mathsf{some}(\mathsf{warden}, P) : \\ (\mathbf{1}_e \multimap_3 \mathbf{0}_p) \multimap \mathbf{0}_p\end{array}
}
{\begin{array}{c}(((y : \mathbf{3}_e, [\text{escaped}])^{\uparrow 1}, [\text{thinks}])^{\downarrow 2}, [\text{a warden}]) \vdash \\ \mathsf{some}(\mathsf{warden}, \lambda x.\mathsf{think}(x, \mathsf{escape}(y))) : \mathbf{0}_p\end{array}}
$$

Figure 9: Failed attempt to derive an inverse scope interpretation for (9)

This step of MA is licit because $\downarrow 1$, representing the blocking strength of *ensured*, does not block $\uparrow 1$, representing the escaping strength of *every prisoner*. Thus, the inverse scope interpretation is possible.

By contrast, consider (9), with the (simplified) f-structure shown in (10) and the appropriately instantiated meaning constructors shown in (28).

(9) A warden thinks that every prisoner escaped.

(10)
$$
\boxed{0}\left[\begin{array}{ll}
\text{PRED} & \text{`think}\langle \boxed{1}, \boxed{2} \rangle\text{'} \\
\text{SUBJ} & \boxed{1}\left[\text{“a warden”}\right] \\
\text{COMP} & \boxed{2}\left[\begin{array}{ll}
\text{PRED} & \text{`escape}\langle \boxed{3} \rangle\text{'} \\
\text{SUBJ} & \boxed{3}\left[\text{“every prisoner”}\right]
\end{array}\right]
\end{array}\right]
$$

(28) $[\text{a warden}] := \lambda P.\mathsf{some}(\mathsf{warden}, P) : (\mathbf{1}_e \multimap_{\uparrow 3} \mathbf{0}_p) \multimap \mathbf{0}_p$

$[\text{thinks}] := \lambda p.\lambda x.\mathsf{think}(x, p) : \mathbf{2}_p \multimap_{\downarrow 2} (\mathbf{1}_e \multimap_{\uparrow 3} \mathbf{0}_p)$

$[\text{every prisoner}] := \lambda P.\mathsf{every}(\mathsf{prisoner}, P) : \forall X.(\mathbf{3}_e \multimap_{\uparrow 1} X_p) \multimap X_p$

$[\text{escaped}] := \mathsf{escape} : \mathbf{3}_e \multimap_{\uparrow 1} \mathbf{2}_p$

The surface scope interpretation is derived in an entirely analogous manner to the surface scope interpretation of (13) as shown in Figure 7, with $\multimap$ elimination the only rule of inference used. In order to derive an inverse scope interpretation, one would have to proceed as shown in Figure 9, introducing an auxiliary assumption early in order to be abstracted later. However, if you do that then at some point the derivation cannot proceed, as shown. In order to get an inverse scope interpretation, y would have to be moved to the outside of the premise structure so that it can be abstracted. But this is not possible because the relevant portion of the structure has the schematic form shown in (29).

(29) $((\Gamma, \Delta)^{\uparrow 1}, \Sigma)^{\downarrow 2}$

Mixed associativity cannot apply because $\downarrow 2$, the blocking strength of *thinks*, blocks $\uparrow 1$, the escaping strength of *every prisoner*. Thus, no inverse scope interpretation is possible.

It should be clear from these examples how the rules round out the SISC: neither *any N*, *every N* nor *no N* can take scope over negation because MA cannot apply to a structure of the form shown in (30), and in fact *no N* cannot take scope from out of the complement of *think* or *ensure* either because MA cannot apply to a structure of the form shown in (31).

(30) $((\Gamma, \Delta)^{\uparrow 0/1/2}, \Sigma)^{\downarrow 3}$

(31) $((\Gamma, \Delta), \Sigma)^{\downarrow 1/2}$

4.2 Reflections

The choice of available modes in this multi-modal Glue system, and the way they interact in the MA rule, are obviously ad-hoc to an extent. As with the SCOPEISLAND features considered in Section 3, these modes can be viewed as placeholders for whatever the comparative strengths of various scope island inducers and escapers turn out to be. I intend to leave open the possibility, for example, that there could be an island inducer stronger than *ensure* but weaker than *think*, or an island escaper stronger than *every N* but weaker than *any N*; I am also open to the possibility that these modes could be predictable from some syntactic or semantic feature.[9]

As a choice of *formal system*, however, multi-modal Glue has one major advantage: given the MA rule and a natural order on the modes (here represented by $<$), the scope island subset constraint follows automatically. Unlike in the system outlined in Section 3, there is no way to give a lexical entry like (22) for a quantifier which can take scope out of the complement of *thinks* but not out of the complement of *ensured*, for example.

On the other hand, it does complicate the underlying logic considerably to move to a multi-modal system, whereas the blocking features-based approach only makes use of established LFG+Glue technology. In the following section we will look at some potential additional motivations for adopting a multi-modal approach.

5 Possible extensions

As alluded to in Section 3, it is an open question what the connection is between the explanations for scope islands and scope freezing. As an example of the latter in English, consider (32), which has a surface scope interpretation but not an inverse scope interpretation, as shown.

(32) Every warden checked no prisoner(s).
$\Rightarrow$ $\mathsf{every}(\mathsf{warden}, \lambda x.\mathsf{not}(\mathsf{some}(\mathsf{prisoner}, \lambda y.\mathsf{check}(x, y))))$

[9] My thanks to a reviewer for pressing this point.

$$\dfrac{\dfrac{\begin{array}{c} y : \mathbf{2}_e \vdash \\ y : \mathbf{2}_e \end{array} \qquad \begin{array}{c} [\text{checked}] \vdash \\ \lambda v.\lambda u.\mathsf{check}(u,v) : \mathbf{2}_e \multimap (\mathbf{1}_e \multimap_{\uparrow 1} \mathbf{0}_p) \end{array}}{\begin{array}{c} (y : \mathbf{2}_e, [\text{checked}]) \vdash \\ \lambda u.\mathsf{check}(u,y) : \mathbf{1}_e \multimap_{\uparrow 1} \mathbf{0}_p \end{array}} \qquad \begin{array}{c} \vdots \\ [\text{every warden}] \vdash \\ \lambda P.\mathsf{every}(\mathsf{warden}, P) : \\ (\mathbf{1}_e \multimap_{\uparrow 1} \mathbf{0}_p) \multimap_{\downarrow 1} \mathbf{0}_p \end{array}}{\begin{array}{c} ((y : \mathbf{2}_e, [\text{checked}]), [\text{every warden}])^{\downarrow 1} \vdash \\ \mathsf{every}(\mathsf{warden}, \lambda u.\mathsf{check}(u,y)) : \mathbf{0}_p \end{array}}$$

Figure 10: Failed attempt to derive an inverse scope interpretation of (32)

$$\not\Rightarrow \quad \mathsf{not}(\mathsf{some}(\mathsf{prisoner}, \lambda y.\mathsf{every}(\mathsf{warden}, \lambda x.\mathsf{check}(x,y))))$$

Because there is no embedded clausal f-structure in the f-structure of (32), shown in (33), there is no choice of scope level and hence no way to account for scope freezing in the blocking features-based approach. Both *every warden* and *no prisoner* have to take $\boxed{0}$ as their scope level.

(33) $\boxed{0}\left[\begin{array}{ll} \text{PRED} & \text{‘check}\langle\boxed{1},\boxed{2}\rangle\text{’} \\ \text{SUBJ} & \boxed{1}\left[\text{“every warden”}\right] \\ \text{OBJ} & \boxed{2}\left[\text{“no prisoner”}\right] \end{array}\right]$

In Gotham (2019) I proposed an account of scope freezing in Glue but, as I mentioned in Section 3, it requires yet another complication of the linear logic fragment used, of a different kind to that discussed in Section 4. Perhaps more seriously, it is not ideally suited to the kind of quantifier-determined scope rigidity exhibited by (32). What I mean by that is that it is not the case in general that direct objects cannot scope over subjects in English—unlike in e.g. German main clauses with canonical SVO order, which is more the point of Gotham (2019). Rather, it seems to be the case that downward-monotonic objects (such as *no N*) cannot scope over upward-monotonic subjects (such as *every N*).

Multi-modal Glue suggests a way we could approach this issue. Look at the proposed meaning constructors for *every* and *no* below, and compare them with those given in Figure 6.

$$\textit{every} \rightsquigarrow \lambda P.\lambda Q.\mathsf{every}(P,Q) : (\uparrow_e \multimap \uparrow_p) \multimap \forall X((\uparrow_e \multimap_{\uparrow 1} X_p) \multimap_{\downarrow 1} X_p)$$
$$\textit{no} \rightsquigarrow \lambda P.\lambda Q.\mathsf{not}(\mathsf{some}(P,Q)) : (\uparrow_e \multimap \uparrow_p) \multimap \forall X((\uparrow_e \multimap X_p) \multimap X_p)$$

In the meaning constructor shown above, *every* has been given a blocking mode index on its final $\multimap$. This makes an inverse scope interpretation of (32) unavailable, as shown by the failed attempt to derive one in Figure 10. A premise structure of the general form shown in (34) is created, meaning that MA cannot apply.

(34) $((\Gamma, \Delta), \Sigma)^{\downarrow 1}$

However, this strategy for explaining the non-ambiguity of (32)—of effectively making *every N* induce a scope island from which *no N* cannot escape—quickly runs into problems. The same structure as (34) would be created in any attempt to derive a *surface* scope interpretation of (35), for example.

(35) No warden checked every prisoner.

The modes in our fragment effectively have two parameters:[10] $\upharpoonright$ vs. $\downharpoonleft$ to express blocking vs. escaping, and 0–3 to express strength thereof. In order to differentiate between (32) and (35), and allow the *no* $>$ *every* scope order in the latter but not the former, there would need to be an additional parameter keeping track of some relevant property, presumably either argument structure, linear order or c-structure embeddedness.

5.1 Extending the fragment further

Suppose that we have the linear logic fragment as defined in Figure 5, except that we have the expanded list of modes shown in (36), and the definition of blocking for the MA rule is as shown in (37).[11] Once again, the use of i or j in a mode index means that choice of parameter is free.

(36) $a\upharpoonright 0, b\upharpoonright 0, c\upharpoonright 0, d\upharpoonright 0, a\upharpoonright 1, b\upharpoonright 1, c\upharpoonright 1, d\upharpoonright 1, a\upharpoonright 2, b\upharpoonright 2, c\upharpoonright 2, d\upharpoonright 2, a\upharpoonright 3, b\upharpoonright 3, c\upharpoonright 3, d\upharpoonright 3,$
$a\downharpoonleft 1, b\downharpoonleft 1, c\downharpoonleft 1, d\downharpoonleft 1, a\downharpoonleft 2, b\downharpoonleft 2, c\downharpoonleft 2, d\downharpoonleft 2, a\downharpoonleft 3, b\downharpoonleft 3, c\downharpoonleft 3, d\downharpoonleft 3$

(37) j blocks $i \Leftrightarrow j = x\downharpoonleft m, i = y\upharpoonright n, m > n$ and $x < y$.

The idea is to use $a/b/c/d$ to encode in the lexical entry for a verb the relative prominence of its arguments. The definition in (37) stipulates that for blocking to occur, the prospective blocker must (in the relevant sense) outrank the relevant escaper on both the alphabetical and numerical parameters. That means we can account for the contrast between (32) and (35) by means of the lexicon shown in Figure 11.[12,13]

No N then can take scope over *every N* when that corresponds to a surface scope interpretation, e.g. of (35). The crucial inferential step is as shown in (38).

[10]This talk of parameters need not be taken literally. In reality, the modes can be simple, with a blocking order defined on them directly. But the notational use of parameters helps with exposition.

[11]For the purposes of (37), $a < b < c < d$, since alphabetical order is 'ascending'.

[12]To retain the account of scope islands from Section 4, it follows that the complement argument must be given an alphabetical parameter that outranks every other, as shown in the lexical entry for *thinks* in Figure 11. Therefore, this is not quite the same notion of syntactic rank as expressed in LFG binding theory, although perhaps the definitions could be changed to bring the two notions into line.

[13]The lexical entry given for *not* in Figure 11 is now of type $p \multimap p$, and permits, but does not require, a quantifier in subject position to take scope over negation. This can be seen as an improvement over the treatment of negation given in Figure 6.

checked V
$\lambda y.\lambda x.\mathsf{check}(x,y) : (\uparrow \text{OBJ})_e \multimap_{c\upharpoonright i} ((\uparrow \text{SUBJ})_e \multimap_{b\upharpoonright j} \uparrow_p)$

every D
$\lambda P.\lambda Q.\mathsf{every}(P,Q) : (\uparrow_e \multimap \uparrow_p) \multimap \forall X((\uparrow_e \multimap_{i\upharpoonright 1} X_p) \multimap_{i\downharpoonright 1} X_p)$

no D
$\lambda P.\lambda Q.\mathsf{not}(\mathsf{some}(P,Q)) : (\uparrow_e \multimap \uparrow_p) \multimap \forall X((\uparrow_e \multimap_{i\upharpoonright 0} X_p) \multimap_{i\downharpoonright 1} X_p)$

warden N
$\mathsf{warden} : \uparrow_e \multimap \uparrow_p$

prisoner N
$\mathsf{prisoner} : \uparrow_e \multimap \uparrow_p$

thinks V
$\lambda p.\lambda x.\mathsf{think}(x,p) : (\uparrow \text{COMP})_p \multimap_{a\downharpoonright 2} ((\uparrow \text{SUBJ})_e \multimap_{b\upharpoonright i} \uparrow_p)$

not Adv
$\mathsf{not} : (\text{ADJ} \in \uparrow)_p \multimap_{b\downharpoonright 3} (\text{ADJ} \in \uparrow)_p$

Figure 11: Partial lexical entries for scope freezing

(38)
$$\frac{((x : \mathbf{1}_e, [\text{checked}])^{b\upharpoonright 0}, [\text{every prisoner}])^{c\downharpoonright 1} \vdash \mathsf{every}(\mathsf{prisoner}, \lambda y.\mathsf{check}(x,y)) : \mathbf{0}_p}{(x : \mathbf{1}_e, ([\text{checked}], [\text{every prisoner}])^{c\downharpoonright 1})^{b\upharpoonright 0} \vdash \mathsf{every}(\mathsf{prisoner}, \lambda y.\mathsf{check}(x,y)) : \mathbf{0}_p}\ \text{MA}$$

Mixed associativity is applicable in (38) because $c \not< b$ and so blocking does not occur according to the revised definition in (37). By contrast, if we attempt to derive an inverse scope interpretation of (32) then we end up with a premise structure of the form shown in (39), to which MA cannot apply.

(39)
$$((x : \mathbf{1}_e, [\text{checked}])^{c\upharpoonright 0}, [\text{every prisoner}])^{b\downharpoonright 1} \vdash \mathsf{every}(\mathsf{prisoner}, \lambda y.\mathsf{check}(x,y)) : \mathbf{0}_p$$

Since *every N* outranks *no N* both according to strength ($1 > 0$) and, in this case, according to argument position ($b < c$), MA is blocked.

6 Discussion

The previous section has shown that it is at least feasible for scope islands and scope freezing to both be accounted for using the same formal tools, but it remains to be seen whether or not this is the best approach. It also remains to be seen what the predictions of any specific implementation of this idea might be. For example, languages that are scope rigid in the sense that inverse scope interpretations are disallowed in general, and not just based on the particular quantifiers involved, could be accommodated within the particular formulation of the modes and structural rules given in (36)–(37) by assigning to every quantifier a meaning constructor of the general form shown in (40).

(40) $\mathsf{quant} : \forall X((\uparrow_e \multimap_{i|0} X_p) \multimap_{j|1} X_p)$

This will ensure that the blocking strength of any quantifier will always be greater than the escaping strength of any other, meaning that the argument position parameter alone will be decisive. But it remains to be seen what the implications of this assumption would be for the kind of extra-clausal scope interactions that our discussion began with ('scope islands'), and whether or not they are borne out.[14] For learnability reasons, our default assumption really ought to be that every language uses the same fragment of linear logic for meaning composition, which makes it crucial to push these kinds of questions early on if we decide that multi-modal Glue is the way to go.

As became progressively clear from Sections 4–5, analysing the data using multi-modal Glue involves incorporating a significant amount of syntactic information into meaning constructors, to the point where many LFG practitioners may feel that we are doing too much categorial grammar within LFG. What I would say to that is that the data force us to do so to some degree, because in Glue meaning constructors are standardly defined based on either f- or s-structure: levels at which certain properties that seem to be crucial for scope possibilities are not defined.

It is still an open question what the best way to account for scope islands in LFG+Glue is, partly because the empirical landscape is not entirely clear, despite decades of work from researchers working in a variety of frameworks. That said, it seems highly likely that the proper explanation for at least some forms of scope rigidity will require a complication of the fragment of linear logic used in Glue beyond simply **LP** (with or without quantification to fix scope level), for the reasons discussed in Section 3 and in Gotham (2019). Just what form that complication should take, though, and what data it should cover, are also open questions.

References

Andrews, Avery D. 2010. Propositional glue and the projection architecture of LFG. *Linguistics and Philosophy* 33, 141–170.

Barker, Chris. 2021. Rethinking Scope Islands. *Linguistic Inquiry* Advance publication.

Charlow, Simon. 2014. *On the semantics of exceptional scope*. Ph. D.thesis, New York University.

[14] There is also the issue of how argument position, which this paper has focussed on, conspires with other factors, such as word order and c-structure prominence, to determine what constitutes 'inverse scope'. In (Gotham 2019, §4.1) this issue was addressed for German by means of a template allowing particular word order configurations to 'reset' the scope constraints. Such an approach could be added to multi-modal Glue as well. My thanks to a reviewer for picking up on this.

Dalrymple, Mary, Lowe, John J. and Mycock, Louise. 2019. *The Oxford Reference Guide to Lexical Functional Grammar*. Oxford: Oxford University Press.

Dalrymple, Mary, Patejuk, Agnieszka and Zymla, Mark-Matthias. 2020. XLE+Glue – A new tool for integrating semantic analysis in XLE. In Miriam Butt and Ida Toivonen (eds.), *Proceedings of the LFG'20 Conference, On-Line*, pages 89–108, Stanford, CA: CSLI Publications.

Fry, John. 1999. Proof Nets and Negative Polarity Licensing. In Mary Dalrymple (ed.), *Semantics and Syntax in Lexical Functional Grammar*, pages 91–116, Cambridge, MA: MIT Press.

Gotham, Matthew. 2019. Constraining Scope Ambiguity in LFG+Glue. In Miriam Butt, Tracy Holloway King and Ida Toivonen (eds.), *Proceedings of the LFG'19 Conference, Australian National University*, pages 111–129, Stanford, CA: CSLI Publications.

Gotham, Matthew and Haug, Dag Trygve Truslew. 2018. Glue semantics for Universal Dependencies. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG'18 Conference, University of Vienna*, pages 208–226, Stanford, CA: CSLI Publications.

Kokkonidis, Miltiadis. 2008. First-Order Glue. *Journal of Logic, Language and Information* 17(1), 43–68.

Lev, Iddo. 2007. *Packed Computation of Exact Meaning Representations*. Ph. D.thesis, Stanford University.

May, Robert. 1977. *The Grammar of Quantification*. Ph. D.thesis, Massachusetts Institute of Technology.

Moot, Richard and Retoré, Christian. 2012. *The Logic of Categorial Grammars*. Lecture Notes in Computer Science, No. 6850, Berlin/Heidelberg: Springer.

Rosén, Victoria, De Smedt, Koenraad, Meurer, Paul and Dyvik, Helge. 2012. An open infrastructure for advanced treebanking. In Jan Hajič, Koenraad De Smedt, Marko Tadić and António Branco (eds.), *META-RESEARCH Workshop on Advanced Treebanking at LREC2012*, pages 22–29, Istanbul.

The ‘productive’ vs. ‘thematic’ prefix distinction in Tetsǫ́t’ıné: an LFG formalization

Alessandro Jaker
University of Toronto

Proceedings of the LFG’21 Conference

On-Line

Miriam Butt, Jamie Y. Findlay, Ida Toivonen (Editors)

2021

CSLI Publications

pages 167–180

http://csli-publications.stanford.edu/LFG/2021

Keywords: Dene, Athabaskan, D-mapping theory, m-structure, OT-LFG

Jaker, Alessandro. 2021. The ‘productive’ vs. ‘thematic’ prefix distinction in Tetsǫ́t’ıné: an LFG formalization. In Butt, Miriam, Findlay, Jamie Y., & Toivonen, Ida (Eds.), *Proceedings of the LFG’21 Conference, On-Line*, 167–180. Stanford, CA: CSLI Publications.

Abstract

In Tetsǫ́t'ıné (ethnologue: CHP), the causative prefix *ł* has both productive and thematic uses. When *ł* is used productively, it adds an argument to the PRED it modifies, and also participates in selection and blocking relations with other prefixes. When *ł* is used thematically—that is, as part of the basic lexical entry of a verb—it appears to be semantically empty, and yet its selection and blocking properties are retained. This paper proposes a unified treatment of both occurrences of *ł*, using *D*-mapping theory (Dalrymple 2015). The *D*-mapping function, by which changes in the f-structure/a-structure mapping are projected from m-structure, is formulated as a violable constraint in OT-LFG. The result is that when *ł* is compatible with the argument structure of the PRED, as in its productive uses, the output of the *D*-mapping function is realized, whereas when *ł* is incompatible with the argument structure of the PRED, as in its thematic uses, *ł* is bleached of its semantic content.

1 Introduction

Tetsǫ́t'ıné is a dialect of Dëne Sųłıné (ethnologue: CHP) spoken in Canada's Northwest Territories. It belongs to the Dene (Athapaskan) language family. In the Dene linguistics literature, a distinction is often made between 'productive' and 'thematic' uses of the same prefixes (Rice 2000: 126-170). Briefly, when a prefix is used productively, it contributes to the semantics and morphosyntactic representation of the verb, and also engages in selection and blocking relationships with other prefixes. When a prefix is used thematically, however—that is, as part of a larger morphological construction—it appears to be semantically empty, and yet its selection and blocking properties are retained.

This paper will focus on a single prefix, the causative voice/valence marker *ł*, which can be used either thematically or productively in Tetsǫ́t'ıné. I will propose a single representation which accounts for both productive and thematic uses of this prefix, and I will propose a mechanism by which this prefix is semantically bleached in its thematic uses. My analysis will rely crucially upon the distinction, made possible in LFG, between f-structure, the level at which morphosyntactic features are realized (Bresnan 2001, Dalrymple 2001), and m-structure, the level at which morphological selection and blocking restrictions are stated (Frank & Zaenen 2004). Data are taken from Jaker & Cardinal's (2020) *Tetsǫ́t'ıné Verb Grammar* (TVG), unless otherwise specified.

1.1 Productive vs. thematic uses of the prefix *ł*

The prefix *ł* is a causative voice/valence prefix which adds an argument to the verb stem it modifies. It is one of three voice/valence prefixes in Dene languages; the others are *d* 'middle voice' and *l* 'causative middle' (Rice 2000: 126-170). Some surface verb forms do not have an overt voice/valence prefix, and such forms are described as 'Ø classifier' in the Dene linguistics literature. These verbs may be either transitive or intransitive. The prefixes *d* and *ł* are productive in that, in my experience, *d* can be added to any transitive verb as part of the reflexive construction, while *ł* can be added to any intransitive verb to make it transitive, provided that the lexical semantics of the verb are compatible. Where the prefixes *d, l,* and *ł* do surface, they always occur immediately preceding the verbal root. Some examples of the prefix *ł* used productively are given in (1), where we can contrast the intransitive verbs in (1.1) (without *ł*) with their corresponding transitive verbs in (1.2) (with *ł*). The subscript numbers in the underlying forms refer to template position numbers (to be explained in §1.2).

(1) Examples of *ł* prefix used productively (changes argument structure)

(1.1) Intransitive verbs, without *ł*

Tetsǫ́t'ıné	**English gloss**
a. /łaH$_{0}$-ñe$_{10}$-ñe$_{11}$-dhër/ → łaįdhër	'he/she/it died'
b. /ne$_{8}$-ye/ → neye	'he/she/it grows'
c. /bes/ → hebes	'it is boiling'
d. /t'éth/ → het'éth	'it is cooking'

(1.2) Transitive verbs, with *ł*

Tetsǫ́t'ıné	**English gloss**
a. /łaH$_{0}$-ñe$_{10}$-ñe$_{11}$-**ł**$_{13}$-dhër/ → łaį**ł**thër	'he/she killed (O)'
b. /ne$_{8}$-**ł**$_{13}$-ye/ → ne**ł**she	'he/she grows (O)'
c. /ł-bes/ → he**ł**bes	'he/she is boiling (O)'
d. /ł-t'éth/ → he**ł**t'éth	'he/she is cooking (O)'

ref: TVG §4.5.1, 5.2.3, author's fieldnotes

In (1), the causative prefix *ł* is added to all of the intransitive forms in (1.1), to generate the corresponding transitive forms in (1.2). The function of *ł* is not always so transparent, however. Indeed, in many cases, this prefix seems to be synchronically meaningless. Consider the examples in (2).

(2) Examples of *ł*-classifier used thematically (i.e. semantically empty)

Underlying form	Surface form	English gloss
a. /ya$_{4}$-**ł**$_{13}$-tı/	ya**ł**tı	‘he/she speaks’
b. /the$_{10}$-**ł**$_{13}$-tą/	the**ł**tą	‘a round container filled with liquid is sitting’
c. /the$_{10}$-**ł**$_{13}$-chúth/	the**ł**chúth	‘a clothlike object is sitting’
d. /the$_{10}$-**ł**$_{13}$-tsį/	the**ł**tsį	‘he/she made (O)’
e. /ná$_{1}$-the$_{10}$-**ł**$_{13}$-t’us/	náthe**ł**t’us	‘he/she punched (O)’
f. /ná$_{1}$-the$_{10}$-**ł**$_{13}$-tthel/	náthe**ł**tthel	‘he/she chopped (O)’

ref: TVG § 6.6.2, 8.2, 8.7.

All of the verbs in (2) exhibit what appears to be causative morphology; however, in none of these examples does there exist an independent morphological base form from which these morphological causatives are derived. Indeed, in many cases—such as with the verbs meaning ‘speak’ and ‘sit’—it is difficult to imagine how these verbs could be derived from a more basic verb with one less argument. In these cases, *ł* is part of the basic lexical entry of these verbs, which in the Dene linguistics literature is called the VERB THEME (see §2). For this reason, the *ł* classifier is said to be THEMATIC in examples such as in (2).

1.2 Morpheme identity and template position

Given that the prefix *ł* sometimes clearly functions as a causative prefix, as in (1) and is sometimes semantically meaningless, as in (2), the question arises as to whether these are both instances of the same prefix, or rather two different (but homophonous) prefixes. In my opinion, there are two arguments as to why these are indeed the same prefix: template position and selectional restrictions. In this section (§1.2) I will discuss template position, while in the next section (§1.3) I will discuss selectional restrictions.

Dene languages are traditionally described as templatic languages. A template is an abstract set of positions or ‘slots’. Under the template model, every prefix contains, as part of its lexical entry, a position number, which assigns it a position within the template (Rice 2000: 9; Jaker, Welch & Rice 2020). The template for Tetsǫ́t’ıné consists of 13 template positions as shown in (3) below.

(3) Tetsǫ́t'ıné verbal template (TVG: 35)

Name	Preverb	Distributive	Iterative	Incorporate	Areal	Object	Deictic Subject	Qualifier	Aspect	Conjugation	Mode	Subject	Classifier	Stem
Position #	1	2	3	4	5	6	7	8	9	10	11	12	13	
Examples	ná	dá	na	shé	ha	se	he	ne	te	ghe	ñe	s	d	t'éth
	xá			ya	ho	ne	ts'e	de	ne	the	ghu	ı	l	k'éth
	xa			gór		ye				hí		ñe	ł	tthél
	ní					nuhe				ñe		híd		tsıl
	nı					hube						uh		tı
						ʔe								zé
						ʔede								ya
						ʔełe								

In the template in (3), all three voice/valence markers *d*, *l*, and *ł* (called 'classifiers' in the Dene linguistics literature) occur in position 13, immediately preceding the stem. The template model thus predicts that no other prefixes can intervene between the voice/valence markers and the stem. Accordingly, note that in the examples where *ł* used productively in (1.2), as well as the examples where *ł* is used thematically in (2), it always occurs immediately preceding the verb stem. Thus, one argument that both productive and thematic uses of *ł* are instances of the same prefix is that they occur in the same linear position.

1.3 Selectional properties are unchanged

Tetsǫ́t'ıné morphology contains numerous discontinuous dependencies, across different template positions, which take the form of selectional and blocking restrictions (TVG: 33-64). In this section, we will discuss one particular set of selectional relations: the relation between the voice/valence marker (in position 13) and the choice of perfective marker (in position 11). Briefly, when the voice/valence marker is absent (so-called 'zero-classifier verbs'), or when the voice/valence marker is *ł*, the perfective prefix is /ñe/. Due to the morphophonemic rules of the language, this usually results in the front high nasal vowel į on the surface. On the other hand, when the voice/valence marker is either *d* ('middle voice') or *l* ('causative middle'), /Ø/, a zero allomorph of the perfective marker occurs instead (TVG: 39-40). This is illustrated in (4).

(4) Voice/valence prefixes select perfective allomorph (based on Jaker 2014)

ñe **Ø/ł**

conjugation$_{10}$ – mode$_{11}$ – subject$_{12}$ – voice/valence$_{13}$ – root

Ø **d/l**

Rice (2000: 169), following earlier work by Hopper & Thompson (1980) suggests that this pattern may be due to a restriction on overtly marking perfectivity in the middle voice. For the purposes of this paper, what is important to note is that, for all three of the voice/valence markers (plus ‘zero’), their selectional properties are unchanged whether they are used productively or thematically. This is illustrated in (5)-(7) below.

(5) *ł* classifier selects *ñe* perfective when used productively

Underlying form	**Surface form**	**English gloss**
a. /łaH$_{0}$-ñe$_{10}$-**ñe$_{11}$**-ł$_{13}$-thër/	ła**į**łthër	‘he/she killed (O)’
b. /łaH$_{0}$-he$_{7}$-ñe$_{10}$-**ñe$_{11}$**-ł$_{13}$-thër/	łáh**įį**łthër	‘they killed (O)’

ref: TVG §6.5.4

(6) *ł* classifier selects *ñe* perfective when used thematically

Underlying form	**Surface form**	**English gloss**
a. /ya$_{4}$-ghe$_{10}$-**ñe$_{11}$**-ł$_{13}$-tı/	ya**į**łtı	‘he/she spoke’
b. /ya$_{4}$-he$_{7}$-ghe$_{10}$-**ñe$_{11}$**-ł$_{13}$-tı/	yah**įį**łtı	‘they spoke’

ref: TVG §4.7.1

(7) *d* and *l* classifiers select Ø perfective (used thematically)

Underlying form	**Surface form**	**English gloss**
/shé$_{4}$-he$_{7}$-ghe$_{10}$-**Ø$_{11}$**-d$_{13}$-tį/	shéh**ee**tį	‘they (2) ate’
/se$_{6}$-he$_{7}$-ghe$_{10}$-**Ø$_{11}$**-l$_{13}$-ts’ün/	seh**ee**lts’ün	‘they kissed me’

ref: TVG §6.3.2, 6.3.3

In both lexical-incremental as well as realizational theories of morphology, it is problematic that a semantically empty prefix should be able to select or block other prefixes. This is because the prefix which does the selecting presumably does so by virtue of the inflectional features which it contributes or expresses, respectively. The fact that semantically empty prefixes can have selectional properties, therefore, suggests that selectional restrictions ought to be stated at a different level of representation than the level at which morphosyntactic features are encoded. For present purposes, however, it is sufficient to note that the *ł* voice/valence marker has the same selectional properties whether it is used productively, as in (5), or thematically, as in (6). The fact that selectional properties are unchanged whether *ł* is used productively or thematically thus provides a second argument that, in both cases, we are dealing with the same prefix.

1.4 Overview of proposal

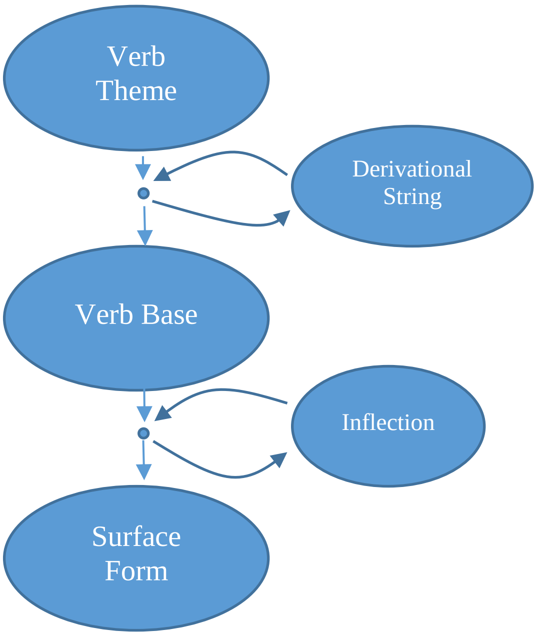
If both thematic and productive uses of *ł* are instances of the same prefix, we are faced with the following basic problem: how can the same prefix sometimes change the argument structure of the verb, and sometimes be semantically empty? I propose that LFG provides a set of formal tools with which to address this problem, by distinguishing two levels of representation: f-structure (Bresnan 2001, Dalrymple 2001), where morphosyntactic features are encoded, and m-structure (Frank & Zaenen 2004), where morphological selectional and blocking restrictions can be stated. In the remainder of this paper, I will assume a morpheme-based or 'lexical-incremental' model of morphology, since I believe that the issues can be described most transparently in such a framework. Specifically, I will claim that the *ł* voice/valence marker has a single lexical entry, whether it is used productively or thematically. However, when the *ł* prefix is part of a larger morphological construction, such as a verb theme or derivational string (see §2), sometimes a clash of features arises at the level of the f-structure/a-structure mapping. When the argument structure projected by the *ł* prefix is in conflict with the argument structure projected by the PRED, the *ł* prefix is bleached of its semantic content. This process of semantic bleaching is formalized in OT-LFG.

2 Interrupted synthesis and word formation

As mentioned earlier, when the prefix *ł* is used thematically—as part of the basic lexical entry of the verb—it is almost always used as part of a larger morphological construction called the VERB THEME (hence the term 'thematic'). In this section, I will provide background on the three main constituent parts involved in Dene word formation: the verb theme, DERIVATIONAL STRING, and INFLECTION.

According to the traditional model of Dene word formation (Whorf 1932; Kari 1979, 1989), which I will call 'Interrupted Synthesis', word formation consists of the recursive interfixation of discontinous strings into other discontinuous strings. Word formation begins with the verb theme, which constitutes the basic lexical entry of the verb. Verb themes always contain a verbal root, and frequently contain an adverbial prefix and voice/valence marker as well. In the next stage of word formation, a derivational string is added to the verb theme, to make the VERB BASE. Derivational strings can be aspectual or non-aspectual (Kari 1979, 1989); often, derivational strings will consist of an adverbial prefix plus a conjugation marker, although other combinations of prefixes as possible. Finally, inflectional prefixes, including subject and object agreement, are added to the verb base, to make a SURFACE FORM. A flow chart illustrating the process of word formation, under this model, is given in (8), while some Tetsǫ́t'ıné examples illustrating the terminative derivational string (which means 'stop doing X') are given in (9).

(8) Interrupted Synthesis model (simplified), based on Kari (1992: 111)



(9) Illustration of the terminative derivational string (TVG: 128-139)

(9)(a)

Inflection **Verb Theme**

ní$_1$ ya$_4$ he$_7$ ñe$_{10}$ ñe$_{11}$ ł$_{13}$ tı

Derivational String

níyahı̨ı̨łtı 'they stopped speaking'

(9)(b)

Inflection **Verb Theme**

ní$_1$ she$_4$ he$_7$ ñe$_{10}$ Ø$_{11}$ d$_{13}$ tı

Derivational String

níshehı̨ı̨tı̨ 'they (DU) stopped eating'

In (9), we see two verb themes /ya$_4$…ł$_{13}$-tı/ ‘speak’ and /shé$_4$…d$_{13}$-tį̨/ ‘eat’, which carry the main lexical meanings of these verbs. To both of these verbs is added the TERMINATIVE derivational string /ní$_1$…ñe$_{10}$/ ‘stop doing X’. Finally, inflectional prefixes are added, such as /he$_7$/ ‘3plS’, and /ñe$_{11}$/ or /Ø$_{11}$/ ‘PERF’. The main point of (9) is that verb themes and derivational strings are discontinuous within the word, but must nevertheless be treated as morphological ‘constructions’ in some sense, which are more than the sum of their parts, in terms of their semantic content.

To summarize, under this model it is assumed that word formation begins with the verb theme, to which derivational strings are added to make the verb base, to which finally inflectional prefixes are added. This is relevant in that the behavior of *ł* may be correlated with the stage of word formation at which it is added. When the prefix *ł* is semantically empty—that is, ‘thematic’—that is because it belongs to the verb theme, which is the basic lexical entry of the verb. When, on the other hand, *ł* used productively, it is added at a later stage of word formation. Therefore, when the thematic use of *ł* leads to a clash of features at the level of f-structure, this clash of features arises within the verb theme itself, as we will see in the following sections.

3 An LFG formalization using *D*-mapping

As mentioned earlier, my analysis will rely crucially upon the distinction, available in LFG, between f-structure (Bresnan 2001, Dalrymple 2001), the level at which morphosyntactic features are expressed, and m-structure (Frank & Zaenen 2004), the level at which morphological selectional and blocking restrictions are stated. Following Dalrymple (2015), I will assume that f-structure is projected from m-structure via the *D*-mapping function. This means that in most cases, as in (10) below, it is not necessary to specify f-descriptions as part of the lexical entry of prefixes. Rather, the f-description can be projected by *D*-rules. Based on information specified in the lexical entry of the *ł* voice/valence marker, an m-structure is projected, as illustrated in (10). In (10), I have labeled the m-structure attribute for *ł* ‘VOICE’, although here this term is used in a broad sense, in that the voice/valence prefixes actually contribute a combination of information about both voice and valence.

(10)(a) Lexical entry for *ł*-classifier

m-str: ($\downarrow_{\mu}$ VOICE) = ł

p-form: /ł/, V_{Prefix}, Level 1, Position 13

(10)(b) m-structure

μ [VOICE ł]

Based on the information contained in the m-structure, the m-structure will project an f-structure via the *D*-mapping function. More precisely, in this particular case, I assume that the *ł* classifier introduces changes at the level of f-structure/a-structure mapping (Dione 2013), as shown in (11).

(11) *D*-mapping rule for causatives

M-VOICE: ł $\Rightarrow$ {'caus < ARG, 'pred < ARG >' >'

AG PT/THM

The rule in (11) introduces an argument which is an agent, and also requires that the internal argument of the PRED be a patient or theme. I assume that the former will be interpreted as a subject and the latter as an object according to Lexical Mapping Theory (LMT) (Bresnan & Zaenen 1990). Crucially, this means that the rule in (11) will be compatible with the lexical entry of unaccusative verbs such as 'die' in (12), but not with unergative verbs such as 'speak' in (13). I further assume that the lexical entry for a verb theme contains a PRED value, which specifies the semantic role(s) of its argument(s), but which is unspecified for grammatical functions, which are filled in according to LMT.

(12) Lexical entry for *łą́ą̨dhır* 'die'

($\uparrow$ PRED) = 'die < ARG >'

PATIENT

łaH: p-form: /łaH/, V_{Prefix}, Level 5, Position 1

thır: p-form: /thır/ ~ /thër/, V_{Root}, Level 1

m-str:

$$\begin{pmatrix} \text{p-form = /thır/} \Rightarrow {}_{\mu}[\text{ASPECT}] =_c \text{IMP} \vee \text{OPT} \\ \text{p-form = /thër/} \Rightarrow {}_{\mu}[\text{ASPECT}] =_c \text{PERF} \end{pmatrix}$$

(13) Lexical entry for *yałtı* 'speak'

(↑ PRED) = 'speak < ARG >'
AGENT

ya: p-form: /ya/, V_{Prefix}, Level 5, Position 4
ł: p-form: /ɬ/, V_{Prefix}, Level 1, Position 13
m-str: ($\downarrow_{\mu}$ VOICE) = ɬ
tı: p-form: /tı/, V_{Root}, Level 1

The lexical entry for 'die' in (12) specifies a patient as its argument, and is thus compatible with the rule in (11). The entry for 'speak' in (13), however specifies an agent. Applying the *D*-mapping rule in (11) to (13) would therefore violate coherence—specifically, it would generate a clash at the level of a-structure. In the next section, I will propose a mechanism by which such potential violations of coherence are repaired, resulting in semantic bleaching of the prefix *ł* when it is used thematically.

4 An OT account of semantic bleaching

Strictly speaking, applying the *D*-mapping rule in (11) to a PRED with a pre-specified agent argument does not predict semantic bleaching—rather, it predicts a clash of features at the level of a-structure. Therefore, an additional step of the analysis is necessary. Specifically, using OT-LFG (e.g. Lee 2001) I propose that the *D*-mapping function in (11) can be re-formulated as a violable constraint. Under this analysis, the *D*-mapping function is in conflict with both coherence as well as the information specified in the lexical entry of the PRED. The three constraints I will use are formalized in (14).

(14) Constraints used in OT-LFG analysis

a) MAX(*D*): The output of every *D*-mapping function must be realized in f-structure and a-structure.
b) MAX(PRED-ARG): For every PRED, every semantic role specified in the lexical entry of the PRED must be realized in the a-structure of the output.
c) COHERENCE(ARG): Every argument is specified for at most one semantic role.

If there were evidence that the f-structures and a-structures projected by different prefixes were ranked differently with respect to faithfulness, the constraint in (14)(a) could be further specified as MAX(*D*-[μ-VOICE: ɬ]). The interaction of these constraints is illustrated in the tableau in (15).

(15) Semantic bleaching where *D*-mapping function is outranked

(↑ PRED) = 'speak <ARG>' AG (↓$_{\mu}$ VOICE) = ł	COHERENCE (ARG)	MAX (PRED-ARG)	MAX(*D*)
SUBJ a. (↑ PRED) = 'speak <ARG>' AG PT	*!		
SUBJ b. (↑ PRED) = 'speak <ARG>' PT		*!	
SUBJ ☞c. (↑ PRED) = 'speak <ARG>' AG			*

The tableau in (15) illustrates the mechanism by which some prefixes can be bleached of their meaning, when that meaning would clash with the meaning of the PRED. Specifically, the semantic role projected by the m-structure of *ł* via the *D*-mapping function, that of patient, is not realized in the output, because it conflicts with the agent role specified in the lexical entry of the PRED. However, even when the output of the *D*-mapping function is unrealized, the m-structure information specified in (10) is still available to be used for the purposes of defining morphological selection and blocking relations. In this way, the f-structure/m-structure distinction in LFG enables us to account for how a prefix can be bleached of its semantic content, yet still retain its selectional properties.

In contrast, (16) illustrates how the *ł* prefix functions in unaccusative verbs. Recall that, in unaccusative verbs, the internal argument of the PRED is a patient, and thus there is no conflict between the lexical specification of the PRED and the output of the *D*-mapping rule. Thus, in (16) we see how the prefix *ł* renders the change from 'die' to 'kill'.

(16) No conflict between *D*-mapping function and PRED in unaccusative verbs

(↑ PRED) = ‘die <ARG>’ PT ($\downarrow_{\mu}$ VOICE) = ł	COHERENCE (ARG)	MAX (PRED-ARG)	MAX (*D*)
SUBJ OBJ a.(↑ PRED) = ‘caus <ARG, die <ARG>’>’ AG AG PT	*!		
SUBJ OBJ ☞b.(↑ PRED) = ‘caus <ARG, die <ARG>’>’ AG PT			
SUBJ OBJ c.(↑ PRED) = ‘caus <ARG, die <ARG>’>’ AG AG		*!	*

As shown in (16), because the internal PRED of this verb is unaccusative, the winning candidate (b) satisfies all three constraints simultaneously: it satisfies coherence, it realizes the lexically specified argument of the internal pred ‘die’, which is a patient, as well as the arguments specified by the *D*-mapping function, which are a patient or theme as the internal argument, and an agent as the external argument. To summarize, in unaccusative verbs such as ‘die’, the voice/valence prefix *ł* renders the change from ‘die’ to ‘kill’ ultimately as a result of its lexical entry in m-structure: lexical entry projects the attribute-value pair [VOICE ł] at m-structure, which in turn activates the *D*-mapping function in (11), which ultimately results in the change in f-structure/a-structure mapping as shown in (16).

5 Summary and conclusion

Like other prefixes in Tetsǫ́t’ıné, the *ł* voice/valence marker has both productive and thematic uses. In its productive uses, it acts as a causative prefix, introducing changes to the f-structure/a-structure mapping, and selects a particular form of the perfective marker—*ñe* or Ø. In its thematic uses, on the other hand, it appears to be semantically empty, and yet its selectional properties are retained.

LFG provides a way to describe this pattern by distinguishing f-structure from m-structure. Causativity is stated in terms of changes to the f-structure/a-structure mapping (Dione 2013), while morphological selection is stated at m-structure (Frank & Zaenen 2004). Under this view, both productive and thematic uses of *ł* involve the same lexical entry. When *ł* is compatible with the argument structure of a verb, the causative meaning is realized; when *ł* is semantically incompatible with a verb's argument structure, it is bleached of its semantic content. Finally, I suggested a formal mechanism by which to model this semantic bleaching, which is to formulate the *D*-mapping function as a violable constraint, within OT-LFG.

References

Bresnan, Joan. 2001. Lexical Functional Syntax. Malden, MA & Oxford: Blackwell.

Bresnan, Joan & Annie Zaenen. 1990. Deep Unaccusativity in LFG. In K. Daiwirek, P. Farrell & E. Mejías-Bikandi (eds.) *Grammatical Relations: A Cross-Theoretical Perspective:* 45-57. Stanford: CSLI Publications

Dalrymple, Mary. 2001. Lexical Functional Grammar. San Diego: Academic Press.

Dalrymple, Mary. 2015. Morphology in the LFG Architecture. *Proceedings of the LFG15 Conference*: 64-83.

Dione, Cheikh Bamba. 2013. Valency Change and Complex Predicates in Wolof: An LFG Account. *Proceedings of the LFG13 Conference*: 232-252.

Frank, Anette & Annie Zaenen. 2004. Tense in LFG: Syntax and Morphology. In Sadler & Spencer (eds.) *Projecting Morphology*: 23-66. CSLI Publications.

Hopper, Paul & Sandra Thompson. 1980. Transitivity in grammar and discourse. *Language* 56: 251-299.

Jaker, Alessandro. 2014. Selection and Blocking in the Northeast Dene Verb. *Proceedings of the LFG14 Conference*: 241-261.

Jaker, Alessandro & Emerence Cardinal. 2020. Tetsǫ́t’ıné Verb Grammar. Fairbanks: ANLC Publications.

Jaker, Alessandro, Nicholas Welch & Keren Rice. 2020. The Na-Dene Languages. In Daniel Siddiqi, Michael Barrie, Carrie Gillon, Jason D. Haugen & Éric Mathieu (eds.) *The Routledge Handbook of North American Languages.* New York & London: Routledge.

Kari, James. 1992. Some concepts in Ahtna Athabaskan word formation. In Mark Aronoff (ed.) *Morphology Now.* Albany, NY: SUNY Press: 107-132.

Lee, Hanjung. Optimization in Argument Expression and Interpretation: a Unified Approach. Stanford: Doctoral Dissertation, Stanford University.

Rice, Keren. 2000. Morpheme Order and Semantic Scope. Cambridge: Cambridge University Press.

On the inventory of grammatical functions in LFG from a Hungarian perspective

Tibor Laczkó
Károli Gáspár University of the Reformed Church in Hungary

Proceedings of the LFG'21 Conference

On-Line

Miriam Butt, Jamie Y. Findlay, Ida Toivonen (Editors)

2021

CSLI Publications

pages 181–201

http://csli-publications.stanford.edu/LFG/2021

Keywords: Hungarian, grammatical functions, COMP, XCOMP, PREDLINK

Laczkó, Tibor. 2021. On the inventory of grammatical functions in LFG from a Hungarian perspective. In Butt, Miriam, Findlay, Jamie Y., & Toivonen, Ida (Eds.), *Proceedings of the LFG'21 Conference, On-Line*, 181–201. Stanford, CA: CSLI Publications.

1 Introduction

Given the architecture, the assumptions and the principles of LFG, grammatical functions (GFs) play a central role in the theory. As a consequence, LFG has always needed a suitable taxonomy of GFs. Bresnan (1982b) offers the following classification in the earliest model of LFG.[1]

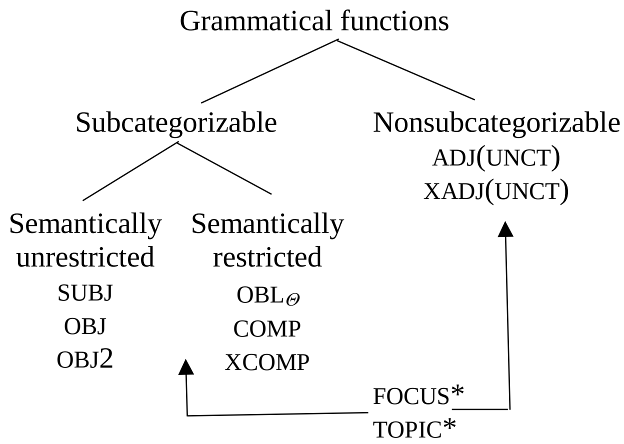


Figure 1. *Classification of grammatical functions (Bresnan 1982b: 287)*

This basic taxonomy of GFs in the clausal domain[2] has remained rather stable, except for one significant change: OBJ2 has been reclassified as semantically restricted: OBJ$_{\Theta}$. We find this modified classification in Bresnan et al. (2016), Börjars et al. (2019) and Dalrymple et al. (2019). However, Alsina et al. (1996, 2005) and several other authors since then have proposed that COMP and (to a lesser extent) XCOMP should be eliminated from the inventory of GFs in LFG. In this paper I point out that the tests used for other languages to support this proposal do not apply to the relevant Hungarian phenomena, as opposed to Szűcs's (2018) claim to the contrary. I also show that some Hungarian facts are straightforwardly analyzable by employing the COMP and XCOMP GFs. In addition, I argue that PREDLINK needs to be added to the inventory of LFG's GFs. The reason why I discuss this GF as well is two-fold. On the one hand, I believe that it is indispensable in the analysis of certain constructions. On the other hand, my view strongly contrasts with some recent GF-reductionist proposals in the LFG literature.

[1] As regards focus and topic, Bresnan remarks that their subcategorizability is parametric, governed by the "subject-oriented" vs. "topic-oriented" nature of languages.

[2] For overviews of proposals with respect to the inventory and nature of GFs in the nominal domain and alternative proposals, see Laczkó (1995, 2004).

The structure of the paper is as follows. In section 2, I give a brief overview of the COMP debate and present the Hungarian facts supporting the retention of this GF, arguing against a recent abandonment proposal. Section 3 is devoted to XCOMP along the same lines. In section 4, I show that there are two Hungarian copula constructions (expressing identity and possession) that strongly call for an analysis using PREDLINK, and I also point out that English identity copula constructions are also best analyzed in this fashion. In section 5, I conclude.

2 COMP

In this section first I discuss the COMP debate in LFG (2.1) and then I concentrate on Hungarian, arguing against Szűcs's COMP-less proposal (2.2.1) and adding further general remarks (2.2.2).

2.1 COMP in general

There are three different views related to COMP.

(A) All clausal complements have the COMP function.
(B) In "mixed" languages certain clausal complements have the regular functions, and other clausal complements are COMPs.
(C) There is no COMP function: all clausal complements have regular (nominal) functions: SUBJ, OBJ and OBL.

Below I discuss the most salient representatives of these views in the literature in the above order.

(A) In the spirit of Bresnan (1982b), Asudeh & Toivonen (2015: 380) give the following description of COMP. "Closed (saturated) complement: a clausal argument which has its own subject." Bresnan et al. (2016: 99) also cite this. Börjars et al. (2019) and Dalrymple et al. (2019) provide a similar definition.[3]

(B) Dalrymple & Lødrup (2000), concentrating on the COMP vs. OBJ contrast in the case of clausal arguments, propose that there are mixed languages in which there are two types of clausal complements: one of them calls for the standard COMP analysis, while the other is more appropriately analyzable as bearing the OBJ function. They use the following five tests (the first four are diagnostics for the OBJ function, and the fifth is a COMP test).

(i) If the argument of the V can be realized by either NPs or CPs, the CP bears OBJ.
(ii) If the NP and CP arguments of the V can be coordinated, the CP bears OBJ.
(iii) If the CP argument can be passivized, it bears OBJ.
(iv) Typically, but not in all languages, if the clausal complement can be involved in an unbounded dependency, it bears OBJ.

[3] However, Dalrymple et al. (2019) subscribe to the view that not all clausal complements have this function, see (B) below.

(v) Typically, if a CP can be the complement of a noun or an adjective, it bears COMP (because Ns and As are intransitive).

On these grounds they claim that German, English and Swedish belong to the mixed type, and they add that Slave also exhibits crucial characteristics of this type. Dalrymple et al. (2019) add some further details to this proposal.

Lødrup (2012) shows that there is a group of verbs in Norwegian whose noun phrase arguments exhibit syntactic behaviours characteristic of clausal arguments rather than noun phrase arguments; therefore, they are more appropriately analyzable as bearing COMP. This is important because it is one of the arguments against COMP (even in a mixed type approach) that it is burdened with the following redundancy: if a constituent has the COMP function, it can only be a CP.[4]

On the basis of agreement, pronominalization and coordination facts, Belyaev et al. (2017) argue that in Moksha Mordvin the majority of clausal complements (factive and eventive propositions) are straightforwardly analyzable as bearing the SUBJ, OBJ and OBL GFs, while a smaller group of CPs (non-factive propositions) are best treated as carrying the COMP function.

(C) Alsina et al. (2005) criticize Dalrymple & Lødrup's (2000) mixed languages approach, and on the basis of Spanish, Catalan and Malayalam data they argue for eliminating COMP from the inventory of GFs in LFG. One of their key arguments is based on Catalan clitcization and subcategorization facts involving clausal complements. Forst (2006), using German and French examples, and Patejuk & Przepiórkowski (2014, 2016), using Polish data, share this view. For useful comparative overviews of these three main approaches, see Patejuk & Przepiórkowski (2016), Szűcs (2018) and Dalrymple et al. (2019). It is also noteworthy here that Patejuk & Przepiórkowski's (2016) view is more reductionist than just abandoning COMP and XCOMP: motivated by Alsina (1996), they suggest a three-way GF-division: SUBJ-OBJ-DEP. The third label is short for dependents, subsuming both OBLs and ADJUNCTs.

2.2 COMP in Hungarian

2.2.1 On Szűcs (2018)

Szűcs (2018) subscribes to the anti-COMP view, and he argues that Hungarian embedded clauses do not need the COMP function at all. After briefly reviewing the COMP-related literature, he claims that the relevant Hungarian data can be adequately analyzed by assuming that finite and non-finite (i.e. infinitival) propositional arguments have the regular SUBJ, OBJ and OBL functions. He has the two standard arguments (shared by the COMP-less approaches) for this claim. On the one hand, he shows that DPs, finite CPs and non-finite Ss can

[4] See Alsina et al. (2005) for instance.

realize the same arguments of a predicate. On the other hand, he claims that these various categorial realizations of the same argument can be coordinated, which justifies the assumption that they share the same GF. Below I highlight the most important aspects of Szűcs's argumentation and I make my comments as we proceed.

As regards Szűcs's first argument, consider his examples in (1)-(4).[5]

(1) *Kati* *fél* *a* *kutyák-tól.*
Kate.NOM fear.PRES.3SG the dogs-from
'Kate fears dogs.'

(2) *Kati* *fél,* *hogy* *a* *kutya*
Kate.NOM fear.PRES.3SG COMP the dog.NOM
megharap-ja.
bite-PRES.3SG.DEFO
'Kate fears that the dog may bite her.'

(3) *Kati* *fél* *kutyá-t* *tart-ani.*
Kate.NOM fear.PRES.3SG dog-ACC keep-INF
'Kate fears keeping a dog.'

(4) *Kati* *a-ttól* *fél,* *hogy* *a* *kutya*
Kate.NOM that-from fear.PRES.3SG COMP the dog.NOM
megharap-ja.
bite-PRES.3SG.DEFO
'Kate fears that the dog may bite her.'

In (1), the second argument of the verb is expressed by an oblique case-marked DP, in (2), it is expressed by a finite clause, and in (3), it is expressed by an infinitival construction. According to Szűcs, they should be treated as sharing the same OBL function. (4) is a special case in that it contains an oblique case-marked pronoun (*attól* 'that.from') that is associated with the same finite clause as we see in (2). Szűcs points out that this type can be analyzed by assuming that the pronoun is the OBL argument and the finite clause is its ADJUNCT associate, as proposed by Rákosi & Laczkó (2005). Szűcs shows that the same parallels as those in (1)-(4) hold for the SUBJ and OBJ functions in Hungarian. Below I only cite his OBJ examples, because for my purposes the OBL and the OBJ cases are important.

(5) *Kati* *étel-t* *akar.*
Kate.NOM food-ACC want.PRES.3SG
'Kate wants food.'

(6) *Kati* *akar-ja,* *hogy* *e-gyünk.*
Kate.NOM want-PRES.3SG.DEFO COMP eat-SBJV.1PL
'Kate wants that we eat.'

[5] In the glosses below COMP stands for complementizer, DEFO for the definite object marker, INF for the infinitival marker, and SBJV for subjunctive mood.

(7) *Kati* *e-nni* *akar.*
Kate.NOM eat-INF want.PRES.3SG
'Kate wants to eat.'

(8) *Kati* *az-t* *akar-ja,* *hogy* *e-gyünk.*
Kate.NOM that-ACC want-PRES.3SG.DEFO COMP eat-SBJV.1PL
'Kate wants (it) that we eat.'

The four types in (1)-(4) and (5)-(8) are entirely parallel. Only one remark is in order here, which will be important in the discussion below. Notice that the verb is marked for a definite object (DEFO) in both (6) and (8). This is obvious in the case of (8), because demonstrative pronouns bearing the OBJ function trigger the definite conjugation on the verb as a rule. (6) demonstrates that *that* clause complements also trigger this conjugation on a transitive verb.

Although it is certainly true that in the case of a considerable number of Hungarian verbs we can find this four-way complement realization, the overwhelming majority of verbs do not have all the four options. What is of great importance, I claim, is that in Hungarian, too, there is a class of verbs that are best analyzed as subcategorizing for clausal complements bearing the COMP GF. Consider the following minimal pair examples.

(9) *Kati* *jelez-te,* *hogy* *induljunk.*
Kate signal-PAST.3SG.DEFO COMP start.SBJV.1PL
'Kate signalled that we should start.'

(10) *Kati* *jelz-ett,* *hogy* *induljunk.*
Kate signal-PAST.3SG COMP start.SBJV.1PL
'Kate signalled that we should start.'

In (9) the verb, in addition to the standard subject agreement inflection (3SG), is also marked for definite object agreement (DEFO). Thus, it is natural to assume that the clausal complement has the OBJ GF, and it triggers object agreement on the verb. This manifests the pattern exemplified in (6): the second argument is expressed by a clausal argument, and it is not associated with a co-occurring OBJ pronoun. However, the same verb with exactly the same semantics can be used without object agreement, see (10). And, crucially, in the case of this verb there is no semantically fully identical OBL pronoun plus clausal complement combination here, i.e., the type in (4) is not available. Given the semantics of the verb in both (9) and (10), I claim that it is not an option to assume that in (10) the clausal constituent is an adjunct.

Now consider (11).

(11) *Kati* *int-ett,* *hogy* *induljunk.*
Kate.NOM wave-PAST.3SG COMP start.SBJV.1PL
'Kate waved (her hand) that we should start.'

In the case of this verb there is no (either OBL or OBJ) nominal alternative realized either by a DP alone, the (1) type, or by the combination of an object or oblique pronoun combined with the clausal constituent, the (4) and (8) type, respectively. And, again, given the semantics of the verb in (11), it is not an option to assume that the clausal constituent is an adjunct. A certain number of verbs of communication in the broad sense exhibit similar properties in Hungarian. In this connection, consider the following quote from Dalrymple & Lødrup (2000: 118) "Foley and Valin (1984) show that the use of a finite clause as a core argument is a marked situation in UG, which is only allowed for verbs of saying in some languages." Dalrymple & Lødrup (2000) claim that this typological generalization supports their approach in the following way. If the clausal complement is syntactically integrated into a sentence, it has the OBJ GF, and if it does not take part in syntactic processes like other core arguments, it has the COMP GF. In a "mixed language" in their terminology these two cases coexist. Lødrup (2012: 386) writes: "COMP differs from the other complement functions by not having their properties; it is a complement that just 'is there', and does not take part in grammatical processes."[6]

On the basis of the foregoing discussion, my main claim is that the clausal complements of a whole range of Hungarian verbs of communication in the broad sense can be most appropriately analyzed along the COMP lines. Of course, it is also possible to claim that all these cases can be handled by assuming that these clausal complements, after all, still have an OBL function, but there are restrictions on their categorial realization. I think that this choice can be taken to be dependent, to a considerable extent, on the theory-internal persuasion of the researcher. The key issue here is whether we intend to capture the relevant facts in the dimension of GFs or in terms of categorial constraints on particular complements of individual predicates. My preference is the GF-based approach, while my external reviewer strongly advocates the categorial approach.[7]

[6] My external reviewer writes: "I do not see why the semantics of the verb in (11) prevents us from assuming that its clausal complement is an adjunct. Alternatively, one could assume that it is an oblique categorially constrained to be a CP." I think this paragraph provides enough language-internal and cross-linguistic justification for my non-adjunct approach. Of course, in principle, it would also be possible to develop an adjunct-based analysis by creating the necessary formal devices for capturing the semantic generalizations and parallels discussed above in general and for encoding that the predicates in question admit (or, rather, "optionally subcategorize for") a particular kind of propositional constituent. As regards the other approach mentioned by my reviewer, this would be the most plausible analysis on COMP-less grounds.

[7] It is also noteworthy in this connection that at LFG21 Péter Szűcs made the following written comment, still accessible on the website of the conference. "It must not be forgotten that *that*-clauses can be relatively freely added to a number of verbs that are communicative only in the very broad sense: *tapsol* (clap), *pislog* (blink), *bólint* (nod), etc. – *János tapsolt, hogy bejöhetünk.* (John clapped that we may enter / John clapped indicating that we may enter.). For these CPs I'd be in favor of a (thematic) adjunct analysis and a similar approach might work for other verbs if there is only a finite CP complement." My reply was as follows. "I also used the expression 'in the broad sense'. At the same time it's my conviction that these verbs are truly and definitely

As far as Szűcs's second COMP-less argument is concerned, he presents the following example among others (and, in my judgement, all his relevant examples, which I cannot discuss here for space limitations, are equally problematic).

(12) *Kati* *fél* *a* *kutyák-tól* *és* *hogy* *az-ok*
Kate.NOM fear.PRES.3SG the dogs-from and COMP that-PL
megharapják.
bite.PRES.3PL.DEF
'Kate fears dogs and that they might bite her.'

In (12) an oblique DP and a CP are conjoined. The claim is that the possibility of this kind of coordination justifies the assumption that these categorially different constituents can be coordinated because they share the same GF, and naturally this GF can only be OBL. It is important to point out that Patejuk & Przepiórkowski (2014, 2016) crucially base their COMP-less approach on similar coordination facts in other languages. As regards (12) (and Szűcs's related examples), my intuitions and the results of a small scale questionnaire question Szűcs's argumentation to a considerable extent. Even his own example is only marginally acceptable. On an OK/?/??/?*/* scale it would rank as ??. It is also noteworthy that the conjoined constituents in (12) are specifically related: the first (DP) conjunct is coreferential with the subject of the second (CP) conjunct. My claim is that if two semantically entirely distinct conjuncts of these two phrasal categories (DP and CP) are coordinated then the result is absolutely ungrammatical, see (13). If we swapped the two conjuncts, the result would be even worse. By (significant) contrast, if in the same example the pronoun plus CP version is used, i.e., type (4), the result is full grammaticality, see (14).

(13) **Kati* *fél* *a* *macskák-tól* *és* *hogy*
Kate.NOM fear.PRES.3SG the cats-from and COMP
a *kutyák* *megharapják.*
the dogs.NOM bite.PRES.3PL.DEF
'Kate fears cats and that the dogs might bite her.'

used in this broad (or very broad) communicative sense. I think this is a productive semantic domain that calls for a systematic treatment along the lines that I sketched. In theory the thematic adjunct option is also available. However, it is my conviction that the *jelez(1)* vs. *jelez(2)* minimal pair […] rather supports the COMP treatment. Compare: (9) *Kati jelezte, hogy induljunk* and (10) *Kati jelzett, hogy induljunk.* On semantic grounds, I can't see why the argument vs. adjunct status of the CP in (10) should be assumed to be different from that of the CP in (9). In both cases the CP expresses the message (the content of the signal)."

(14) *Kati* *fél* *a* *macskák-tól* *és* *a-ttól,* *hogy*
Kate.NOM fear.PRES.3SG the cats-from and that-from COMP
a *kutyák* *megharapják.*
the dogs.NOM bite.PRES.3PL.DEF
ca. 'Kate fears cats and the possibility that the dogs might bite her.'

My conclusion is that this construction type cannot be used as evidence for abandoning the COMP GF. This holds at least for the variety of Hungarian in which the grammatical status of the foregoing key examples is as I have pointed out.

My general remark on Szűcs's two arguments based on Hungarian data for abandoning COMP from the inventory of LFG GFs is that they are not convincing. On the contrary, they can be used to argue against his proposal.

As regards his first argument, the potential categorial diversity for the realization of the same GF, I have shown that there is a group of semantically (and cross-linguistically) identifiable verbs that can only take a CP complement. In this case the most natural assumption in a "mainstream" LFG framework is that the given complement carries the COMP function. Any other solution in a COMP-less approach seems to me to be less plausible for the following reason. The most straightforward COMP-less solution is that the verbs in question subcategorize for OBL, but the category of their OBL argument is constrained to CP. In my view it is a rather unusual situation that a GF cannot be realized by its default category (or categories). In Hungarian OBLs are canonically expressed by either (oblique) case-marked DPs or by postpositional phrases. Of course, it can be claimed that the semantics of the argument is responsible for this constraint: these are propositional arguments. However, in theory it would also be possible to use a derived nominal counterpart of the verb of such a CP, and this event nominal could be used in an oblique case-marked DP or in a PP. This alternative, however, is not available here.

As to Szűcs's second argument, the conjoinability of CP complements with categorially different complements, appears to backfire. CP complements by themselves (i.e. without pronominal support) seem to strongly reject coordination with non-CP complements**.** Thus, according to the logic of Szűcs's argumentation this lack of conjoinability actually supports the assumption that these non-conjoinable CPs bear a different GF: COMP.[8]

[8] My external reviewer, advocating the COMP-less approach, remarks that despite my claim to the contrary, the non-conjoinability here can be simply captured in the categorial dimension: CPs are not compatible with non-CPs, so we do not need to invoke the GF dimension with COMP. My response to this observation is that there are several cases in Hungarian in which conjoinability has to be accounted for by assuming a GF shared by different phrasal categories. The most salient example of this is the natural conjoinability of oblique case-marked DPs and PPs when they share either an OBL or an ADJUNCT GF. (They are different categories because they exhibit different

Dalrymple et al. (2019: 32) write: "until convincing arguments can be made that all COMPs in languages such as English, German, and Norwegian can be reanalyzed in terms of other grammatical functions, COMP cannot be abandoned on the basis of being redundant."[9] My fundamental claim is that Szűcs's arguments as they stand are not convincing enough; therefore, in Hungarian "COMP cannot be abandoned on the basis of being redundant."

2.2.2 Further remarks

In this section I make two additional remarks. (A) is about the oblique domain and (B) is about the subject and object domains.

(A) Below I repeat one of the five tests employed by Dalrymple & Lødrup (2000) from section 2.1, the COMP test.

(v) Typically, if a CP can be the complement of a noun or an adjective, it bears COMP (because Ns and As are intransitive).

In Hungarian there are deverbal nouns of the "simple event or result" types that can be argued to have a complement, and this complement can only be expressed by CPs. Consider the following examples.[10]

---

morpho-phonological properties.) Given this fact, the COMP-less approach would need to give a reason why CPs allegedly bearing the same GF cannot be conjoined with the other two categories. At LFG21 Péter Szűcs's second important written remark was similar to my external reviewer's. "As for the coordination data […], I really think a careful empirical investigation is required. I expect much variation here. A potential pitfall is that one might erroneously assume that GFs are the only relevant factors in coordination. This is very tempting for an LFG-practitioner, but in reality it may well be that GFs are just one factor out of many (c-structure categories, discourse structure, etc.)." My reply was as follows. "Your 'thought-provoking' 2018 paper made me start thinking about these phenomena (thanks for this motivation…). I readily accept your claim that coordination factors may not be reduced to the GF dimension. However, *you* used coordination examples to argue for abandoning COMP (a GF dimension). I took a look at your data and argumentation, and my claim is that, at least in the variety of Hungarian I speak and I am familiar with, these data rather support keeping COMP. Yes, there may be great variation here. As I briefly pointed out in the talk, there may even be dialectal differences here."

[9] My external reviewer writes: "one should really turn this around. Given that the simplicity criterion favors a framework with fewer theoretical concepts over one with more theoretical concepts, the burden of proof is on the side of the proponents of COMP. One could more appropriately say: *Until convincing arguments can be made that certain phenomena cannot be explained without COMP as a GF, the GF COMP should not be introduced in the inventory of GFs*." I think these two quotes, from Dalrymple et al. (2019) and from my reviewer, perfectly characterize the antagonistic with-COMP vs. without-COMP perspectives in LFG. I subscribe to the view of the with-COMP camp.

[10] In the glosses DEV stands for deverbal nominalizing suffix.

(15) *Kati* *jelz-és-e,* *hogy* *induljunk*
Kate.NOM signal-DEV-POSS.3SG COMP start.SBJV.1PL
'Kate's signal(ling) that we should start.'

(16) *a* *gondol-at,* *hogy* *János* *távoz-ott*
the think-DEV COMP John.NOM leave-PAST.3SG
'the thought that John left'

(17) *a* *kérd-és,* *hogy* *ki* *távoz-ott*
the ask-DEV COMP who.NOM leave-PAST.3SG
'the question of who left'

(15) can be taken to be the nominal counterpart of (10), and I think we can draw a straightforward parallel here. In the case of (10), I have argued that it is reasonable to assume that the CP, spelling out the content of the message expressed by signalling, is a complement bearing COMP. On these grounds it also stands to reason that the CP in (15) is a CP complement of the noun head, again, bearing COMP. Note that in the case of (15), just like in the case of (10), the only categorial option is CP, and the semantic correspondence between the two CPs is also obvious. As (16) and (17) demonstrate, the head noun typically imposes constraints on the actual type of the required CP: we cannot exchange the two CPs in these examples (*gondolat* 'thought' requires a declarative CP, while *kérdés* 'question' calls for an interrogative CP).[11]

I believe that the facts in the OBL domain in Hungarian amply support the idea that COMP needs to be retained. The crucial points are as follows. *(i)* There is a semantically identifiable group of verbs that can only take a CP complement, most naturally assumed to bear COMP, see (11) and the discussion of its relevance above. *(ii)* Certain (fundamentally) "result" deverbal nouns can also be assumed to subcategorize only for CP COMPs. *(iii)* Coordination facts also show that CP complements are not really conjoinable with oblique case-marked DPs (or PPs), see (12)-(13) and their discussion above. *(iv)* In addition, CP COMPs cannot bear all the same discourse functions as their DP/PP OBL counterparts. Consider the examples in (18), (19) and (20), and also compare them with (1), (4) and (2), respectively.

(18) *A* *kutyák-tól* *csak* *Kati* *fél.*
the dogs-from only Kate.NOM fear.PRES.3SG
ca. 'As far as dogs are concerned, only Kate is afraid of them.'

(19) *A-ttól,* *hogy* *a* *kutya* *megharap-ja,*
that-from COMP the dog.NOM bite-PRES.3SG.DEFO
csak *Kati* *fél.*
only Kate.NOM fear.PRES.3SG
ca. 'As far as getting bitten by the dog is concerned, only Kate is afraid of that.'

[11] Naturally, Szűcs's comment cited in Footnote 7 is valid in this case, too, and my reply is also the same as that I cited there.

(20) *Hogy a kutya megharap-ja,
COMP the dog.NOM bite-PRES.3SG.DEFO
csak Kati fél.
only Kate.NOM fear.PRES.3SG
ca. 'As far as getting bitten by the dog is concerned, only Kate is afraid of that.'

Recall from the discussion of (1), (2) and (4) that, non-finite propositional complementation aside, in the Hungarian system verbs like *fél* 'fear, be afraid of' can take as complements oblique case-marked DPs, as in (1) and (18), a similarly oblique case-marked pronoun with a CP associate, as in (4) and (19), and a CP on its own, as in (2) and (20). (18)-(20) contain sentences with a contrastive topic and a classic *csak* ('only') focus constituent. As (18) and (19) demonstrate the DP complement alone and the corresponding pronoun with its CP associate can bear the contrastive topic DF. By contrast, (20) shows that a CP alone cannot be a contrastive topic. I think this is a strong additional argument for retaining COMP in LFG's GF inventory.[12]

(B) The subject–object domain is different from the oblique domain discussed in (A) above in one important respect. Although there are full parallels between the four potential argument realization types, compare (1)-(4) and (5)-(8), the type illustrated by (6) in the subject–object domain is special. Naturally, it can be analyzed in exactly the same stand-alone CP fashion as the oblique counterpart in (2). However, given that Hungarian is a subject and object pro-drop language, there is an additional analytical option here: it can also be assumed that in this type we are dealing with a pro-dropped subject or object, in which case we can analyze this construction in the same way as the PRON + CP type exemplified in (8). I leave it to future research to investigate the theoretical ramifications of this potential analytical duality.

3 XCOMP

In this section first I briefly characterize XCOMP (3.1) and then I concentrate on Hungarian, arguing against Szűcs's XCOMP-less proposal (3.2).

[12] My external reviewer makes the following comment. "This sentence is an implicit acknowledgement that we don't need COMP." Of course, they are right from their COMP-less perspective, where basically all the relevant facts need to be captured in categorial terms. However, I still claim that from my with-COMP perspective this is a valid argument. Let me also add a minor technical point here. It seems to me that this specific constraint on contrastive topics is more straightforwardly capturable in the formal apparatus of LFG in the GF dimension: (CONTR-TOPIC) ≠ (COMP).

3.1 XCOMP in general

LFG's XCOMP is an "open (unsaturated) predicate complement" (Asudeh & Toivonen 2015: 380), realized by categorially varied constituents whose shared property is that they do not have an overt, c-structurally expressed subject, and their subject, present in f-structure, is functionally controlled by an appropriate controller from outside the constituent, hence its openness. XCOMP constituents are typically headed by non-finite verbs (infinitives and participles), see Szűcs's (2018) Hungarian example and its English translation from section 2.2, repeated below for convenience.

(7) *Kati* *e-nni* *akar.*
Kate.NOM eat-INF want.PRES.3SG
'Kate wants to eat.'

Here the XCOMP constituent is a VP headed by an infinitive in both languages and its unexpressed subject argument is functionally controlled by the overt subject of the finite matrix verb. Predicative APs and NPs can also bear this function, which will be important in sections 3.2 and 4 below.

As regards the XCOMP GF, LFG practitioners in the pro-COMP camp obviously assume the standard status of XCOMP in the GF inventory of the theory. Interestingly, Falk (2005) goes even further and he proposes additional open GFs: XOBJ$_{\Theta}$ and XOBL$_{\Theta}$. I do not think that this extension is warranted by Hungarian data.

In the anti-COMP camp there is no absolute consensus about XCOMP. For instance, Forst (2006), from an implementational perspective, argues for abandoning COMP and for keeping XCOMP. By contrast, Alsina et al. (2005: 41) write "XCOMP should probably go the same way as COMP", but they do not substantiate this claim. Patejuk & Przepiórkowski (2016) argue that the same kinds of coordination facts justify abandoning XCOMP as they capitalize on in the case of getting rid of COMP. They develop an alternative and implementationally tested analysis of functional control into closed GFs like OBJ or OBL.

3.2 XCOMP in Hungarian

Szűcs (2018) also claims that XCOMP, just like COMP, can be dispensed with in the analysis of Hungarian. Recall that in section 2.2, when I discussed his arguments for abandoning COMP, realized by CPs, I showed that he assumes that non-finite (infinitival) S-s can also bear the regular (SUBJ, OBJ and OBL) GFs, just like CPs. In the case of his example in (7), repeated in 3.1 above, he assumes that the infinitival constituent has the OBJ (and not the XCOMP) function, and he points out that control into this OBJ can be handled along the lines proposed by Patejuk & Przepiórkowski (2016). In the case of his other relevant example, repeated here for convenience, he assumes that the infinitival VP bears OBL, and control works in the same way.

(3) *Kati* *fél* *kutyá-t* *tart-ani.*
Kate.NOM fear.PRES.3SG dog-ACC keep-INF
'Kate fears keeping a dog.'

Szűcs provides the same two arguments for abandoning XCOMP as he provides for abandoning COMP: (i) categorial complement realization variability and interchangeability and (ii) the conjoinability of categorial unlikes.

As regards (i), categorial variability, I think his argument here is even weaker than in the case of COMP, because there are a great number of verbs that can only take infinitival complementation, see my randomly selected example in (21), where the order of categorial realization types follows that in (5)-(8).

(21) a. **Kati* *próbál-ja* *a* *koncentrál-ás-t.*
Kate.NOM try-PRES.3SG.DEFO the concentrate-DEV-ACC
ca. '*Kate is trying concentration.'

b. **Kati* *próbál-ja,* *hogy*
Kate.NOM try-PRES.3SG.DEFO COMP
koncentrál-j-on.
concentrate-SBJV-3SG
lit. 'Kate is trying that she should concentrate.'

c. *Kati* *próbál* *koncentrál-ni.*
Kate.NOM try.PRES.3SG concentrate-INF
'Kate is trying to concentrate.'

d. **Kati* *próbál-ja* *az-t,* *hogy*
Kate.NOM try-PRES.3SG.DEFO that-ACC COMP
koncentrál-j-on.
concentrate-SBJV-3SG
lit. 'Kate is trying the thing that she should concentrate.'

As (21c) shows, the complement can only be realized by an infinitival construction. The Hungarian verbs *igyekszik* 'endeavour (to do sg)', *habozik, hezitál, tétovázik,* all three: 'hesitate (to do sg)', *baszik* (vulgar) 'literally: fuck; rudely refuse (to do sg)' behave similarly.

As regards (ii), conjoinability, the argument is as weak as in the case of COMP. Below I show Szűcs's relevant example.

(22) *Kati* *étel-t* *és* *a-zzal*
Kate.NOM food-ACC and that-with
jóllak-ni *akar.*
satisfied.become-INF want-PRES.3SG
'Kate wants food and to be satisfied with it.'

Just like in the case of Szűcs's COMP coordination example in (12) in section 2.2, this example is unacceptable according to my intuitions and my small-scale survey. Moreover, here, too, the conjoined constituents are semantically linked. The object NP of the matrix verb is coreferential with the oblique complement of the infinitive. My remark here, too, is that if the two conjuncts are semantically entirely independent, such constructions are absolutely ungrammatical, see (23).

(23) **Kati* *étel-t* *és* *Pali-val* *sétál-ni* *akar.*
Kate.NOM food-ACC and Paul-with walk-INF want-PRES.3SG
'Kate wants food and to go for a walk with Paul.'

Szűcs also mentions "subject-to-object raising" constructions in Hungarian. Consider his key example in (24).

(24) *Kati-t* *boldog-nak / zseni-nek* *tart-om.*
Kati-ACC happy-DAT genius-DAT consider-PRES.1SG
'I consider Kate happy / a genius.'

In this sentence 'Kate' undoubtedly has the (non-thematic) OBJ function, and the non-SUBJ semantic argument of the verb can be realized by a predicatively used AP ('happy') or NP ('a genius'). In this case, Szűcs (2018: 335) writes: "the (X)OBJ_{Θ} seems to be an appropriate function for raising in Hungarian and XCOMP is not needed." He assumes, agreeing with Patejuk & Przepiórkowski (2016), for instance, that the X in the function name can be omitted if an appropriate treatment of functional control (into closed GFs) is developed. Even so, my problem with Szűcs's alternative GF proposal is that, as far as I know, the OBJ_{Θ} GF has not been proposed in any LFG analysis of any phenomenon in Hungarian. Therefore, its inclusion in the set of Hungarian GFs would require substantial justification. As things stand now, Szűcs gets rid of XCOMP in the analysis of this functional control construction type by introducing a GF otherwise unattested in this language so far. Moreover, it is an additional and equally serious problem with Szűcs's proposal that in his analysis of raising constructions he is forced to assume that not only predicative noun phrases but adjectival phrases can also bear his newly introduced OBJ_{Θ} GF, which is a rather unorthodox category–function combination.[13]

[13] My external reviewer makes the following comment. "The observation that, in a framework without COMP or XCOMP, as in that defended by Szűcs (2018) and Patejuk & Przepiórkowski (2016), the predicative adjective phrase of consider-type verbs is assigned one of the GFs OBJ, OBJ_{Θ}, or OBL can hardly be taken as an argument for COMP or XCOMP, however unexpected it may be for someone who assumes the standard LFG inventory of GFs to call a predicative adjective phrase an OBJ, OBJ_{Θ}, or OBL. It is an obvious consequence of removing COMP and XCOMP from the inventory of GFs that the remaining GFs, particularly OBJ, OBJ_{Θ} and OBL, will have to be used to designate grammatical functions that, in the standard framework, are labeled as COMP or XCOMP." My brief response to this observation is as follows. It seems to me that my reviewer's view of the nature of LFG's GFs is rather simplistic. Of course, it is understandable that a COMP/XCOMP-less approach needs to use one of the three remaining GFs (other than SUBJ). However, if my reviewer assumes, as they state, that

4 PREDLINK

In their XLE implementational platform, Butt et al. (1999) propose a new GF: PREDLINK for a uniform treatment of copula constructions in English, German and French. It is interesting to see how this GF figures in most recent authoritative books on LFG. There is no mention at all of PREDLINK in Bresnan et al. (2016). Börjars et al. (2019: 155) mention this GF only once in a "Reading" section as an alternative of XCOMP in the analysis of copula constructions. Dalrymple et al. (2019) compare the PREDLINK and the XCOMP analyzes of certain copula constructions (2019: 32-33, 194-197).[14] This (rather minimal) coverage of PREDLINK[15] saliently contrasts with the standard, mainstream LFG view of the status of COMP and XCOMP in the same three books.

The two major general LFG strategies for the treatment of copula constructions (CCs) across languages are represented by Butt et al. (1999) and Dalrymple et al. (2004). In the former approach, CCs are treated in a uniform manner functionally. The copula is always assumed to be a two-place predicate. It subcategorizes for a subject (SUBJ) argument, which is uncontroversial in any analysis of these constructions, and the other constituent is invariably assigned a special, designated function designed for the second, "postcopular" argument of the predicate: PREDLINK. As opposed to this approach, in Dalrymple et al.'s (2004) view, the SUBJ & PREDLINK version is just one of the theoretically available options. In addition, they postulate that the copula can be devoid of a PRED feature (and, consequently, argument structure) and in this use it only serves as a pure carrier of formal verbal features: tense and agreement. Finally, it can also be used as a one-place "raising" predicate, associating the XCOMP function with its propositional argument and also assigning a non-thematic SUBJ function.

In Laczkó (2021) I analyze five CCs in Hungarian: attribution/classification, identity, location, existence and possession. I subscribe to the view, advocated by Dalrymple et al. (2004) and also by Nordlinger & Sadler (2007), among others, that the best LFG strategy is to examine all CCs individually, and to allow for diversity and systematic variation both in c-structure and in f-

---

it is unproblematic to analyze an AP as possibly bearing either of the two OBJ functions, then for me this is tantamount to using these GF labels without minimally taking into consideration the general(ly acknowledged) grammatical (syntactic and morphosyntactic) properties of OBJs and APs. While I admit that this is a possible alternative approach to GFs in LFG, I strongly subscribe to the view I am defending in this paper.

[14] In Chapter 6 of Laczkó (2021) I present a comprehensive assessment of main approaches to copula constructions in English, with a detailed and systematic comparison of LFG and the Chomskyan mainstream.

[15] Even when PREDLINK is discussed, and thereby its existence in LFG is acknowledged, its actual status in the GF inventory is not addressed.

structure representations across and even within languages. This means that I reject Butt et al.'s (1999) and Attia's (2008) uniform PREDLINK approach at the f-structure level. In my analysis the copula has five distinct lexical forms. I show that this five-way distinction is strongly justified by the fact that each of these types has a partially different set of properties.[16] I make two fundamental claims. (i) It is highly implausible to analyze all the five Hungarian CCs in the same, uniform functional way. (ii) PREDLINK needs to be employed for the most feasible LFG analysis of the possession and identity types.

Let us first take a brief look at my analysis of Hungarian possession CCs in Laczkó (2021). Consider the following example (Laczkó 2021: 318).

(25) *Az igazgató-nak van szóvivő-je.*
the director-DAT is spokesperson-his.NOM
'The director has a spokesperson.'

I claim that this special CC type is best analyzed along the PREDLINK lines.[17] My intuitive assumption is that the function of the copula here is to link the possessor and the possessed entity at the clause level. In other words, the copula "raises" the possessive relationship expressed within DPs to a sentential level. The crucial parts of my representation of the lexical form of the possession copula is shown in (26).[18]

(26) *van*, V (↑ PRED) = 'BE_{poss} < (↑SUBJ) (↑PREDLINK) >'
possessee possessor
(↑ PREDLINK CASE) $=_c$ dat

Now let us take a brief look at identity CCs exemplified in (27).

(27) a. *Az igazgató volt a szóvivő.*
the director.NOM was the spokesman.NOM
'The director was the spokesman.'
b. *van*, V (↑ PRED) = 'BE_{ident} < (↑ SUBJ) (↑ PREDLINK) >'
(↑ PREDLINK CASE) $=_c$ nom

I believe that not only my PREDLINK analysis of this Hungarian CC in Laczkó (2021) is the best solution, but the English counterpart (see the translation in (27a)) is also most appropriately analyzed along the same lines for the following reasons. *(A)* It would be highly implausible to assume that the

[16] For an overview of these details, see Table 6.3 in Laczkó (2021: 321).
[17] For my full argumentation for the assumption that the PREDLINK analysis is superior to assuming that the dative possessor bears OBL, see Section 6.3.2.5 in Laczkó (2021).
[18] The detailed representation formalizes the following properties of Hungarian possession CCs. (i) The possessee SUBJ must be indefinite and third person (singular or plural). (ii) The possessor PREDLINK's case must be dative (which is one of the case markers of possessors within possessive DPs, the other being nominative). (iii) The copula must be focused unless there is another focused constituent in the sentence.

postcopular fully referential DP constituent has the (sentential) PRED feature, and the copula as a co-head only contributes the usual morphosyntactic features (tense and agreement). An XCOMP "raising" analysis would suffer from the same problem, because the constituent in question would have the (sentential) PRED feature. *(B)* All the other three standard non-SUBJ GFs would be implausible to varying extents. I think that the two object functions (OBJ and OBJ_{Θ}) would not be meaningful options, because it hardly makes theoretical sense to assume that the copula is a transitive verb.[19] Thus, the remaining choice would be OBL, see Footnote 21 in Patejuk & Przepiórkowski (2016: 547). However, I think that this would just be the best of the inappropriate solutions in the "straightjacket environment" of the canonical inventory of GFs in LFG for three reasons. *(i)* The category/form of the second argument is not at all oblique-like. *(ii)* There is number agreement between the subject and this argument.[20] *(iii)* The two arguments are "on a par" in that they can swap their GFs (which naturally follows from the identificational/equative role of the predicate). Compare the following sentence with the English translation in (27a). *The spokesman was the director*.[21]

[19] Note that Bresnan (1982c), among others, assumes that the postcopular noun phrase in *there*-constructions bears OBJ (and *there* bears SUBJ).

[20] There is person and number agreement between the SUBJ possessum and the PREDLINK possessor in Hungarian possession CCs. (This agreement is present within possessive DPs and, as I pointed out above, we can assume that the possessum–possessor relation is "raised" to the clausal level by the possession copula, including the agreement dimension.) I think that this shared agreement property of the two Hungarian CCs that I analyze by employing PREDLINK lends additional support to this PREDLINK concept.

[21] My external reviewer makes the following remarks. "It seems that PREDLINK would be reserved for two constructions involving the copula: the identity construction and the possession construction. But one fails to see what the two uses of PREDLINK have in common: in one construction this GF is nominative and in the other one it is dative. I get the impression that the only reason for wanting to add this GF to the inventory is that it is a closed GF, without a functionally controlled subject, which means it cannot be XCOMP, generally taken to be an open GF, and the author feels it is unintuitive to use any of the existing closed GFs (OBJ, OBL, etc.). If one accepts the idea that what makes a GF open or closed is not the name that we give to the GF but whether it is associated with a control equation that identifies its subject with a GF of the embedding verb, this discussion becomes irrelevant. We could call it OBJ or OBL: it is a closed function if there is no control equation to go with it and it is an open function if there is a control equation establishing identity between its subject and a GF of the controlling verb." My response is as follow. As I showed, the constituent that I assume to have tbe PREDLINK function is a DP. As I also point out, they share a special agreement property. True, they bare different cases. However, both nominative case and dative case (in this particular use) are "structural" (i.e. non-oblique) cases. As regards the reviewer's repeated point that any standard closed function (other than SUBJ) can be used instead in an unproblematic manner, I can only repeat my response to a previous comment of theirs: this is tantamount to using these GF labels without minimally taking into

5 Concluding remarks

In this paper I showed that Hungarian does not provide convincing evidence for eliminating COMP and XCOMP from the inventory of GFs in LFG. On the contrary, it provides evidence for retaining these functions. In addition, I argued that PREDLINK is also needed for principled theory-internal reasons at least in the analysis of certain copula constructions (identity and possession in Hungarian and identity in English).

My view of the GF inventory is not reductionist; on the contrary, it is expansionist. I readily admit that the reductionist approach is also fully legitimate in LFG, and principled alternative analyzes can be developed of the same phenomena that have traditionally been treated in terms of the mainstream GF inventory (see my external reviewer's comments).[22] However, on the basis of the Hungarian facts discussed here my theory-internal choice is the classical LFG approach to GFs.

It is a frequently repeated reductionist claim that dropping COMP and XCOMP has the favourable side-effect that LFG's Lexical-Mapping Theory can be made more streamlined and principled. However, in my view first a broad consensus on the number and nature of GFs in the inventory should be achieved (and adding GFs is a likely option here, see PREDLINK, for instance) and it is only after this that the argument–function mapping system should be (re)developed.

Acknowledgements

For very useful comments, I am grateful to the audience of my talk at LFG21, in particular to Péter Szűcs, Adam Przepiórkowski and Alex Alsina. I am indebted to my internal and external reviewers for their valuable comments, which helped me to make the presentational and argumentative aspects of the paper stronger. As usual, all remaining errors and shortcomings are my sole responsibility.

References

Alsina, Alex, KP Mohanan & Tara Mohanan. 1996. Untitled submission to the LFG List. 3 Sept 1996.

consideration the general(ly acknowledged) grammatical (syntactic and morphosyntactic) properties of OBJs.

[22] In the last paragraph of the review my external reviewer writes the following. "My conclusion is that the paper presents no solid arguments for including the three GFs under discussion in the inventory of GFs. It seems to me that it would probably be easier to rewrite this paper so that it presents arguments in favor of eliminating the GFs COMP, XCOMP, and PREDLINK than to revise it in such a way that the phenomena presented can be shown to provide arguments for supporting the claim that these GFs are needed." I am not sure that I agree with this conclusion. It seems that the with-COMP vs. without-COMP debate continues.

Alsina, Alex, KP Mohanan & Tara Mohanan. 2005. How to get rid of the COMP. In Miriam Butt & Tracy Holloway King. eds. *The Proceedings of the LFG'05 Conference*. Stanford, CA: CSLI Publications, 21-41.
Alsina, Alex. 1996. *The Role of Argument Structure in Grammar. CSLI Lecture Notes 62*. Stanford, CA: CSLI Publications.
Asudeh, Ash & Ida Toivonen. 2015. Lexical-functional grammar. In Bernd Heine & Heiko Narrog. eds. *The Oxford Handbook of Linguistic Analysis*. *Second Edition.* Oxford: Oxford University Press, 373-406.
Attia, Mohammed. 2008. A unified analysis of copula constructions. In Miriam Butt & Tracy Holloway King. eds. *The Proceedings of the LFG'08 Conference*. Stanford, CA: CSLI Publications, 89-108.
Belyaev, Oleg, Anastasia Kozhemyakina & Natalia Serdobolskaya. 2017. In defense of COMP complementation in Moksha Mordvin. In Miriam Butt & Tracy Holloway King. eds. *The Proceedings of the LFG'17 Conference*. Stanford, CA: CSLI Publications, 84-13.
Börjars, Kersti, Rachel Nordlinger & Louisa Sadler. 2019. *Lexical-Functional Grammar. An Introduction.* Cambridge: Cambridge University Press.
Bresnan, Joan. 1982a. ed. *The Mental Representation of Grammatical Relations*. Cambridge, MA: MIT Press.
Bresnan, Joan. 1982b. Control and complementation. In Bresnan (1982a), 282-390.
Bresnan, Joan. 1982c. The passive in lexical theory. In Bresnan (1982a), 3-84.
Bresnan, Joan, Ash Asudeh, Ida Toivonen & Stephen Wechsler. 2016. *Lexical-Functional Syntax*. Wiley Blackwell.
Butt, Miriam, Tracy Holloway King, Maria-Eugenia Niño & Frédérique Segond. 1999. *A Grammar Writer's Cookbook*. Stanford: CSLI Publications.
Dalrymple, Mary, Helge Dyvik & Tracy Holloway King. 2004. Copular complements: Closed or open*?* In Miriam Butt & Tracy Holloway King. eds. *Proceedings of the LFG04 Conference*. Stanford, CA: CSLI Publications, 188-198.
Dalrymple, Mary & Helge Lødrup. 2000. The grammatical functions of complement clauses. In Miriam Butt & Tracy Holloway King. eds. *The Proceedings of the LFG'00 Conference*. Stanford, CA: CSLI Publications, 104-121.
Dalrymple, Mary, John J. Lowe & Louise Mycock. 2019. *The Oxford Reference Guide to Lexical Functional Grammar*. Oxford: OUP.
Foley, William A. & Robert D. Van Valin. 1984. *Functional Syntax and Universal Grammar*. Cambridge: Cambridge University Press.
Forst, Martin. 2006. COMP in (parallel) grammar writing. In Miriam Butt & Tracy Holloway King. eds. *The Proceedings of the LFG'06 Conference*. Stanford, CA: CSLI Publications, 222-239.
Laczkó, Tibor. 1995. *The Syntax of Hungarian Noun Phrases – A Lexical-Functional Approach.* Frankfurt am Main: Peter Lang.

Laczkó, Tibor. 2004. Grammatical functions, LMT, and control in the Hungarian DP revisited. In Miriam Butt & Tracy Holloway King. eds. *The Proceedings of the LFG'04 Conference*. Stanford, CA: CSLI Publications,313-333.

Laczkó, Tibor. 2012. On the (un)bearable lightness of being an LFG style copula in Hungarian. In Miriam Butt & Tracy Holloway King. eds. *The Proceedings of the LFG'12 Conference*. Stanford, CA: CSLI Publications, 341-361.

Laczkó, Tibor. 2021. *Lexicalising Clausal Syntax: The Interaction of Syntax, the Lexicon and Information Structure in Hungarian.* Amsterdam: John Benjamins.

Lødrup, Helge. 2012. In search of a nominal COMP. In Miriam Butt & Tracy Holloway King. eds. *The Proceedings of the LFG'12 Confere*nce. Stanford, CA: CSLI Publications, 383-404.

Nordlinger, Rachel & Sadler, Louisa. 2007. Verbless clauses: Revealing the structure within. In Grimshaw, Jane, Joan Maling, Chris Manning, Jane Simpson & Annie Zaenen. eds. *Architectures, Rules and Preferences: A Festschrift for Joan Bresnan.* Stanford, CA: CSLI Publications, 139-160.

Patejuk, Agnieszka & Adam Przepiórkowski. 2014. Control into selected conjuncts. In Miriam Butt & Tracy Holloway King. eds. *The Proceedings of the LFG'14 Confere*nce. Stanford, CA: CSLI Publications, 448-460.

Patejuk, Agnieszka & Adam Przepiórkowski. 2016. Reducing grammatical functions in LFG. In Arnold, Doug, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller. eds. *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar*. Stanford: CSLI Publications, 541-559.

Rákosi, György & Tibor Laczkó. 2005. Verbal category and nominal function – Evidence from Hungarian subject clauses. In Miriam Butt & Tracy Holloway King. eds. *The Proceedings of the LFG'05 Conference*. Stanford, CA: CSLI Publications, 353-370.

Szűcs, Péter. 2018. A COMP-less approach to Hungarian complement clauses. In Miriam Butt & Tracy Holloway King. eds. *The Proceedings of the LFG'18 Conference*. Stanford, CA: CSLI Publications, 325-342.

A constraint-based approach to anaphoric and logophoric binding in Mandarin Chinese and Cantonese

Chit Fung Lam
University of Manchester

Proceedings of the LFG'21 Conference

On-Line

Miriam Butt, Jamie Y. Findlay, Ida Toivonen (Editors)

2021

CSLI Publications

pages 202–222

http://csli-publications.stanford.edu/LFG/2021

Keywords: binding, anaphor, logophor, reflexive, Chinese, Cantonese, adnominal possession, Lexical-Functional Grammar

Lam, Chit Fung. 2021. A constraint-based approach to anaphoric and logophoric binding in Mandarin Chinese and Cantonese. In Butt, Miriam, Findlay, Jamie Y., & Toivonen, Ida (Eds.), *Proceedings of the LFG'21 Conference, On-Line*, 202–222. Stanford, CA: CSLI Publications.

Abstract

This paper proposes an LFG constraint-based approach to binding in Mandarin Chinese and Cantonese. We illustrate the power of LFG's f-structure in developing a formal model which is, in essence, a unifying proposal integrating syntactic anaphoric binding with pragmatically-rooted but grammaticised logophoric binding. The anaphoric-binding component of our model resolves the local binding of complex reflexives and that of simplex reflexives, whereas the logophoric-binding component handles the long-distance binding of simplex reflexives. Our view that Chinese binding is best explained by a dual system encompassing syntactic (anaphoric) and pragmatic (logophoric) aspects is in line with Huang and Liu (2001). While it is not easy for a syntactic theory to accommodate logophoric binding, the LFG formalism has a high degree of flexibility, allowing it to model both types of binding while maintaining its formal, mathematical rigour. Our constraint-based proposal offers an alternative binding theory in response to recent Minimalist proposals on Chinese binding (e.g., Giblin, 2016; Reuland, Wong & Everaert, 2020), opening up a cross-theoretical dialogue. We establish the notion of grammaticised logophoricity in Chinese binding in connection with crosslinguistic studies. Empirically, we examine a range of data to clarify properties of Chinese reflexives and settle past debates, in particular, the animacy debate in relation to typological research on adnominal possession. The comparison between Mandarin Chinese and Cantonese contributes to the comparative study of binding phenomena in Sinitic languages.

1 Introduction[1]

Chinese anaphora has continued to fascinate linguists despite decades of research (e.g., Tang, 1989; Huang & Tang, 1991; Xue, Pollard & Sag, 1994; Cole & Wang, 1996; Huang & Liu, 2001; Pan & Hu, 2003; Giblin, 2016; Charnavel, Huang, Cole, & Hermon, 2017; Charnavel & Y.-J. Huang, 2018; Sperlich, 2019; Reuland, Wong, & Everaert, 2020). One of the most intriguing aspects is the reflexive *ziji*, whose long-distance (LD) binding seems to be elusive to the locality requirement of anaphoric binding (Chomsky, 1981).

Past research on the LD binding of *ziji* can be broadly divided into two perspectives: (derivational) syntax-based approaches (e.g., feature-agreement systems by Tang, 1989; Huang & Tang, 1991; Giblin, 2016; Reuland et al., 2020) vs discourse-functional approaches (e.g., self-ascription theory by Pan, 1997; neo-Gricean pragmatic theory by Y. Huang, 2016). Each of these studies seems to explain a part of the overall picture. There is also a predominant focus on Mandarin Chinese, leaving other Chinese varieties seldom discussed. To resolve issues of Chinese anaphora, what we need, perhaps, is a unifying proposal that: i) considers insights from both syntactic and functional perspectives; ii) provides a formal, explicit system that explains the binding of different pronouns (not just *ziji*); and iii) considers more Chinese varieties.

[1] I am extremely grateful to Prof Kersti Börjars and Prof Eva Schultze-Berndt for their very helpful comments and suggestions as this work developed. My deep gratitude to Prof Mary Dalrymple for her insightful comments and invaluable advice on the LFG binding theory and formalism. Many thanks to Prof John Payne for his advice as I was preparing for the presentation. I am very grateful to the two anonymous reviewers of this paper, three reviewers of the conference abstract, and audience of LFG21, especially Dr Jamie Findlay, Dr Péter Szűcs, Dr James Donaldson, and Dr Rebecca Dinkel for their helpful comments. Special thanks to Ziling Bai and Lin Zhang for assistance when I required additional native-speaker judgement. Any errors in this paper are mine.

This paper focuses on the binding properties of four 3rd person singular Mandarin Chinese (MC) and Cantonese (CC) reflexives. Like other pronouns in MC and CC, their spoken forms do not express distinction in gender.

	Complex reflexive	Simplex reflexive
Mandarin Chinese (MC)	*taziji*	*ziji*
Cantonese (CC)	*keuhjihgei*	*jihgei*

We argue that LFG's f-structure provides the formal environment for a unifying proposal integrating syntactic anaphoric binding as well as pragmatically-rooted but grammaticised logophoric binding. For anaphoric binding, we demonstrate that MC and CC do not uphold the widely assumed f-commanding relation between the antecedent and anaphor (Dalrymple, 1993, 2015); nor are the binding patterns captured by the four binding domains (Co-argument Domain, Minimal Complete Nucleus, Minimal Finite Domain, Root Domain) LFG posits for typologically diverse languages (Dalrymple, 1993, 2015). For logophoric binding, we expand on Dalrymple's (2015) proposal as we develop constraints to differentiate the various types of logophoric binding in regard to Sells's (1987) logophoric taxonomy, which has been shown to be valuable to binding in Sinitic languages (see Cole et al., 2001).

2 Properties of MC & CC reflexives

2.1 Grammatical functions and basic patterns (local vs LD binding)

The complex reflexives *taziji* (MC) and *keujihgei* (CC) are locally bound: in (1a), *taziji* is bound by *Lisi* rather than *Zhangsan*; likewise, in (1b), *keujihgei* is bound by *Gafai* instead of *Amihng*.

(1) a. zhangsan$_i$ shuo [lisi$_j$ changchang biaoyang **taziji**$_{*i/j}$]
Zhangsan say Lisi always praise C.SELF[2]
'Zhangsan says that Lisi always praises himself.' (MC)
b. amihng$_i$ wah [gafai$_j$ sihngyaht jaan **keuihjihgei**$_{*i/j}$]
Amihng say Gafai always praise C.SELF
'Amihng says that Gafai always praises himself.' (CC)

We will discuss the formal constraint capturing the local-binding relation in section 5. In comparison, the simplex reflexives *ziji* (MC) and *jihgei* (CC) are subject to both local and LD binding: in (2a), *ziji* is bound by *Lisi* or *Zhangsan*, depending on the context; a similar situation applies to *jihgei* in (2b).

(2) a. zhangsan$_i$ shuo [lisi$_j$ changchang biaoyang **ziji**$_{i/j}$]
Zhangsan say Lisi always praise SELF
'Zhangsan says that Lisi always praises him(self).' (MC)
b. amihng$_i$ wah [gafai$_j$ sihngyaht jaan **jihgei**$_{i/j}$]
Amihng say Gafai always praise SELF
'Amihng says that Gafai always praises him(self).' (CC)

All four reflexives can be assigned the grammatical functions (GFs) of OBJ, OBJ$_\theta$, OBL$_\theta$, (embedded) SUBJ or POSS.[3] When functioning as (embedded) SUBJ

[2] In this paper, we gloss simplex reflexives as SELF and complex reflexives as C.SELF.
[3] The simplex reflexives, *ziji* and *jihgei*, can also be used as adverbials with the meaning of "by oneself". Our analysis will not cover this usage.

or POSS, the reflexive is bound by potential antecedent(s) in higher clause(s), meanwhile observing its local or LD binding properties:

(3) a. zhangsan$_i$ shuo lisi$_j$ renwei [**ziji**$_{i/j}$/**taziji**$_{*i/j}$-de pengyou hen qinfen]
Zhangsan say Lisi think SELF-POSS friend very be.diligent
'Zhangsan says Lisi thinks his friend is very diligent.' (MC)

b. amihng$_i$ wah gafai$_j$ yihngwaih [**jihgei**$_{i/j}$/**keuihjihgei**$_{*i/j}$-ge pahngyauh
Amihng say Gafai think SELF/C.SELF-POSS friend
hou kahnlihk]
very diligent
'Amihng says that Gafai thinks his friend is very diligent.' (CC)

2.2 Animate-antecedent controversy

Past studies debate whether the antecedent of MC *ziji* and *taziji* must be animate: see Tang (1989), Huang & Liu (2001) and Huang et al. (2009) for affirmative views; Pan (1997) and Charnavel & Y.-J Huang (2018) for negative views. We have observed the following tendencies.

First, when the reflexive bears a non-POSS function, such as OBJ in (1a), it needs to be bound by an animate antecedent. While the reason for this animacy requirement is not entirely transparent, we conjecture that it is related to the common observation that the syntax of particular constructions correlates with certain semantic meanings; in this case, an object-*(ta)ziji* 'self' construction in general denotes an agent-patient relation with a shared identity between the agent and patient such that the agent performs a certain action on themselves. The notion of "agenthood" often correlates with the concept of "volition", with the prototypical agent being an entity with a high level of volition (see e.g., Jackendoff, 1990 on thematic roles); thus, the prototypical agent is an animate entity. The shared identity between the agent and patient enforced by an object-*(ta)ziji* 'self' construction would in turn entail that the patient is an animate entity. We believe this syntax-semantics correlate has misled some studies to posit that *(ta)ziji* is inherently animate (e.g., Tang, 1989). We will see that this is not an accurate postulation. Before that, we shall point out that our analysis does not aim to account for cases of personification, such as (4), where *(ta)ziji* refers to an entity which is construed to be animate only in metaphorical usage:

(4) yueliang$_i$ na wuyun lai zhegai **(ta)ziji**$_i$
moon take dark.cloud come cover (C.)SELF
'The moon covered herself with dark clouds.' (MC; adapting Tang, 1989: 96)

On the other hand, when the reflexive bears a POSS function, it can encode a range of relationships commonly attested in the typology of adnominal possession (e.g., Koptjevskaja-Tamm, 2002; Haspelmath, 2017). They include ownership, body-part, kinship, part-whole relations, etc. We have observed that while most types of POSS reflexives require animate antecedents – in particular those encoding ownership, body-part, and kinship relations – POSS reflexives expressing part-whole relations (e.g., possessed quality)[4] can be bound by an inanimate antecedent, such as (5):

[4] The type of part-whole relations we focus on is the possessed quality type. Although in the typological literature, body-part relations (e.g., *my hand*) are sometimes classified as a subtype of part-whole relations, we make a distinction between them in this paper.

(5) a. [[zhangsan$_i$ zhizuo de] mei-zhang shuqian]$_j$ dou you
Zhangsan make DE every-CL bookmark all have
(ta)ziji$_{*i/j}$-(de) dute de xingzhuang
(C.)SELF-(POSS) unique DE shape
'Every bookmark that Zhangsan made has its unique shape.' (MC)

b. gongqiao zai shuimianshang touxia **(ta)ziji**-de daoying
arch.bridge on water.surface cast (C.)SELF-POSS shadow
'The arch bridge casts its shadow on the water surface'
(MC; adapted from Pan, 1997: 12)

In (5a), *(ta)ziji* relates *xingzhuang* 'shape' (part) to its inanimate antecedent *shuqian* 'bookmark' (whole). In (5b), we assume that *gongqiao* 'arch bridge' and its own shadow form a part-whole relation in a broader sense. The use of *(ta)ziji* in part-whole adnominal possession constitutes an important counterexample to the postulation that *(ta)ziji* is inherently animate.

The above generalisations are extended to the CC reflexives *jihgei* and *keuhjihgei*. We skip the data here. In section 5, we assume that the different semantic notions expressed by constructions of POSS (part-whole), POSS (non-part-whole) and non-POSS reflexives are grammaticalised such that POSS (part-whole) reflexives obey different syntactic constraints from the other reflexives.

2.3 Subject orientation

We concur with most past studies that MC reflexives *taziji* and *ziji* need to be bound by SUBJ antecedents (e.g., see Huang et al., 2009):[5]

(6) a. zhangsan$_i$ song (gei) lisi$_j$ yi-zhang **(ta)ziji**$_{i/*j}$-de xiangpian
Zhangsan give to Lisi one-CL (C.)SELF-POSS picture
'Zhangsan gives Lisi a picture of himself.'
(MC; Charnavel et al., 2017: 2341)

b. zhangsan$_i$ [cong lisi$_j$ chu] tingshuo wangwu$_k$ bu xihuan **ziji**$_{i/*j/k}$
Zhangsan [from Lisi place] hear Wangwu not like SELF
'Zhangsan heard from Lisi that Wangwu did not like him/himself.'
(MC; Pollard & Xue, 1998: 296)

c. zhangsan$_i$ gaosu lisi$_j$ **taziji**$_{i/*j}$-de shenshi
Zhangsan tell Lisi C.SELF-POSS life.story
'Zhangsan told Lisi the story of his life.' (MC; Huang & Tang, 1991: 282)

On the other hand, it has been found that the Cantonese complex reflexive *keuhjihgei* is not subject-oriented, even though subject-orientation holds for the simplex reflexive *jihgei* (see Matthews & Yip, 2013; Yip & Tang, 1998):

(7) amihng$_i$ bei-jo gafai$_j$ yat-jeung {**keuihjihgei**$_{i/j}$/**jihgei**$_{i/*j}$}-ge seung
Amihng give-PFV Gafai one-CL C.SELF/SELF-POSS photo
'Amihng has given Gafai a photo of himself.' (CC)

[5] Some studies explore the possibility of interpreting subject orientation as c-command orientation; in other words, a configurational rather than grammatical-relation concept. They often use BA constructions to discuss the possibility (e.g., Charnavel et al., 2017). However, this treatment would not explain why in double-object constructions, e.g., (6a) and (6c), where both subject and object c-command the reflexive, the antecedent is the subject but not the object. We maintain the view that subject orientation should be interpreted as a grammatical-relation concept and the idiosyncrasy of BA constructions awaits further investigation.

2.4 LD binding and blocking effects

It is well-known that the LD binding of *ziji* is susceptible to a range of blocking effects, which prevent it from being bound by a potentially available antecedent. This section summarises a few blocking effects in the literature.

First, it has been observed that an intervening 1st or 2nd person pronoun blocks a 3rd-person NP from being LD bound by *ziji* (see e.g., Tang, 1989). In (8), *Zhangsan* is blocked by *wo*/*ni* 'I/you' from being an LD antecedent of *ziji*:

(8) zhangsan$_i$ juede wo/ni$_j$ dui **ziji**$_{*i/j}$ mei xinxin
Zhangsan feel I/you to SELF no confidence
'Zhangsan feels that I/you have no confidence in myself/yourself.'
(MC; Tang, 1989: 108)

As shown below, the 1st or 2nd person pronoun does not need to be a SUBJ to cause the blocking (see e.g., Xue et al., 1994; Huang & Tang, 1991):

(9) zhangsan$_i$ gaosu wo$_j$ [lisi$_k$ dui **ziji**$_{*i/*j/k}$ mei xinxin]
Zhangsan tell me Lisi to SELF no confidence
'Zhangsan told me that Lisi has no confidence in himself.' (MC)

On the other hand, a 3rd person NP does not block the LD binding of a 1st or 2nd person NP (see e.g., Xu, 1993). This contrast is known as "person asymmetry" of blocking effects.

(10) wo$_i$ juede zhangsan$_j$ hui taoyan **ziji**$_{i/j}$
I feel Zhangsan will hate SELF
'I feel that Zhangsan will hate me/himself.' (MC)

When more than one instance of *ziji* is in the sentence, a potential LD antecedent of *ziji* blocks another potential LD antecedent further away (see e.g., Pan, 2001):

(11) John$_i$ renwei Bill$_j$ zhidao Mark$_k$ ba **ziji**1-de shu jiegei-le
John think Bill know Mark BA SELF-POSS book lend-PFV
ziji2-de pengyou
SELF-POSS friend
'John thinks Bill knows Mark lends self's book to self's friends.'
(MC; Pan, 2001: 303-304)

The available readings include (a) to (g), whereas (h) and (i) are unavailable:
(a) ziji1 = ziji2 = John; (b) ziji1 = ziji2 = Bill; (c) ziji1 = ziji2 = Mark
(d) ziji1 = Mark; ziji2 = Bill; (e) ziji1 = Mark; ziji2 = John;
(f) ziji1 = John; ziji2 = Mark (g) ziji1 = Bill; ziji2 = Mark;
*(h) ziji1 = John; ziji2 = Bill; *(i) ziji1 = Bill; ziji2 = John

It has been reported that the deictical use of a 3rd person NP causes blocking (see e.g., Huang & Liu, 2001):

(12) zhangsan$_i$ shuo ta$_j$[deictical use] qipian-le **ziji**$_{*i/j}$
Zhangsan say he deceive-PFV SELF
'Zhangsan says that he has deceived himself.' (MC; Huang & Liu, 2001: 147)

We have observed that the LD binding of CC *jihgei* is also susceptible to the above blocking effects. On the contrary, no blocking effects have been observed for locally-bound *ziji* and *jihgei*. For instance, (13) shows that the $1^{st}/2^{nd}$ person NP blocking effect does not appear for locally-bound *ziji* and *jihgei* (see also Huang & Liu 2001):[6]

(13) a. zhangsan$_i$ gaosu wo **ziji**$_i$-de mimi
Zhangsan tell I SELF-POSS secret
'Zhangsan told me about his secret.' (MC)

b. amihng$_i$ tuhng ngoh gong **jihgei**$_i$-ge beimaht
Amihng to I tell SELF-POSS secret
'Amihng told me about his secret.' (CC)

The contrast between LD- and locally-bound simplex reflexives in terms of the availability of blocking effects seems to suggest they involve different binding mechanisms. We will argue that this hypothesis is on the right track.

3 Past proposals on LD binding

Past studies centre on the binding behaviour of *ziji*. Early studies include syntactic proposals leveraging movement-based feature-agreement mechanisms: e.g., Tang (1989), Cole and Wang (1996), Huang and Tang (1991). According to these proposals, LD binding involves successive-cyclic steps of movement in LF, each forming a local binding, satisfying Principle A. The blocking effects are explained by feature agreement: during movement, traces left by *ziji* must agree with their local subject; thereby, all subjects, local and non-local, agree with *ziji* in person and number. However, the LF-movement account suffers empirical problems since there are observations which cannot be explained by feature agreement alone, such as person asymmetry, deictical blocking, and blocking by another LD antecedent.

The shortcomings of the feature-agreement proposals called for alternative accounts from a discourse-functional perspective. Huang and Liu (2001) proposed a dual system which views locally-bound *ziji* as a syntactic anaphor, conforming to Principle A, and analyses LD-bound *ziji* as a logophor to "designate the individual [...] whose speech, thoughts, feelings, or general state of consciousness are reported" (Clements, 1975: 141). The antecedent of an LD-bound *ziji* is considered the "speaker" or "virtual speaker" (e.g., thinker, feeler, knower, experiencer) of the complement clause where the reflexive is found. The blocking effects are explained functionally as effects of a perceptual strategy to avoid perspective conflicts. For details on how all the blocking effects can be explained by this perceptual strategy, please refer to Huang and Liu (2001: 161–165). As a summary, the blocking effect of a $1^{st}/2^{nd}$ person pronoun, e.g., (8), is induced because the $1^{st}/2^{nd}$ person pronoun anchors the perspective to the external speaker/addressee, while *ziji* as a logophor designates the perspective of the internal speaker (i.e., matrix subject). Consequently, there is a perspective conflict, blocking LD binding. In comparison, when the matrix subject is a $1^{st}/2^{nd}$ person pronoun, as in (10), it anchors the perspective to that of the external speaker, but in this case *ziji* as a

[6] If the blocking effects were in place, we would expect the interpretations of *ziji*$_i$ (MC) and *jihgei*$_i$ (CC) to be unavailable (see (9)); in other words, the speaker would have to use non-reflexive *ta* (MC) or *keuih* (CC) to refer to the matrix subject. The fact that *ziji*$_i$ and *jihgei*$_i$ are available readings entails that there is no blocking effect.

logophor also refers to this external speaker; in other words, there is no perspective conflict. The deictical blocking in (12) is similarly explained by how the deictical NP is anchored to the perspective of the external speaker; thus, the LD binding of *ziji* by the internal speaker *Zhangsan* is ruled out. The blocking by another LD antecedent in (11) is likewise accounted for as an effort to avoid conflicting perspectives caused by different logophoric *ziji* which anchors the utterance to varying perspectives.

Huang and Liu (2001: 156) provide a logophoric theory incorporating Sells's (1987) taxonomy which classifies the antecedents of logophors into three primitive roles: SOURCE (the intentional agent of the communication), SELF (the one whose mental state or attitude the proposition describes), and PIVOT (the one with respect to whose time-space location the content of the proposition is evaluated). They explore the possibility that these roles can be reduced to the notion of *de se* in the sense of Chierchia (1989) with the assumption that SOURCE and SELF satisfy a stronger *de se* requirement than PIVOT since it is observed that PIVOT can be licensed once the external speaker takes the perspective or empathises with the internal protagonist (see also Cole et al., 2001; Pan, 2001). Sells's classification has a useful application in capturing variations among Sinitic languages, as we will see in section 5 that LD binding in CC must be licensed by SOURCE or SELF, but not PIVOT.

Huang and Liu's proposal seems to provide a more satisfactory account for LD binding compared to earlier studies. They hypothesise that logophoricity can be integrated into syntax by postulating SourceP, SelfP, and PivotP as CP-type functional phrases in LF representations. However, from a theory-internal perspective, as admitted by Huang and Liu (2001: 178), their formalism of LF syntax does not in itself capture the blocking effects. From the present perspective, as remarked by Sperlich (2019: 23), Huang and Liu's machinery is not supported by current Minimalist theory.

Recent Minimalist studies on Chinese anaphora have regained interest in agreement-based proposals, amid crosslinguistic proposals (e.g., Reuland, 2011) which posit Agree to be the main machinery in binding relations while abandoning Principle A. One of these proposals is Giblin's (2016) Agree-based account of LD *ziji*. Giblin analyses *ziji* to be ϕ-feature deficient and syntactically bound using the mechanism of Contiguous Agree. This system can explain blocking caused by unmatched person values, e.g., (8). Reuland et al. (2020) also incorporate Giblin's agreement system in their proposal. However, like the earlier proposals leveraging movement-based agreement, Giblin's agreement-based account is not sufficient in explaining the wide range of blocking effects, especially those unrelated to issues of feature agreement.

After reviewing the above proposals, we conclude that agreement-based accounts for LD binding suffer empirical problems, and despite the inadequacies of Huang and Liu (2001), a logophoric account is preferred based on empirical considerations. Although one may argue that it is possible to produce a nonuniform proposal for LD binding embracing both agreement and logophoric accounts, as Giblin (2016) suggests, we question, by Occam's razor, why it is necessary to introduce an additional agreement system if a logophoric account is already sufficient. We will devise an LFG proposal where LD binding is explained logophorically. We will discuss how the (grammaticised) notions of SOURCE, SELF, and PIVOT are formally introduced into our syntactic structure (f-structure) in a mathematically well-defined manner as well as how our system can potentially capture blocking effects by suspension of logophoric constraints, which cannot be modelled in Huang and Liu's (2001) LF syntax.

4 Grammaticised logophoricity

In the following sections, we will present our LFG binding system, where we preserve the insight of Huang and Liu (2001) that LD binding in MC is logophoric binding.[7] We shall extend their insight to explore how a logophoric reflexive is formally bound in its logophoric domain[8] which is created by a logocentric predicate.[9] Before that, we shall address one more issue: do MC and CC demonstrate "pure" logophoricity or "grammaticised" logophoricity?

Logophoricity is in itself a pragmatic concept. Cross-linguistically, languages exhibit varying degrees of logophoricity. According to Culy (1994), pure logophoric languages are those containing special morphological and/or syntactic forms employed only in logophoric domains. For example, the logophoric pronouns in Babungo are to be used only in logophoric domains but not in other contexts. They are considered "true" logophoric pronouns and Babungo is regarded as a pure logophoric language. On the other hand, as discussed by Huang and Liu (2001), the local binding of *ziji* is unrelated to logophoricity. From this perspective, the logophoric use of *ziji* in LD binding is an extended use of the reflexive. Neither are *ziji* and *jihgei* "true" logophoric pronouns on a par with Babungo's logophoric pronouns, nor are MC and CC pure logophoric languages. In fact, Culy (1994) observes that while many languages show degrees of logophoricity, pure logophoric languages are only found in Africa.

We argue MC and CC exhibit grammaticised logophoricity. To elaborate, we build on Dalrymple's (2015) argumentation in her study of Yag Dii where she holds that Yag Dii exhibits grammaticised logophoricity. She argues, citing Clements (1975: 141), that the antecedent of a "true" logophoric pronoun is the individual "whose speech, thoughts, feelings, or general state of consciousness are reported" (see also Sells, 1987); in other words, the antecedent is identified by semantic/pragmatic means, not syntactically. Therefore, if one finds that the identification of a logophoric antecedent has syntactic requirements, one can conclude that the language displays grammaticised logophoricity in contrast to pure logophoricity. One of the important pieces of evidence Dalrymple provides for Yag Dii is that the antecedent of a BI (logophoric) pronoun must be a syntactic SUBJ. Therefore, Yag Dii demonstrates grammaticised logophoricity. As mentioned earlier, the antecedent of *ziji* (MC) and *jihgei* (CC) also has a SUBJ requirement, as shown

[7] We will extend this mechanism to the LD binding in CC.

[8] Following Y. Huang (2000: 183), the concept of "logophoric domain" can be defined pragmatically or syntactically. Pragmatically, a logophoric domain is a stretch of discourse where the perspective of the internal protagonist is being represented. Syntactically, a logophoric domain begins in a clause subordinate to the one where the logophoric antecedent is identified.

[9] Y. Huang (2000: 184) explains that there are two common forms of "logocentric licensers": (i) logocentric predicates (ii) logocentric complementizers. We discuss in section 5 the types of logocentric predicates assumed in MC and CC, but logocentric complementizers are not found in these languages.

in (6) and (7). Based on this evidence, we conclude that MC and CC show grammaticised logophoricity. In fact, as remarked by Dalrymple, this kind of SUBJ requirement is commonly found among (partially) grammaticised logophoric systems, including Icelandic (Bresnan, 2016; Sells, 1987):

(14) *Eg heyrði fra Joni$_i$ að Maria hefði boðið **ser**$_i$
I heard from John that Maria had-SBJN invited him
'I heard from John that Maria had invited him (John).'(Maling, 1984: 233)

Bresnan (2016: 266) attributes the ungrammaticality of (14) to the violation of the SUBJ requirement that applies to the logophor *ser*. We can contrast the grammaticised logophoric systems of MC/CC, Yag Dii, and Icelandic with the pure logophoric system of Ewe where there is no SUBJ requirement:

(15) Kɔmi xɔ agbale tso Kofi$_i$ gabɔ be wo-a-va me kpe na **ye**$_i$
Kwami get letter from Kofi side that PRO-T-come cast block for LOG
'Kwami got a letter from Kofi$_i$ saying that he should come cast blocks for him$_i$.' (Clements, 1975: 160)

The observation that logophoric binding in MC and CC is a grammaticised one has important implications on how we formalise the binding. Since the antecedent cannot be defined in purely pragmatic terms, at least some of the logophoric constraints need to be stated for the syntactic structure (f-structure).

5 Our LFG constraint-based binding system

We analyse LD-bound *ziji* and *jihgei* as grammaticised logophors. Both are subject to the same blocking effects. Conversely, we have not observed blocking effects for locally-bound *ziji* and *jihgei*, and neither do they need to comply with any *de se* requirements. We agree with Huang and Liu (2001) that the local binding of *ziji* (and *jihgei*) should be modelled differently from LD binding. Our constraint-based binding system contains two key components:

Component 1: Anaphoric Binding
- Local binding of complex reflexives – *taziji* (MC) and *keuhjihgei* (CC)
- Local binding of simplex reflexives – *ziji* (MC) and *jihgei* (CC)

Component 2: Logophoric Binding
- LD binding of simplex reflexives – *ziji* (MC) and *jihgei* (CC)

As a preview, in (16), we provide a schematic overview of the lexical entries of the reflexives. It illustrates how we organise the constraints for anaphoric binding and different types logophoric binding, namely SOURCE-binding, SELF-binding, PIVOT-binding, and binding by the discourse speaker. *Ziji* and *jihgei* contain constraints for both anaphoric and logophoric binding organised in a disjunctive manner, whereas *taziji* and *keuhjihgei* are only capable of anaphoric binding. We use 'REFL-PRO' as the semantic form of an anaphoric reflexive, and 'LOG-PRO' as that of a logophoric reflexive. Later, our anaphoric binding constraints will use the FN attribute (Dalrymple et al., 2019: 154) to refer to this semantic form as we delimit the binding domain.[10]

[10] A common LFG notation for reflexives is to use 'PRO' as the PRED value together with the attribute-value pair <PRONTYPE, REFL> (see e.g., Dalrymple et al., 2019).

(16) Schematic overview of the lexical entries of the reflexives:

<table>
<tr>
<td>Lexical entry of taziji (MC):
(↑PRED) = ‘REFL-PRO’
Constraints for anaphoric binding</td>
<td rowspan="2">Lexical entry of ziji (MC):
{ (↑PRED) = ‘REFL-PRO’
Constraints for anaphoric binding (local-binding)
| (↑PRED) = ‘LOG-PRO’
{Constraints for SOURCE-/SELF-binding
| Constraints for PIVOT-binding
| Constraints for reference to discourse speaker }}</td>
<td rowspan="2">Lexical entry of jihgei (CC):
{ (↑PRED) = ‘REFL-PRO’
Constraints for anaphoric binding (local-binding)
| (↑PRED) = ‘LOG-PRO’
{Constraints for SOURCE-/SELF-binding
| Constraints for reference to discourse speaker }}</td>
</tr>
<tr>
<td>Lexical entry of keuhjihgei (CC):
(↑PRED) = ‘REFL-PRO’
Constraints for anaphoric binding</td>
</tr>
</table>

5.1 Anaphoric-binding component: local binding

LFG assumes that binding relations are stated in f-structural terms, and posits that binding requirements should be specified lexically instead of on a language-by-language or universal basis. (17) is the general equation of anaphoric-binding, adapted from Dalrymple (1993, 2015), to be included in the lexical entries of the reflexives.

(17) $$(\uparrow_\sigma \text{ANT}) = ((\underset{\text{OFFPATH}}{\text{GF*}} \quad \text{GF}_{\text{pro}} \uparrow)\ \text{ANTE})_\sigma$$

$\uparrow$ is the f-structure of the reflexive, $\uparrow_\sigma$ is the semantic structure corresponding to $\uparrow$, and ($\uparrow_\sigma$ ANT) refers to the antecedent. (GF* GF_{pro} $\uparrow$) is an inside-out path reaching the binding domain within which the antecedent is found. LFG assumes that the antecedent f-commands the reflexive (Dalrymple, 1993). We will see that this does not hold for MC and CC. ANTE is an outside-in path from the binding domain encoded by (GF* GF_{pro} $\uparrow$) to the antecedent. The binding domain is delimited by the off-path constraint OFFPATH acting on the path GF* to limit the reflexive's search for an antecedent. Cross-linguistically, there are four common binding domains (Dalrymple, 1993): Co-argument Domain, Minimal Complete Nucleus, Minimal Finite Domain, and Root Domain. We will see that they do not capture binding in MC and CC. Given the equation (17), our task is to derive ANTE and OFFPATH for the reflexives in MC and CC. We have identified four characteristics that are important for deriving ANTE and OFFPATH:

i. Does the antecedent need to be a SUBJ? (Section 2.3)
ii. Can the antecedent be further embedded within an f-commanding GF?
iii. Is there any animacy restriction on the antecedent? (Section 2.2)
iv. If the answer to (iii) is "yes", does the animacy restriction of the antecedent have any implications on the binding domain?

Our discussion below will focus on *taziji*. We will discuss how the constraints of *taziji* can be adapted for *keuhjihgei* and the local binding of *ziji* and *jihgei*.

5.1.1 Constraints for reflexives taking non-POSS functions

As discussed in section 2.2, when the reflexive bears a non-POSS function, it has to be bound by an animate antecedent. In other words, there is animacy restriction on the antecedent. Moreover, the antecedent can be further embedded within a f-commanding GF. (18) contains examples of *(ta)ziji* adapted from Tang (1989: 100). As discussed in section 2.3, *(ta)ziji* is SUBJ-oriented. The antecedent of *(ta)ziji* can be further embedded within SUBJ:

(18) a. [[zhangsan$_i$-de baba$_j$-de] aoman]$_k$ hai-le **(ta)ziji**$_{*i/j/*k}$
Zhangsan-POSS dad-POSS arrogance harm-PFV (C.)SELF
'The arrogance of Zhangsan's dad has harmed himself.' (MC)

b. [zhangsan$_i$ nayan zuo]$_j$ dui **(ta)ziji**$_{i/*j}$ bu li
Zhangsan that.way do to (C.)SELF not advantageous
'That Zhangsan behaved in such a manner did him no good.' (MC)

c. [zhangsan$_i$ zuoshi xiaoxin de taidu]$_j$ jiu-le **(ta)ziji**$_{i/*j}$
Zhangsan do.thing careful DE attitude save-PFV (C.)SELF
'Zhangsan's cautious attitude saved him.' (MC)

The ANTE path for (18a) is [SUBJ POSS], for (18b) is [SUBJ SUBJ], and for (18c) [SUBJ ADJ ∈ SUBJ]. We generalise the ANTE path for *(ta)ziji* to be (19), where we have added animacy restrictions on the GF along the ANTE path:

(19) ANTE_(TA)ZIJI ≡

$$\{ \underset{(\to \text{ANIMATE}) =_c +}{\text{SUBJ}} \mid \underset{(\to \text{ANIMATE}) =_c -}{\text{SUBJ}} \ \underset{(\to \text{ANIMATE}) =_c +}{\text{POSS}} \mid \text{SUBJ}\ \underset{(\to \text{ANIMATE}) =_c +}{\text{SUBJ}} \mid \underset{(\to \text{ANIMATE}) =_c -}{\text{SUBJ}}\ \text{ADJ}\in\ \underset{(\to \text{ANIMATE}) =_c +}{\text{SUBJ}} \}$$

Because the antecedent can be embedded within SUBJ, it does not need to f-command the reflexive. In CC, these patterns are also observed for *keuhjihgei* and locally-bound *jihgei*. Nevertheless, *keuhjihgei* is not SUBJ-oriented. Therefore, while the ANTE path for the locally-bound *jihgei* is the same as MC (*ta)ziji*, we have removed the SUBJ requirement for *keuhjihgei*:

(20) a. **ANTE_(TA)ZIJI_JIHGEI** ≡

$$\{ \underset{(\to \text{ANIMATE}) =_c +}{\text{SUBJ}} \mid \underset{(\to \text{ANIMATE}) =_c -}{\text{SUBJ}} \ \underset{(\to \text{ANIMATE}) =_c +}{\text{POSS}} \mid \text{SUBJ}\ \underset{(\to \text{ANIMATE}) =_c +}{\text{SUBJ}} \mid \underset{(\to \text{ANIMATE}) =_c -}{\text{SUBJ}}\ \text{ADJ}\in\ \underset{(\to \text{ANIMATE}) =_c +}{\text{SUBJ}} \}$$

b. **ANTE_KEUHJIHGEI** ≡

$$\{ \underset{(\to \text{ANIMATE}) =_c +}{\text{GF}} \mid \underset{(\to \text{ANIMATE}) =_c -}{\text{GF}} \ \underset{(\to \text{ANIMATE}) =_c +}{\text{POSS}} \mid \text{GF}\ \underset{(\to \text{ANIMATE}) =_c +}{\text{SUBJ}} \mid \underset{(\to \text{ANIMATE}) =_c -}{\text{GF}}\ \text{ADJ}\in\ \underset{(\to \text{ANIMATE}) =_c +}{\text{SUBJ}} \}$$

We now derive the off-path constraint OFFPATH for *taziji*. First, non-POSS *taziji* must be bound by the "closest" animate antecedent (see section 2):

(21) a. zhangsan$_i$ shuo [lisi$_j$ renwei [wangwu$_k$ zeguai **taziji**$_{*i/*j/k}$]]
Zhangsan say Lisi think Wangwu blame C.SELF
'Zhangsan says Lisi thinks Wangwu blames himself.' (MC)

b. zhangsan$_i$ shuo [lisi$_j$ renwei [**taziji**$_{*i/j}$ hen qinfen]]
Zhangsan say Lisi think C.SELF very be.diligent
'Zhangsan says Lisi thinks he is very diligent.' (MC)

The binding domain for (21a) is the f-structure (OBJ↑) and that for (21b) is (COMP SUBJ↑). With *taziji* being an embedded SUBJ, the Minimal Complete Nucleus cannot be the correct binding domain. Otherwise, the domain for (21b) would be the f-structure (SUBJ↑), wherein there is no valid antecedent. We may formulate the off-path constraint as ¬[(→SUBJ PRED FN) ≠ REFL-PRO]. The constraint states that none of the attributes corresponding to the path GF* of (17) may contain a non-reflexive SUBJ. Thus, in (21b), while (COMP SUBJ↑) is valid with COMP having the SUBJ *taziji*, (COMP COMP OBJ↑) is invalid as the outermost COMP contains the non-reflexive SUBJ *Lisi*. Thus, *taziji* is bound by *Lisi* but not *Zhangsan*.

Additional evidence suggests that the animacy requirement of non-POSS *taziji* has implications on the binding domain such that the constraint ¬[(→SUBJ PRED FN) ≠ REFL-PRO] needs to be further revised:

(22) John$_i$ shuo [na-ben shu$_j$ hai-le **taziji**$_{i/*j}$]
John say that-CL book harm-PFV C.SELF
'John says that book has harmed him.' (MC; Pan & Hu, 2003: 153)

The binding domain for (22) is the f-structure at the end of (COMP OBJ↑) such that with *shu* 'book' being an inanimate entity, instead of resulting in ungrammaticality, *taziji* will continue to search for its antecedent outside the embedded clause.[11] We revise the off-path constraint as (23):

(23) ¬[(→ SUBJ PRED FN) ≠ REFL-PRO & (→ SUBJ ANIMATE) = +]

The revised constraint ensures that when the path GF* of (17) contains an inanimate SUBJ, the reflexive's search for its antecedent will continue to an outer f-structure. In (22), (COMP OBJ↑) is the valid binding domain: although COMP does not fulfil the first part of the disjunctive rule,[12] it satisfies the second part by not containing an animate SUBJ.

The same off-path constraint is applied to non-POSS locally-bound *ziji*, non-POSS *keuhjihgei* and non-POSS locally-bound *jihgei*.

5.1.2 Constraints for reflexives taking POSS functions

As discussed in section 2.2, we have observed that most types of POSS reflexives (except part-whole type) require animate antecedents. These types of *taziji*, locally-bound *ziji*, *keuhjihgei* and locally-bound *jihgei* share the same ANTE as their non-POSS counterparts. (24) is an example of kinship POSS:

(24) zhangsan$_i$ shuo [na-fan hua$_j$ shanghai-le [**taziji**$_{i/*j}$-de mama]]
Zhangsan say that-CL words hurt-PFV C.SELF-POSS mum
'Zhangsan says that those words have hurt his mum.' (MC)

The reflexive is a POSS embedded within another GF (e.g., OBJ). We add this observation to the off-path constraint (23) and revise it as (25). (25) is applicable to all locally-bound reflexives in MC and CC bearing non-POSS GF or POSS GF (except the part-whole type).

(25) **OFFPATH** ≡ ¬[(→{SUBJ|POSS} PRED FN) ≠ REFL-PRO & (→SUBJ ANIMATE)= +]

As discussed in section 2.2, POSS reflexives expressing a part-whole relation (POSS$_{\text{part-whole}}$) can be bound by an inanimate antecedent:

(26) zhangsan$_i$ shuo [na-ben shu$_j$ you [**taziji**$_{*i/j}$-de tese]]
Zhangsan say that-CL book have C.SELF-POSS feature
'Zhangsan says that book has its own features.' (MC)

---

[11] We can compare with the situation in English where the animacy requirement of a reflexive does not have implications on its binding domain. *Himself* in (i) results in ungrammaticality:
(i) Peter$_i$ said the book has harmed {*himself$_i$ / him$_i$}.

[12] By De Morgan's Law ¬[P & Q] ⇔ ¬P ∨ ¬Q, the constraint is equivalent to the following disjunctive rule: ¬[(→SUBJ PRED FN) ≠ REFL-PRO] ∨ ¬[(→SUBJ ANIMATE) = +].

The binding domain for (26) is (OBJ POSS↑). OFFPATH does not predict the correct result since ¬[(→ SUBJ ANIMATE) = +] would result in the wrong admission of COMP in the binding domain, predicting *Zhangsan* to be the binder. What it requires is OFFPATH_POSS$_{\text{part-whole}}$ as stated in (27), which has removed the disjunctive option of OFFPATH designed for the animacy requirement:

(27) **OFFPATH_POSS$_{\text{part-whole}}$** $\equiv \neg[(\rightarrow\{\text{SUBJ}|\text{POSS}\}\ \text{PRED FN}) \neq \text{REFL-PRO}]$

(26) shows that when a POSS$_{\text{part-whole}}$ reflexive is bound by an f-commanding GF, there is no animacy requirement on the binder GF. However, we have observed that when the binder GF becomes non-f-commanding because it is further embedded within an f-commanding GF, the embedded binder GF is restricted to be animate, as shown in (28):

(28) a.

[xiaoming$_i$-de	hua]$_j$	zhanxianchu	**taziji**$_{i/*j}$-de	xingge
Xiaoming-POSS	word	show	C.SELF-POSS	personality

‘Xiaoming’s words have shown his personality.’ (MC)

b.

[zhuozishang$_i$-de	diaochua]$_j$	you	**taziji**$_{*i/j}$-de	dute	fengge
table-POSS	carving	have	C.SELF-POSS	unique	style

‘The carvings of the table have their unique style.’ (MC)

In (28b), POSS$_{\text{part-whole}}$ *taziji* cannot be bound by a non-f-commanding inanimate GF, although binding by a non-f-commanding animate GF is licensed in (28a). Thus, we posit the ANTE path for POSS$_{\text{part-whole}}$ of different reflexives to be (29), which is modified from (20):

(29) a. **ANTE_(TA)ZIJI_JIHGEI_POSS$_{\text{part-whole}}$** $\equiv$

$$\{\ \text{SUBJ}\ \Big|\ \underset{(\rightarrow\text{ANIMATE}) =_c -}{\text{SUBJ}}\ \ \underset{(\rightarrow\text{ANIMATE}) =_c +}{\text{POSS}}\ \Big|\ \text{SUBJ}\ \ \underset{(\rightarrow\text{ANIMATE}) =_c +}{\text{SUBJ}}\ \Big|\ \underset{(\rightarrow\text{ANIMATE}) =_c -}{\text{SUBJ}}\ \ \text{ADJ}\in\ \ \underset{(\rightarrow\text{ANIMATE}) =_c +}{\text{SUBJ}}\ \}$$

b. **ANTE_KEUHJIHGEI_ POSS$_{\text{part-whole}}$** $\equiv$

$$\{\ \text{GF}\ \Big|\ \underset{(\rightarrow\text{ANIMATE}) =_c -}{\text{GF}}\ \ \underset{(\rightarrow\text{ANIMATE}) =_c +}{\text{POSS}}\ \Big|\ \text{GF}\ \ \underset{(\rightarrow\text{ANIMATE}) =_c +}{\text{SUBJ}}\ \Big|\ \underset{(\rightarrow\text{ANIMATE}) =_c -}{\text{GF}}\ \ \text{ADJ}\in\ \ \underset{(\rightarrow\text{ANIMATE}) =_c +}{\text{SUBJ}}\ \}$$

5.1.3 Summarising anaphoric-binding constraints

Our analysis assumes a different GF for possessive reflexives that encode part-whole relations, which we term as “POSS$_{\text{part-whole}}$”. Empirically, these reflexives illustrate different binding patterns; therefore, they embody a different set of binding constraints than reflexives taking non-POSS functions and those bearing POSS functions indicating other types of possessive relations. In this paper, we use POSS to represent any type of possessor, including part-whole ones. We leave for future research the theoretical status of POSS$_{\text{part-whole}}$ in LFG corresponding to a wealth of typological research on adnominal possession, as well as the question of whether we should sub-classify other types of POSS as different grammatical functions (e.g., POSS$_{\text{kinship}}$, POSS$_{\text{ownership}}$) in MC and CC. (30) summarises our anaphoric-binding constraints for each reflexive. The constraints are written as a disjunctive rule. The first disjunctive option targets at the situation where the reflexive takes a non-POSS$_{\text{part-whole}}$ function, whereas the second disjunctive option is for the situation when the reflexive bears the POSS$_{\text{part-whole}}$ function.

(30) a. Anaphoric-binding constraints for *taziji*, locally-bound *ziji*, and *jihgei*:

$$\{\ \neg(\mathrm{POSS}_{\text{part-whole}}\uparrow)\ \Rightarrow\ (\uparrow_\sigma\ \mathrm{ANT}) = ((\underset{\textbf{OFFPATH}}{\mathrm{GF}^*}\ \ \mathrm{GF}_{pro}\uparrow)\ \ \textbf{ANTE\_(TA)ZIJI\_JIHGEI})_\sigma$$

$$|\ (\mathrm{POSS}_{\text{part-whole}}\uparrow)\ \Rightarrow\ (\uparrow_\sigma\ \mathrm{ANT}) = ((\underset{\textbf{OFFPATH\_POSS}_{\textbf{part-whole}}}{\mathrm{GF}^*}\ \ \mathrm{GF}_{pro}\uparrow)\ \ \textbf{ANTE\_(TA)ZIJI\_JIHGEI\_POSS}_{\textbf{part-whole}})_\sigma\}$$

b. Anaphoric-binding constraints for *keuhjihgei*:

$$\{\ \neg(\mathrm{POSS}_{\text{part-whole}}\uparrow)\ \Rightarrow\ (\uparrow_\sigma\ \mathrm{ANT}) = ((\underset{\textbf{OFFPATH}}{\mathrm{GF}^*}\ \ \mathrm{GF}_{pro}\uparrow)\ \ \textbf{ANTE\_ KEUHJIHGEI})_\sigma$$

$$|\ (\mathrm{POSS}_{\text{part-whole}}\uparrow)\ \Rightarrow\ (\uparrow_\sigma\ \mathrm{ANT}) = ((\underset{\textbf{OFFPATH\_POSS}_{\textbf{part-whole}}}{\mathrm{GF}^*}\ \ \mathrm{GF}_{pro}\uparrow)\ \ \textbf{ANTE\_ KEUHJIHGEI\_POSS}_{\textbf{part-whole}})_\sigma\ \}$$

5.2 Logophoric-binding component: long-distance binding

This section develops constraints for logophoric binding. (31) is our general binding equation modified from Dalrymple (2015: 1116):

(31) $$(\uparrow_\sigma\ \mathrm{ANT}) = ((\underset{(\rightarrow \mathrm{LOG})}{\mathrm{GF}_{log}}\ \ \underset{\neg(\rightarrow \mathrm{LOG})}{\mathrm{GF}_{pro}{}^*}\ \uparrow)\ \mathrm{PATH})_\sigma$$

In line with Dalrymple (2015), we posit a LOG feature in the f-structure to mark the logophoric domain where the logophoric reflexive, *ziji* or *jihgei*, must appear. LOG is not the antecedent. We will see how this feature is introduced by a logocentric predicate.[13] We define GF_{log} as (32), which is essentially a clausal function:

(32) $\mathrm{GF}_{log} \equiv \{\mathrm{COMP}\ |\ \mathrm{XCOMP}\}$

PATH is the outside-in path from the f-structure immediately containing GF_{log} to the antecedent. In most circumstances, as we shall see, PATH is a single SUBJ. We will expand on Dalrymple's (2015) proposal by integrating insights from Huang and Liu (2001) as we develop constraints to differentiate the different types of logophoric binding with reference to Sells's (1987) logophoric taxonomy: SOURCE, SELF, and PIVOT binding.

5.2.1 SOURCE as antecedent

We adopt the definition of Huang and Liu (2001: 156) that a SOURCE-type antecedent is "the intentional agent of communication." We assume the logophoric domain for SOURCE-binding is marked by a verb of speech functioning as the logocentric predicate. In (33), our coreferential indexation only concerns the logophoric interpretation (i.e., LD binding).

(33) a. [$_{\text{SOURCE}}$ xiaoming]$_i$ shuo xiaomei hen touyan **ziji**$_i$
Xiaoming say Xiaomei very hate SELF
'Xiaoming says Xiaomei hates him very much.' (MC)

b. [$_{\text{SOURCE}}$ amihng]$_i$ wah ameih hou jang **jihgei**$_i$
Amihng say Ameih very hate SELF
'Amihng says Ameih hates him very much.' (CC)

[13] See footnote 9 and Y. Huang (2000) for more information on logocentric predicates.

The following is the lexical entry of the logocentric predicate *say*:

(34) *shuo/wah* V (↑PRED) = ‘SAY < SUBJ, COMP >’
((↑COMP LOG) = +
(↑SUBJ LOG-ANT) = SOURCE)

(↑COMP LOG) = + marks the logophoric domain as LOG +. The logophoric domain is the complement clause of *say*. LOG-ANT is a feature added to the f-structure of SUBJ to mark it as a SOURCE-type antecedent. The inclusion of logophoric information in the syntactic f-structure corresponds to our analysis that MC and CC illustrate grammaticised logophoricity (see section 4).

The two logophoric constraints are marked as optional. Their optionality is governed by discourse-logophoric conditions (e.g., perspectivity, *de se* attitudes) that we discussed previously with reference to Huang and Liu’s (2001) explanation of the blocking effects. A blocking effect occurs when there is illicit reference to a potential LD antecedent. As such, in our formal system, blocking is understood as the suspension of the two logophoric constraints, thereby causing the absence of the essential logophoric domain needed for LD binding. To formally model the suspension mechanism, we need to relate the two f-structural constraints to a formal representation of discourse where we state the various discourse conditions (e.g., conditions to avoid perspective conflicts). This goes beyond the scope of our paper. However, we now see how blocking effects can potentially be resolved in our constraint-based model, which is an advantage over Huang and Liu’s (2001) derivational approach which, as admitted by Huang and Liu (2001: 178), cannot in itself capture blocking effects. Before discussing the binding constraints in the lexical entries of *ziji* and *jihgei*, we will first examine SELF-binding.

5.2.2 SELF as antecedent

We subscribe to Huang and Liu’s (2001: 156) definition of SELF-type antecedent that refers to “the one whose mental state/attitude the proposition describes”. Like SOURCE-binding, we assume that a logocentric predicate (e.g., verb of feeling/thinking) marks the logophoric domain.

(35) a. [$_{SELF}$ xiaoming]$_i$ hen gaoxing xiaomei xihuan **ziji**$_i$
Xiaoming very be.happy Xiaomei like SELF
‘Xiaoming is very happy that Xiaomei likes him.’ (MC)
b. [$_{SELF}$ amihng]$_i$ hou hoisam ameih jungyi **jihgei**$_i$
Amihng very be.happy Ameih like SELF
‘Amihng is very happy that Ameih likes him.’ (CC)

(36) *gaoxing/hoisam* V (↑PRED) = ‘BE.HAPPY < SUBJ, COMP >’
((↑COMP LOG) = +
(↑SUBJ LOG-ANT) = SELF)

LOG-ANT is a feature added to the f-structure of SUBJ to mark it as a SELF-type antecedent. In most circumstances, a logophoric antecedent is SUBJ, but there are logocentric predicates that optionally allow POSS embedded within SUBJ to be the antecedent, e.g., *biaoshi* ‘indicate’. By default, the antecedent of a logophor is an animate entity. We have observed that when both SUBJ and embedded POSS are animate, the logophoric antecedent is SUBJ.

(37) a. [[$_{\text{SELF}}$ zhangsan]$_i$-de baogao]$_j$ biaoshi tamen dui **ziji**$_{i/*j}$ mei xinxin
Zhangsan-POSS report indicate they to SELF no confidence
'Zhangsan's report indicates that they had no confidence in him.'
(MC; Huang & Liu, 2001: 187)

b. [$_{\text{SELF}}$ [zhangsan]$_i$-de mama]$_j$ biaoshi tamen dui **ziji**$_{*i/j}$ mei xinxin
Zhangsan-POSS mum indicate they to SELF no confidence
'Zhangsan's mum indicates that they had no confidence in her.'

(38) *biaoshi* V $(\uparrow\text{PRED}) = \text{'INDICATE} < \text{SUBJ, COMP} >\text{'}$
$(\ (\uparrow\text{COMP LOG}) = +$
$\{\ (\uparrow\text{SUBJ ANIMATE}) = + \Rightarrow (\uparrow\text{SUBJ LOG-ANT}) = \text{SELF}$
$\mid [(\uparrow\text{SUBJ ANIMATE}) = - \ \&\ (\uparrow\text{SUBJ POSS ANIMATE}) = +] \Rightarrow$
$(\uparrow\text{SUBJ POSS LOG-ANT}) = \text{SELF}\ \}\ )$

We now examine the constraints in the lexical entries of *ziji* and *jihgei* which are responsible for SOURCE and SELF binding:

(39)
$$\{(\uparrow_\sigma \text{ANT}) = ((\underset{(\rightarrow\text{LOG})}{\text{GF}_{log}}\quad \underset{\neg(\rightarrow\text{LOG})}{\text{GF}_{pro}{}^{*}\uparrow})\quad \underset{\substack{(\rightarrow\text{ANIMATE}) =_c + \\ (\rightarrow\text{LOG-ANT}) =_c \{\text{SOURCE} \mid \text{SELF}\}}}{\text{SUBJ}}\quad)_\sigma$$
$$\mid (\uparrow_\sigma \text{ANT}) = ((\underset{(\rightarrow\text{LOG})}{\text{GF}_{log}}\quad \underset{\neg(\rightarrow\text{LOG})}{\text{GF}_{pro}{}^{*}\uparrow})\quad \underset{(\rightarrow\text{ANIMATE}) =_c -}{\text{SUBJ}}\quad \underset{\substack{(\rightarrow\text{ANIMATE}) =_c + \\ (\rightarrow\text{LOG-ANT}) =_c \{\text{SOURCE} \mid \text{SELF}\}}}{\text{POSS}}\quad)_\sigma\ \}$$
$$\neg((\text{SUBJ (POSS)} \uparrow)\ \text{LOG})$$

The disjunctive constraints stipulate that *ziji* is bound by an antecedent, which is SUBJ or embedded POSS, found in the f-structure immediately containing GF_{log}, subject to the LOG-ANT feature and animacy requirements. Thus, SOURCE/SELF binding is achieved by the interaction of the lexical constraints of a logocentric predicate with those of a reflexive. $\neg((\text{SUBJ (POSS)} \uparrow)\ \text{LOG})$ prevents SOURCE or SELF-bound reflexive from appearing as SUBJ (or embedded POSS) in the highest clause within the logophoric domain. So, we consider the local binding of e.g., *Xiaoming$_i$ says [ziji$_i$-POSS friend not go]* as anaphoric binding.[14] As we take blocking effects as independent evidence for logophoric binding, our stance is empirically corroborated by the blocking-effect asymmetry between local and LD binding that local binding is not susceptible to blocking; thus a lack of independent evidence to motivate logophoric binding (section 2.4).[15] The constraint does not affect the LD logophoric binding of reflexives e.g., *Xiaoming$_i$ says [Zhangsan likes ziji$_i$]*.

5.2.3 PIVOT as antecedent

We adopt Huang and Liu's (2001: 156) definition of PIVOT antecedent as "the one with respect to whose time-space location the content of the proposition is

[14] This treatment is in a sense similar to that of Reinhart and Reuland (1993) where anaphoric binding is prioritised over logophoric binding, although we approach binding from a different analytical tradition and our concept of logophoric binding is different from theirs.
[15] In general, we adopt a cautious approach regarding when to propose logophoric binding. We maintain the view that in a language where there are no morphologically distinct forms as logophors, if one wants to argue that an anaphoric form has a dual identity as both anaphor and logophor, one must identify strong empirical evidence to prove its logophoric identity. In MC/CC, the strongest evidence for LD *ziji*/*jihgei* comes from the blocking effects, which would be difficult to explain without the logophoric account.

evaluated". There are differences between MC and CC in that PIVOT does not license logophoric binding in CC. A similar result was reported for a Teochew variety spoken in Singapore where PIVOT does not license binding (Cole et al., 2001).

(40) a. zhangsan lai kan **ziji**$_i$ de shihou, [$_{\text{PIVOT}}$ lisi$_i$] zheng zai kan shu
Zhangsan come see SELF DE moment Lisi now at read book
'Lisi was reading when Zhangsan came to visit him.'
(MC, Huang & Liu, 2001, p. 156)

b. *amihng laih taam **jihgei**$_i$ ge sihhauh, ameih$_i$ haihdouh tai-gan syu
Amihng come see SELF GE moment Ameih at read-DUR book
Intended: 'Ameih was reading when Amihng came to visit her.' (CC)

No logocentric predicate is required for PIVOT binding. Formally, we do not posit any LOG feature marking for PIVOT binding. (41) shows the constraints in the lexical entry of *ziji* for PIVOT binding:

(41) $\neg((\text{GF}^* \ \text{GF}\uparrow) \ \text{GF}^* \ \text{LOG})$
$(\uparrow_\sigma \text{ANT}) \neq ((\underset{\textbf{OFFPATH}}{\text{GF}^*} \ \text{GF}_{pro} \uparrow) \ \textbf{ANTE\_(TA)ZIJI\_JIHGEI})_\sigma$
$(\uparrow_\sigma \text{ANT}) = ((\text{GF}^* \ \text{GF}_{pro} \uparrow) \ \text{SUBJ})_\sigma$
$(\uparrow \text{ANT-TYPE}) = \text{PIVOT}$

The first constraint requires there to be no LOG feature in the f-structure of the sentence. In other words, there is no formal marking of logophoric domain by any logocentric predicates as the logocentric predicates in our system are either SOURCE- or SELF-predicates. The second constraint is a negative version of our anaphoric binding constraint, containing the previously seen components (20) and (25). It requires *ziji* not to be bound by any local antecedent, which otherwise constitutes anaphoric binding. See e.g., (40a) where *ziji* is not bound locally but by an entity somewhere else. The third constraint requires *ziji* to be bound by SUBJ, as is required in PIVOT-binding, where the speaker takes the perspective of a sentence-internal protagonist. PIVOT binding is not licensed by a logocentric verb, which otherwise assigns the LOG-ANT feature to the f-structure of the antecedent. The last constraint adds information to the f-structure of *ziji* that its antecedent is a PIVOT.

5.2.4 Discourse speaker as antecedent

The last type of logophoric binding relates to the observation that *ziji* and *jihgei* can refer to an antecedent in the discourse, which can be the external speaker or a discourse speaker a few sentences away (e.g., extended indirect speech). This is regarded as, cross-linguistically, a significant property of logophors in both pure and grammaticised logophoric systems (see e.g., Bresnan et al., 2016: 269; Culy, 1994; Maling, 1984; Sells, 1987). (42) is an extended indirect speech where *ziji* is interpreted as referring to *Xiaoming*:

(42) xiaoming$_i$ zai xiang … (a few sentences) … zhangsan jide
Xiaoming now think Zhangsan remember
xiaomei shuo-guo na-ge ren dui **ziji**$_i$-de chuxian gandao yiwai
Xiaomei say-PFV that-CL person to SELF-DE appear feel surprised
'Xiaoming$_i$ is now thinking... (a few sentences)… Zhangsan remembered Xiaomei said the person was surprised about his$_i$ appearing.' (MC)

We posit the following constraints in the lexical entries of *ziji* and *jihgei* for this type of binding:

(43) $\neg((\text{GF}^{*}\ \text{GF}\uparrow)\ \text{GF}^{*}\ \text{LOG})$

$(\uparrow_{\sigma}\ \text{ANT}) \neq ((\text{GF}^{*}\quad \text{GF}_{pro}\quad \uparrow)\quad \text{GF}^{*}\quad \text{GF})_{\sigma}$

$(\uparrow\text{ANT-TYPE}) = \text{DISCOURSE-SPEAKER}$

The first constraint requires there to be no LOG feature in the f-structure of the sentence. The second constraint requires the reflexive not to be bound by any entity within the sentence. The last constraint encodes the information that the reflexive refers to a discourse speaker.

5.3 Illustration of f-structures generated by our binding system

(44) is a CC sentence with three possible binding interpretations. See (16) for how we organise the various anaphoric and logophoric binding constraints in the lexical entry of the reflexive.

(44) amihng$_{i}$ wah-gwo [ameih$_{j}$ yanseung [**jihgei**$_{i/j/k}$-ge choihwah]]
Amihng say-PFV Ameih appreciate SELF-POSS talent
'Amihng$_{i}$ has said that Ameih$_{j}$ appreciates his/her$_{i/j/k}$ talent.' (CC)

With the constraints in our binding system, we generate the following (abbreviated) f-structures, each of which represents a referential possibility of *jihgei*. We use the subscripts – *i*, *j*, *k* – as an informal proxy to specify the coreferential relations. A more formal representation would show the coreferential relations in the form of semantic structures projected from the f-structures. Each type of binding relation is encoded with the appropriate f-structural information. The *i* interpretation in (45) belongs to SOURCE binding where *jihgei* is bound by *Amihng* along the path ((COMP OBJ POSS ↑) SUBJ). The *j* interpretation in (46) displays anaphoric binding with *jihgei* bound by *Ameih* along the path ((OBJ POSS ↑) SUBJ). The *k* interpretation in (47) displays binding by an external discourse speaker, for example, in extended indirect speech.

(45)
$$\begin{bmatrix} \text{PRED} & \text{`SAY}\ \langle\text{SUBJ, COMP}\rangle\text{'} \\ \textbf{SUBJ} & \begin{bmatrix}\text{PRED} & \text{`AMIHNG}_{i}\text{'} \\ \text{ANIMATE} & + \\ \text{LOG-ANT} & \text{SOURCE}\end{bmatrix} \\ \textbf{COMP} & \begin{bmatrix} \text{LOG} & + \\ \text{PRED} & \text{`APPRECIATE}\ \langle\text{SUBJ, OBJ}\rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix}\text{PRED} & \text{`AMEIH'}\end{bmatrix} \\ \textbf{OBJ} & \begin{bmatrix}\text{PRED} & \text{`TALENT}\langle\text{POSS}\rangle\text{'} \\ \textbf{POSS} & \begin{bmatrix}\text{PRED} & \text{`LOG-PRO}_{i}\text{'}\end{bmatrix}\end{bmatrix}\end{bmatrix}\end{bmatrix}$$

(46)
$$\begin{bmatrix} \text{PRED} & \text{`SAY}\ \langle\text{SUBJ, COMP}\rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix}\text{PRED} & \text{`AMIHNG'}\end{bmatrix} \\ \text{COMP} & \begin{bmatrix} \text{PRED} & \text{`APPRECIATE}\ \langle\text{SUBJ, OBJ}\rangle\text{'} \\ \textbf{SUBJ} & \begin{bmatrix}\text{PRED} & \text{`AMEIH}_{j}\text{'}\end{bmatrix} \\ \textbf{OBJ} & \begin{bmatrix}\text{PRED} & \text{`TALENT}\langle\text{POSS}\rangle\text{'} \\ \textbf{POSS} & \begin{bmatrix}\text{PRED} & \text{`REFL-PRO}_{j}\text{'}\end{bmatrix}\end{bmatrix}\end{bmatrix}\end{bmatrix}$$

(47)
$$\begin{bmatrix} \text{PRED} & \text{`SAY}\ \langle\text{SUBJ, COMP}\rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix}\text{PRED} & \text{`AMIHNG'}\end{bmatrix} \\ \text{COMP} & \begin{bmatrix} \text{PRED} & \text{`APPRECIATE}\ \langle\text{SUBJ, OBJ}\rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix}\text{PRED} & \text{`AMEIH'}\end{bmatrix} \\ \text{OBJ} & \begin{bmatrix}\text{PRED} & \text{`TALENT}\langle\text{POSS}\rangle\text{'} \\ \text{POSS} & \begin{bmatrix}\text{PRED} & \text{`LOG-PRO}_{k}\text{'} \\ \text{ANT-TYPE} & \text{DISCOURSE-SPEAKER}\end{bmatrix}\end{bmatrix}\end{bmatrix}\end{bmatrix}$$

6 Conclusion

This paper illustrates the power of the LFG machinery as it develops a constraint-based system capable of differentiating various types of anaphoric and logophoric binding in MC and CC. The LFG formalism has a high level of flexibility allowing it to model both types of binding, while maintaining its formal, mathematical rigour. Our constraint-based approach offers an alternative binding theory in response to the recent Minimalist proposals on Chinese binding (e.g., Giblin, 2016; Reuland et al., 2020), opening up a cross-theoretical dialogue. We have established the notion of grammaticised logophoricity in MC and CC in connection with crosslinguistic studies. Empirically, we have re-examined data of MC to clarify the properties of MC reflexives and settle the animacy-antecedent debate with reference to the typological literature on adnominal possession. The comparison between MC and CC contributes to the comparative study of binding phenomena in Sinitic languages.

References

Bresnan, Joan, Ash Asudeh, Ida Toivonen, and Stephen Wechsler. 2016. *Lexical-Functional Syntax*. John Wiley & Sons.

Charnavel, Isabelle, and Yu-Jing Huang. 2018. 'Inanimate Ziji and Condition A in Mandarin'. In *Proceedings of the 35th West Coast Conference on Formal Linguistics*, edited by Wm. G. Bennett, Lindsay Hracs, and Dennis Ryan Storoshenko, 132–41.

Charnavel, Isabelle, C-T James Huang, Peter Cole, and Gabriella Hermon. 2017. 'Long-Distance Anaphora: Syntax and Discourse'. In *The Wiley Blackwell Companion to Syntax*, 2nd Edition, 1–82. John Wiley & Sons.

Chierchia, Gennaro. 1989. 'Anaphora and Attitudes De Se'. In *Semantics and Contextual Expression*, edited by R. Bartsch, J. van Benthem, and P. van Emde Boas, 1–32. Berlin, Boston: De Gruyter.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris Publications.

Clements, George N. 1975. 'The Logophoric Pronoun in Ewe: Its Role in Discourse'. *Journal of West African Languages* 10: 141–77.

Cole, Peter, Gabriella Hermon, and Cher Leng Lee. 2001. 'Grammatical and Discourse Conditions on Long Distance Reflexives in Two Chinese Dialects'. *Long-Distance Reflexives* 33: 1–46.

Cole, Peter, and Chengchi Wang. 1996. 'Antecedents and Blockers of Long-Distance Reflexives: The Case of Chinese Ziji'. *Linguistic Inquiry*, 357–90.

Culy, Christopher. 1994. 'Aspects of Logophoric Marking'. *Linguistics* 32 (6).

Dalrymple, Mary. 1993. *The Syntax of Anaphoric Binding*. 36. Center for the Study of Language (CSLI).

———. 2015. 'Obligatory Nonlocal Binding'. *Natural Language & Linguistic Theory* 33 (4): 1089–1120.

Dalrymple, Mary, John Lowe, and Louise Mycock. 2019. *The Oxford Reference Guide to Lexical Functional Grammar*. Oxford University Press.

Giblin, Iain. 2016. 'Agreement Restrictions in Mandarin Long-Distance Binding'. PhD Thesis, MIT.

Koptjevskaja-Tamm, Maria. 2002. 'Adnominal Possession in the European Languages: Form and Function'. *STUF - Language Typology and Universals* 55 (2).

Haspelmath, Martin. 2017. 'Explaining Alienability Contrasts in Adpossessive Constructions: Predictability vs. Iconicity'. *Zeitschrift Für Sprachwissenschaft* 36 (2): 193–231.
Huang, C-T James, Yen-Hui Audrey Li, and Yafei Li. 2009. *The Syntax of Chinese*. Cambridge University Press.
Huang, C-T James, and C-S Luther Liu. 2001. 'Logophoricity, Attitudes, and Ziji at the Interface'. In *Long-Distance Reflexives*, edited by Peter Cole, Gabriella Hermon, and C-.T James Huang, 33:141–95. New York: Academic Press.
Huang, C-T James, and C-C Jane Tang. 1991. 'The Local Nature of the Long-Distance Reflexive in Chinese'. In *Long-Distance Anaphor*, edited by Jan Koster and Eric Reuland, 19:263–82. Cambridge University Press.
Huang, Yan. 2000. *Anaphora: A Cross-Linguistic Study*. Oxford Studies in Typology and Linguistic Theory. Oxford: Oxford University Press.
———. 2016. 'Aspects of Anaphora in Chinese and in Some Germanic, Romance, and Slavic Languages, the "Syntactic" Versus "Pragmatic" Language Typology, and Neo-Gricean Pragmatics'. In *Pragmemes and Theories of Language Use*, edited by Keith Allan, Alessandro Capone, and Istvan Kecskes, 9:21–43. Springer International Publishing.
Jackendoff, Ray. 1990. *Semantic Structures*. Current Studies in Linguistics 18. Cambridge, Mass: MIT Press.
Maling, Joan. 1984. 'Non-Clause-Bounded Reflexives in Modern Icelandic'. *Linguistics and Philosophy*, 211–41.
Matthews, Stephen, and Virginia Yip. 2013. *Cantonese: A Comprehensive Grammar*. Routledge.
Pan, Haihua. 1997. *Constraints on Reflexivization in Mandarin Chinese*. Garland Publishing.
———. 2001. 'Why the Blocking Effect'. In *Long-Distance Reflexives*, edited by Peter Cole, Gabriella Hermon, and C-.T James Huang, 33:279–316. Syntax and Semantics 33. New York: Academic Press.
Pan, Haihua, and Jianhua Hu. 2003. 'Prominence and Locality in the Binding of Mandarin Complex Reflexive "Ta-Ziji"(s/He-Self)'. *Journal of Chinese Linguistics Monograph* 19: 152–70.
Pollard, Carl, and Ping Xue. 1998. 'Chinese Reflexive Ziji: Syntactic Reflexives vs. Nonsyntactic Reflexives'. *Journal of East Asian Linguistics* 7 (4): 287–318.
Reinhart, Tanya, and Eric Reuland. 1993. 'Reflexivity'. *Linguistic Inquiry* 24 (4): 650–720.
Reuland, Eric J. 2011. *Anaphora and Language Design*. Linguistic Inquiry Monographs 62. Cambridge, Mass: The MIT Press.
Reuland, Eric, Sally Chi Ho Wong, and Martin Everaert. 2020. 'How the Complexity of Mandarin *Zi-Ji* Simplifies the Grammar'. *Linguistic Inquiry* 51 (4): 799–814.
Sells, Peter. 1987. 'Aspects of Logophoricity'. *Linguistic Inquiry* 18 (3): 445–79.
Sperlich, Darcy. 2019. 'Syntactic and Pragmatic Theories of Chinese Reflexives'. *Lingua* 221: 22–36.
Tang, Chih-Chen Jane. 1989. 'Chinese Reflexives'. *Natural Language & Linguistic Theory* 7 (1): 93–121.
Xu, Liejiong. 1993. 'The Long-Distance Binding of Ziji'. *Journal of Chinese Linguistics* 21 (1): 123–42.
Xue, Ping, Carl Pollard, and Ivan A Sag. 1994. 'A New Perspective on Chinese Ziji'. In *Proceedings of the Thirteenth West Coast Conference on Formal Linguistics*.
Yip, Virginia, and Gladys Tang. 1998. 'Acquisition of English Reflexive Binding by Cantonese Learners'. In *Morphology and Its Interfaces in Second Language Knowledge*, edited by Harald Clahsen and Lydia White, 165–93. John Benjamins Publishing Company.

(Almost) everything is oblique in West Circassian

Yury Lander
HSE University

Oleg Belyaev
Lomonosov Moscow State University/Institute of Linguistics RAS/
Pushkin State Russian Language Institute

Irina Bagirokova
Institute of Linguistics RAS

Proceedings of the LFG'21 Conference

On-Line

Miriam Butt, Jamie Y. Findlay, Ida Toivonen (Editors)

2021

CSLI Publications

pages 223–242

http://csli-publications.stanford.edu/LFG/2021

Keywords: polysynthesis, agreement, West Circassian, case marking, argument structure, applicative

Lander, Yury, Belyaev, Oleg, & Bagirokova, Irina. 2021. (Almost) everything is oblique in West Circassian. In Butt, Miriam, Findlay, Jamie Y., & Toivonen, Ida (Eds.), *Proceedings of the LFG'21 Conference, On-Line*, 223–242. Stanford, CA: CSLI Publications.

Abstract

This paper provides a novel analysis of grammatical function marking in West Circassian (West Caucasian, Russia), a polysynthetic language whose syntactic features do not quite fit the standard types of polysynthesis described in the literature. We argue that there is a straightforward connection between GF status, verbal indexing, and case marking. Namely, subjects are Absolutive-marked and never indexed, while all other core arguments are indexed and marked by Oblique case. We show that the description of argumenthood in West Circassian only requires the features SUBJ, OBJ_θ and OBL_θ; OBJ is not needed. SUBJ is most often the $\mathcal{S}/\mathcal{P}$ argument, but 1st and 2nd person pronouns, being indexed on the verb, map to OBJ_θ in all semantic roles. Therefore, GF assignment in West Circassian is dependent on person, and it is possible to have subjectless sentences. We provide a sketch formalization of this analysis and discuss its wider implications.

1 Introduction

While grammatical functions are viewed as theoretical primitives in LFG, the exact patterns of mapping from semantic roles to GFs, and the extent to which languages may choose to use elements of the universally available inventory of GFs, are the subject of much debate. It is widely accepted that languages with non-accusative alignment types can have a GF mapping that differs from that in syntactically accusative languages; for instance, Manning (1996) proposes that SUBJ and OBJ (CORE in his terminology) in ergative languages receive an inverse mapping; namely, the absolutive argument is mapped to SUBJ, while the ergative argument is OBJ/CORE. Falk (2006) splits the grammatical function SUBJ into two functions: $\widehat{\text{GF}}$ (the most prominent argument, corresponding to the traditional $\mathcal{A}/\mathcal{S}$ "subject") and PIV (responsible for licensing long-distance dependencies). The identification of PIV with either $\widehat{\text{GF}}$ or OBJ produces the traditional syntactically accusative/ergative typology, but other patterns of PIV assignment are possible and indeed, according to Falk, attested in various languages of the world.

In this paper, we argue that the data of West Circassian, a polysynthetic West Caucasian language of southern Russia, suggest a complex relationship between case marking, verbal indexing and syntactic behaviour that, in LFG terms, should be analyzed as a rather unusual pattern of GF assignment. Specifically, we suggest that verbal[1] indexing and case marking directly correlate with GF status: all

[†]We are grateful to the audience of LFG2021, especially Ash Asudeh, Mary Dalrymple, and Brian O'Herin, for insightful comments and discussion. In this paper, the formal analysis is due to Belyaev, while the data and informal generalizations are due to Lander and Bagirokova. These, in turn, are heavily based on the ideas of Yakov Testelets (p. c.) and analyses in Arkadiev et al. (2009) and Beliaeva (2006) and Lander and Bagirokova (2017). All remaining errors are ours.
Belyaev's research was performed according to the Development Program of the Interdisciplinary Scientific and Educational School of Moscow University "Preservation of the World Cultural and Historical Heritage".

[1]West Circassian has no well-defined lexical class of verbs as opposed to nouns and adjectives;

indexed arguments have the grammatical function OBJ_θ and are Oblique-marked; the sole non-indexed argument marked by Absolutive is SUBJ;[2] all other (non-indexed and non-Absolutive) arguments are marked by postpositions. This analysis entails some surprising effects, such as the fact that, since 1st and 2nd person $\mathcal{S}/\mathcal{P}$ arguments are always indexed and Oblique-marked (where allowed by the morphology), they should be treated as OBJ_θ; thus, GF assignment is dependent on person, and sentences without SUBJ are possible. The existence of such sentences is consistent with earlier claims, such as Kibort (2006), but the person-motivated, rather than lexical, split is highly unusual.

The paper is organized as follows. In section 2, we describe the West Circassian pattern of case marking and verbal indexing. In 3, we describe the syntactic properties of the core arguments, demonstrating that indexing and case marking correlate with certain syntactic diagnostics that reveal their GF status. In 4, we provide a generalization of our analysis with respect to different types of arguments and case marking patterns in West Circassian.

2 Indexing and case marking in West Circassian

2.1 Indexing

West Circassian (also known as Adyghe, although the same term is applied to Circassian languages – West Circassian and Kabardian – in general) is a polysynthetic language which uses both case marking and verbal indexing of core arguments.[3] Indexing is expressed by a set of verbal prefixes, whose main allomorphs are shown in Table 1. The system of indexing is ergative: there is a set of absolutive indices that refer to $\mathcal{S}/\mathcal{P}$ arguments, and an ergative set for coindexing ergative ($\mathcal{A}$) arguments. Markers labeled as "IO" in Table 1 are used together with applicative prefixes to coindex arguments of various oblique semantic roles (recipients, goals, locations, etc.); they are largely identical to ergative markers. Importantly, 3SG and 3PL absolutive markers are left unmarked.[4]

the terms *verb* and *verbal* are thus used informally, to refer to predicative forms. Statements about verbs equally apply to nominals serving as clausal predicates.

[2]Falk's (2006) split subject analysis is not needed for West Circassian, because while there is some evidence for syntactic processes that specifically target $\widehat{\text{GF}}$ ($\mathcal{A}/\mathcal{S}$) in Lander and Testelets (2017), it is unclear and could also have a semantic explanation.

[3]In what follows, we will sometimes use the typological term "flagging" (Haspelmath 2019) to refer to Absolutive/Oblique case marking. It is in fact debatable whether Absolutive and Oblique should be treated as case markers proper, as will be discussed below.

[4]This is sometimes described in the literature as zero marking. As we show in this paper, this view is incorrect: 3rd person $\mathcal{S}/\mathcal{P}$ arguments are indeed *not* indexed on the verb, which correlates with their flagging and syntactic properties.

Table 1: West Circassian argument indexing prefixes

	ABS	IO	ERG
1SG	*sə-*	*s-*	
1PL	*tə-*	*t-*	
2SG	*wə-*	*p-/w-*	
2PL	*ŝʷə-*	*ŝʷ-*	
3SG	—	*∅-*	*ə-/jə-*
3PL	—	*a-*	
REL/RFL	*zə-*		

2.2 Case marking

Overt case marking in West Circassian is optional (depending on referentiality, see Arkadiev and Testelets 2019). When it appears, the system is two-term: either the Absolutive (*-r*, often called Nominative) or the Oblique (*-m*, also *-š’/j* with certain pronouns and *-me* in the plural; often called Ergative) is used. The Instrumental and Adverbial, also shown in the table, are peripheral cases that display somewhat different properties compared to the Absolutive and Ergative; they are not generally used to mark core arguments, but the Adverbial marks the internal head in relative clauses (see section 3.1.1). Under the traditional view of the West Circassian flagging system (see e.g. Rogava and Keraševa 1966; Kumakhov and Vamling 2019), the Absolutive is used on $\mathcal{S}/\mathcal{P}$ arguments, while the Oblique is used on $\mathcal{A}$, as well as on all arguments that have been introduced by applicative prefixes. As an example of both flagging and indexing, see (1).[5]

(1) *č̣’ale-m*$_i$ *pŝaŝe-m*$_j$ *laʁe-xe-r* *we*$_k$
boy-OBL girl-OBL dish-PL-ABS you.SG

qə- b-$_k$*də- ∅-*$_j$*r- jə-*$_i$ *tə -ʁe -x*
DIR-2SG.IO-COM-3SG.IO-DAT-3SG.ERG-give-PST-PL

‘The boy gave the dishes to the girl with you (sg.).’

In this example, the 3rd person $\mathcal{P}$, ‘dishes’, is not indexed in the verb but flagged by Absolutive case; the 3rd person agent, ‘boy’, is flagged by Oblique and coindexed by the ergative prefix *jə-*. Two additional arguments – comitative, ‘with you’, and recipient, ‘to the girl’, – are introduced by applicative prefixes used together with “indirect object” indices. The recipient is also expressed

[5]The formatting of examples follows the Leipzig Glossing Rules (https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf), using the following abbreviations: ABS – absolutive; ADV – adverbial case; ADD – additive; COM – comitative; DAT – dative; DIR – directive (verbal orientation marker); ERG – ergative (indexing prefix); FUT – future tense; IMP – imperative; INS – instrumental; IO – indirect object; LOC – locative; MOD – modal; NEG – negative; OBL – oblique; PL – plural; PST – past tense; RE – refactive; REL – relative; SG – singular.

by an independent NP that is flagged by Oblique. Crucially, the applicatives in West Circassian are quite different from their namesakes in many other languages (Bresnan and Moshi 1990): namely, they do not change the syntactic status of the core arguments, but merely introduce additional secondary objects into the verbal valency frame.

Non-$\mathcal{S}/\mathcal{P}$ arguments that are not indexed on the verb are mainly expressed by postpositional phrases[6] where the complement is marked by Oblique, see (2).

(2) *ʔeše-deʁʷ-jə* *w-jə-š'əč̣'aʁ-ep* ***mə ʔʷefə-m paje***
weapon-good-ADD 2SG.IO-POSS-necessity-NEG this deed-OBL for
'You don't even need a good weapon **for this deed**.'

3 Syntactic properties of arguments

3.1 Subjecthood

A natural question that arises here is how exactly flagging and indexing are related to grammatical function assignment and subjecthood. West Circassian has a rich inventory of valency-increasing operations but has no real valency-decreasing operations (Lander and Letuchiy 2017); therefore, there are no constructions where it could be argued, for example, that the direct object is promoted to subject status. On the syntactic ergativity of valency-changing operations, also see Letuchiy (2012). Overall, it seems that only the absolutive ($\mathcal{S}/\mathcal{P}$) can be singled out as having a special subject (pivot) function; for an overview of arguments in favour of syntactic ergativity, see Ershova (2019). All other arguments can be treated as secondary objects or obliques, as argued in Lander (2005). This can be demonstrated by two syntactic tests: the behaviour of internally-headed relativization and the "raising-like" construction with the verb 'must'.[7]

3.1.1 Internally-headed relativization

West-Circassian has both internally- and externally-headed relative clauses (thoroughly described in Lander 2012; also see Lander and Daniel 2019 for an overview of the use of relative prefixes in these constructions). Externally-headed relative clauses are prenominal, where $\mathrm{NP}_{\mathrm{rel}}$ is not expressed by a full NP (3a); the external head receives the Oblique case from the matrix verb. In contrast, the head is inside the relative clause in internally-headed relative clauses; it is always marked by the Adverbial suffix. As seen in (3b), it is the verb of the relative clause that receives external case marking.

[6]With the exception of temporal, and partly locative, expressions, which can also be marked by Oblique while not being indexed.

[7]There are also some other constructions that contrast between the absolutive and other arguments which include very specific constraints on relativization not discussed here, see Lander (2010), Lander (2012), and Ershova (2019) for details.

(3) a. [*a-ç̌'e* *ḳʷe-š't*] ***çəfə-m*** *sə-λ-e-χʷə*
that-INS go-FUT person-OBL 1SG.ABS-LOC-DYN-search

b. [*a-ç̌'e* ***çəf-ew*** *ḳʷe-š'tə-**m***] *sə-λ-e-χʷə*
that-INS person-ADV go-FUT-OBL 1SG.ABS-LOC-DYN-search

'I am looking for **a person** [who would go there.]'

(Arkadiev et al. 2009)

The word order in internally-headed relative clauses is somewhat more restricted than in main clauses. Namely, the Adverbial-marked internal head normally cannot separate the Absolutive-marked NP from the verb (Beliaeva 2006; Lander 2010; Lander 2012):

(4) a. *təʁʷaḳʷ-ew* ***dəŝe-r*** *zə-ʔe.pə-teqʷə-ʁe-r*
thief-ADV gold-ABS REL.IO-LOC-disperse-PST-ABS

'the thief who dropped **the gold**', lit. 'out of whose hands the gold fell'

b. ****dəŝe-r*** *təʁʷaḳʷ-ew* *zə-ʔe.pə-teqʷə-ʁe-r*
gold-ABS thief-ADV REL.IO-LOC-disperse-PST-ABS

No such restrictions exist for OBL-marked NPs, regardless of their semantic role:

(5) ***thamate-m*** *qebar-ew* *q-ə-ʔʷete-š'tə-m*
chief-OBL news-ADV DIR-3SG.ERG-tell-FUT-OBL

'the news that **the chief** would tell' (Lander 2012, 250)

This means that the Absolutive noun phrase has a designated position somewhere in the clause structure (at least in internally-headed relative clauses), at a relatively low level, while the position of Oblique-marked arguments is free. While the rule itself could be explained in terms of case, it is more economical to intepret it in terms of a privileged syntactic status of the Absolutive NP, with Absolutive merely serving to flag the GF SUBJ: Case assignment in West Circassian is always fully predictable from semantic roles and verbal marking, and it is never lexically idiosyncratic.

3.1.2 Raising-like construction

Another construction that displays the pivot status of the Absolutive is the raising-like construction with the verb 'must' (Testelets 2009, 688). This verb takes an Adverbial-marked complement clause and may (for some speakers) agree in number with the Absolutive argument of the subordinate clause:[8]

[8]Predicates in West Circassian may always – optionally – agree with Absolutive arguments of their own clauses in number. This could in itself be taken as a piece of evidence in favour of the subject status of the Absolutive, although clause-internal agreement by itself may be case-driven.

(6) a. ***a-xe-r*** *qe-ŝʷe-n-x-ew* *š'ə.tə*-***x***
that-PL-ABS DIR-dance-MOD-PL-ADV must-PL
'**They** should dance.'

b. *a-š'* ***pjəsme-xe-r*** *ə-txə-n-x-ew* *š'ə.tə*-***x***
that-OBL letter-PL-ABS 3SG.ERG-write-MOD-PL-ADV must-PL
'S/he must write **letters**.'

c. ****a-xe-me*** *ʔʷef* *a-ṣ̂ə-n-ew* *š'ə.tə*-***x***
they-PL-OBL.PL work 3PL.ERG-do-MOD-ADV must-PL
(intended: '**They** should work.')

This rule could also be formulated in terms of case rather than grammatical function ("agree with the clause-level absolutive, or with the absolutive of your COMP"). However, given the lack of independent evidence in favour of the pivot status of other arguments, this is more complex than simply stating that the Absolutive-marked NP is the subject. The behaviour of this construction also converges with the behaviour of internally-headed relative clauses. Furthermore, according to Falk's (2006, 78) Pivot Condition, all paths that link arguments across clauses must terminate in a PIV. Hence, long-distance agreement (or functional control) by itself presents enough evidence in favour of the pivot status of the Absolutive.[9]

3.1.3 Analysis

There are two ways in which the pivot status of the Absolutive argument may be analysed in LFG. The simpler would be, in terms of Manning (1996), to treat it as the SUBJ. A more complex analysis, following Falk (2006), is to postulate that the Absolutive is PIV, while maintaing that all clauses also have the $\widehat{\text{GF}}$ function that corresponds to the traditional notion subject ($\mathcal{A}/\mathcal{S}$). In this case, PIV would be identified with $\widehat{\text{GF}}$ in intransitive clauses and with PIV in transitive ones.

The latter solution is, of course, technically possible for West Circassian, but there are no good examples of constructions which are syntactically $\widehat{\text{GF}}$-oriented. Reflexives may seem to target $\mathcal{S}/\mathcal{A}$, but they are better described as targeting the more agent-like argument. Specifically, in the Potential construction, where Ergative indexing of $\mathcal{A}$ is replaced by IO indexing, $\mathcal{A}$ still has binding priority over $\mathcal{P}$.

Furthermore, adopting Falk's (2006) analysis implies maintaining the traditional grammatical function OBJ as opposed to $\widehat{\text{GF}}$ and OBJ$_\theta$. However, this creates an artificial split between $\mathcal{S}/\mathcal{A}$ ("ergative") arguments (which would have to be $\widehat{\text{GF}}$) and other indexed arguments (which would have to be OBJ$_\theta$). In morphological terms, the only difference is that the latter require applicative prefixes, while

[9] A full analysis of the construction with the verb 'must' is outside the scope of this paper. Regardless of whether it is a case of long-distance agreement or functional control, the data clearly present evidence in favour of the pivothood of the Absolutive.

the former is indexed directly; the morphology itself is largely identical (see Table 1). In syntactic terms, we have seen above that the only distinction that can be drawn among core arguments in West Circassian is two-way: between Absolutive and Oblique-marked arguments. Since the number of the latter in a clause is not syntactically restricted, and they always receive indexing depending on their semantic role, they should be viewed as semantically restricted objects (OBJ$_\theta$). Therefore, using OBJ or $\widehat{\text{GF}}$ is redundant; all that is required is a three-way distinction:[10]

SUBJ/PIV $\mathcal{S}/\mathcal{P}$, Absolutive-marked, not indexed on the verb;

OBJ$_\theta$ Oblique-marked, indexed on the verb (both $\mathcal{A}$ and introduced by applicatives);

OBL$_\theta$ postposition-marked, not indexed on the verb.

Abandoning the distinction between OBJ and OBJ$_\theta$ may seem like a radical move, given that, in most LFG work, OBJ$_\theta$ only appears in the presence of a primary object (OBJ). However, such analyses have been proposed before. For example, Dahlstrom (2009) claims that some verbs in Meskwaki (Algonquian) select only SUBJ and OBJ$_\theta$. More radically, Börjars and Vincent (2008) propose abandoning the distinction altogether, treating all objects as OBJ$_\theta$. We do not go so far as to claim that OBJ is universally useless as a GF; what we claim is that it is unnecessary for West Circassian, which only specially distinguishes the subject among the term arguments.

Thus, (1), repeated here, may be analyzed as having the f-structure in (7).

[1, repeated] *č̣’ale-m$_i$* *pŝaŝe-m$_j$* *laʁe-xe-r* *we$_k$*
boy-OBL girl-OBL dish-PL-ABS you.SG

qə- b-$_k$də- ∅-$_j$r- jə-$_i$ tə -ʁe -x
DIR-2SG.IO-COM-3SG.IO-DAT-3SG.ERG-give-PST-PL

‘The boy gave the dishes to the girl with you (sg.).’

[10]Indexed arguments could also be viewed as OBL$_\theta$, which would perhaps be more palatable to a traditional view, because having OBJ$_\theta$ without OBJ seems unusual. However, OBL$_\theta$ is required to distinguish non-indexed arguments – usually expressed by PPs – from Absolutive and Oblique arguments. Furthermore, verbal coindexing is a standard criteria for term (direct, non-oblique) status, see Dalrymple, Lowe, and Mycock (2019, 16).

(7)

$$\left[\begin{array}{ll} \textsc{pred} & \text{‘give<}\textsc{subj}\ \textsc{obj}_{\textsc{ag}}\text{ , }\textsc{obj}_{\textsc{goal}}\text{ , }\textsc{obj}_{\textsc{com}}\text{>’} \\ \textsc{tense} & \textsc{past} \\ \textsc{dir} & \textit{qə} \\ \textsc{subj} & \left[\begin{array}{ll} \textsc{pred} & \text{‘dish’} \\ \textsc{pers} & 3 \\ \textsc{num} & \textsc{pl} \end{array}\right] \\ \textsc{obj}_{\textsc{ag}} & \left[\begin{array}{ll} \textsc{pred} & \text{‘boy’} \\ \textsc{pers} & 3 \\ \textsc{num} & \textsc{sg} \end{array}\right]_{i} \\ \textsc{obj}_{\textsc{goal}} & \left[\begin{array}{ll} \textsc{pred} & \text{‘girl’} \\ \textsc{pers} & 3 \\ \textsc{num} & \textsc{sg} \end{array}\right]_{j} \\ \textsc{obj}_{\textsc{com}} & \left[\begin{array}{ll} \textsc{pred} & \text{‘PRO’} \\ \textsc{pers} & 2 \\ \textsc{num} & \textsc{sg} \end{array}\right]_{k} \end{array}\right]$$

In (7), it can be seen that the grammatical function OBJ is not used. $\mathcal{P}$ is SUBJ, while $\mathcal{A}$ is OBJ_{AG}, not different from other Oblique-marked, indexed arguments.

3.2 1st and 2nd person arguments

Thus far, our analysis has presented a rather regular, straigtforward relationship between case-marking, verbal indexing and GF status in West Circassian. However, there is one seeming exception from this pattern: 1st and 2nd person arguments. As seen in Table 1, unlike 3rd person arguments which are unmarked when $\mathcal{S}/\mathcal{P}$ (i.e. SUBJ in our analysis), these are always overtly indexed on the verb, even when corresponding to $\mathcal{S}/\mathcal{P}$. Furthermore, 1st and 2nd person arguments are never marked by either Oblique or Absolutive in the core functions:

(8)

we	*sə-b-de-haš’xə-ʁ-ep*	*se*
you.SG	1SG.ABS-2SG.IO-LOC-laugh.at-PST-NEG	I

‘I did not laugh at you.’

If uniformity of semantic role to GF mapping is assumed, we can conclude that full pronouns do not morphologically distinguish case, but are OBJ_{θ} when $\mathcal{A}$ (as the agent of 8) and SUBJ when $\mathcal{S}/\mathcal{P}$ (as the object of 8), just as 3rd person NPs. However, unlike full NPs, they are always indexed. This would mean that 1st and 2nd person pronouns are exceptions from the generalization on the direct connection between GF status, case marking and indexing.

The actual status of 1st and 2nd person $\mathcal{S}/\mathcal{P}$ is more complicated, however. First, if a lexical noun, quantifier or adjective is used in place of the full pronoun, or as a postposed modifier of the pronoun, it always stands in the Oblique, even where the Absolutive is expected:

(9) *zeč̣'e-**m**-jə tə-adəg*
all-OBL-ADD 1PL.ABS-Circassian
'We all are Circassians.'[11]

(10) a. *te č̣'ale-xe-**m** tə-qe-ḳʷa-ʁ*
we boy-PL-OBL 1PL.ABS-DIR-go-PST
'We boys came.'

b. **te č̣'ale-xe-**r** tə-qe-ḳʷa-ʁ*
we boy-PL-ABS 1PL.ABS-DIR-go-PST

(Arkadiev et al. 2009, 81)

Thus, lack of case marking on pronouns is a fact of morphology. However, when the syntactic context allows overt case marking to surface, it contradicts our expectations in always being Oblique.

Second, unmarked 1st and 2nd person pronouns in $\mathcal{S}/\mathcal{P}$ position allow an internal head to appear between them and the predicate of the relative clause – something which, as we just saw in section 3.1.1, is not allowed for Absolutive (SUBJ) arguments:

(11) [***we** çəf-ew wə-zə-λeʁʷə-ʁe-r*] *ʔʷəč̣'ə-ž'ə-ʁ*
thou person-ADV 2SG.ABS-REL.ERG-see-PST-ABS leave-RE-PST
'The man who saw **you** left.' (Beliaeva 2006)

In the logic of our proposal, (9) and (11) show that 1st and 2nd person $\mathcal{S}/\mathcal{P}$ arguments, regardless of their overt expression, are thematically restricted objects (OBJ_θ). This means that intransitive clauses with 1st/2nd person $\mathcal{S}$ arguments, such as (9), have no SUBJ, but only an OBJ_{ABS},[12] as in (12).

(12)
$$\begin{bmatrix} \text{PRED} & \text{'Circassian}\langle\text{OBJ}_{\text{ABS}}\rangle\text{'} \\ \text{TENSE} & \text{PRES} \\ \text{OBJ}_{\text{ABS}} & \begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{SPEC} & \textit{all} \\ \text{PERS} & 1 \\ \text{NUM} & \text{PL} \end{bmatrix} \end{bmatrix}$$

Hence, the Subject Condition (Bresnan and Kanerva 1989) does not hold in West Circassian, but this idea, in itself, is not new. For instance, in Falk's (2006, 184) analysis of Choctaw, verbs like 'afraid' select only OBJ_θ and OBJ. Kibort (2006)

[11]As suggested by Ash Asudeh (p.c.), the syntactic peculiarity of this sentence could be expressed in English as *Circassians us all*, with the difference that *all* in West Circassian is not an appositive modifier of *us* but an oblique NP occupying an argument position.

[12]Although Absolutive is not a proper thematic role, we use this label as a cover term for $\mathcal{S}$ and $\mathcal{P}$. In this example, using THEME is also possible, but would require mapping the agreement prefix to different grammatical functions with transitive and intransitive verbs, somewhat complicating the analysis.

analyses certain impersonal sentences in Polish as being truly subjectless. What distinguishes West Circassian from these cases is that this mapping pattern is not lexical, but determined by the person of the verb's arguments. Such a person-dependent GF mapping has not, to our knowledge, previously been proposed in the literature.

4 Discussion

4.1 Generalization

The discussion above, and our analysis, can be summarized in the following way. Grammatical function assignment in West Circassian follows a person-based split. 3rd person arguments are assigned to SUBJ if $\mathcal{S}/\mathcal{P}$ and to OBJ$_\theta$ if $\mathcal{A}$ or introduced by an applicative prefix. In the 1st and 2nd persons, the mapping pattern is different: all core arguments and applicative-introduced arguments are mapped to OBJ$_\theta$. For all persons, oblique arguments that are not introduced by applicatives have the status of OBL$_\theta$. This pattern is summarized in Table 2.

Table 2: GF mapping in West Circassian

<table>
<tr><th></th><th>$\mathcal{S}/\mathcal{P}$</th><th>$\mathcal{A}$</th><th>appl.</th><th>obl.</th></tr>
<tr><td>1/2p.</td><td>OBJ$_\theta$</td><td colspan="2" rowspan="2">OBJ$_\theta$</td><td rowspan="2">OBL$_\theta$</td></tr>
<tr><td>3p.</td><td>SUBJ</td></tr>
</table>

Morphosyntactic encoding is almost completely determined by GF status.[13] SUBJs are case-marked by Absolutive and never indexed on the verb. OBJ$_\theta$s are case-marked by Oblique and always indexed on the verb. OBL$_\theta$s are marked by postpositions and never indexed on the verb. This is summarized in Table 3.

Table 3: Morphosyntactic marking in West Circassian

	case	**index**
SUBJ	Absolutive	✗
OBJ$_\theta$	Oblique	✓
OBL$_\theta$	postposition	✗

[13]The only exception is the existence of indexed PPs (Lander 2015; Lander 2016, 3509), which we do not discuss here. This would be relatively straightforward to formally implement, but would make the generalizations on argument mapping and case marking more complicated.

4.2 Formalization

A full LFG formalization of this analysis requires a more thorough formalization of West Circassian morphology, which is not currently available. Nevertheless, a set of rules and definitions can be sketched using sublexical morphology as used e.g. in Bresnan et al. (2016). The verb consists of the base stem (which we will not analyze at this point) together with a number of coreferencing prefixes:[14]

(13)
$$\begin{array}{llccccc} \mathrm{V} & \rightarrow & (\mathrm{V}_{\mathrm{cref}}) & (\mathrm{V}_{\mathrm{dir}}) & \mathrm{V}_{\mathrm{appl}}{}^{*} & (\mathrm{V}_{\mathrm{cref}}) & \mathrm{V}_{\mathrm{stem}} \\ & & \begin{array}{c}(\uparrow \text{OBJ}_{\text{ABS}}){=}\downarrow \\ (\downarrow \text{PERS}) =_c 1|2\end{array} & \uparrow{=}\downarrow & (\uparrow(\downarrow \text{PCASE})){=}\downarrow & (\uparrow \text{OBJ}_{\text{AG}}){=}\downarrow & \uparrow{=}\downarrow \end{array}$$

The three $\mathrm{V}_{\mathrm{cref}}$ nodes stand for the positions of prefixes that can index arguments without additional applicative morphology. These are, first, the 1st and 2nd person $\mathcal{S}/\mathcal{P}$ arguments; second, the "ergative" prefix (ERG in 1).[15] The annotations on the nodes ensure that only OBJ_θ arguments receive verbal indexing.

The internal structure of the $\mathrm{V}_{\mathrm{appl}}$ node is akin to a PP. It consists of a $\mathrm{V}_{\mathrm{cref}}$ node[16] followed by a $\mathrm{V}_{\mathrm{post}}$ node:

(14) $\mathrm{V}_{\mathrm{appl}} \rightarrow \mathrm{V}_{\mathrm{cref}}\ \mathrm{V}_{\mathrm{post}}$

Turning to the lexical content of these nodes, $\mathrm{V}_{\mathrm{stem}}$ introduces the PRED value and morphosyntactic features such as tense, mood, etc.

(15) *tə-ʁe-x* $\mathrm{V}_{\mathrm{stem}}$
$$\begin{array}{l} (\uparrow \text{PRED}) = \text{'give} \langle \text{OBJ}_{\text{AG}}\ \text{SUBJ}\ \text{OBJ}_{\text{GOAL}}\ \text{OBJ}_{\text{COM}} \rangle\text{'} \\ (\uparrow\ \text{TENSE}) = \text{PAST} \\ (\uparrow\ \text{SUBJ}\ \ \text{NUM}) = \text{PL} \end{array}$$

The cross-reference prefixes act as agreement markers and, optionally, as incorporated pronouns (since this is a *pro*-drop language):

(16) *ə-* $\mathrm{V}_{\mathrm{cref}}$
$$\begin{array}{l} ((\uparrow \text{PRED}) = \text{'PRO'}) \\ (\uparrow \text{PERS}) = 3 \\ (\uparrow \text{NUM}) = \text{SG} \end{array}$$

Finally, $\mathrm{V}_{\mathrm{post}}$ nodes define the PCASE feature that ensures correct grammatical function mapping in the way that it is usually done in LFG analyses of semantically null/case-like adpositions. For example, the following lexical entry defines *de-* as a comitative applicative prefix:

(17) *de-* $\mathrm{V}_{\mathrm{post}}$ $(\uparrow \text{PCASE}) = \text{OBJ}_{\text{COM}}$

At clause level, we assume a non-configurational structure:[17]

[14] The role of the "directive" prefix ($\mathrm{V}_{\mathrm{dir}}$) is not relevant here. In general, the view of West Circassian morphology presented herein is simplified and only serves expository purposes.

[15] Note that the latter prefix does not have a person specification, because markers of any person can appear in this position; it is only the absolutive position that is reserved for 1st and 2nd person arguments only.

[16] We ignore the minor morphological differences between "direct object" and "indirect object" prefixes in Table 1 for the purposes of this discussion.

[17] This is an oversimplification, given the facts described in section 3.1.1. However, a full analysis of West Circassian clause structure requires a separate study that is beyond the scope of this paper.

(18) S $\rightarrow$ $\underset{(\uparrow \textsc{gf})=\downarrow}{\text{NP*}}$ $\underset{\uparrow=\downarrow}{\text{V}}$

Nouns have a simple internal structure that consists of the stem and an optional "case" (Absolutive/Oblique) marker. The stems only introduce PRED and NUM features.[18]

(19) N $\rightarrow$ $\underset{\uparrow=\downarrow}{\text{N}_{\text{stem}}}$ $\underset{\uparrow=\downarrow}{\text{N}_{\text{case}}}$

(20) *č̣'ale* N_{stem} $(\uparrow \textsc{pred}) = \text{‘boy’}$
$(\uparrow \textsc{num}) = \textsc{sg}$

Absolutive and Oblique markers directly encode the grammatical function of the NP (SUBJ for "Absolutive" *-r*, $\textsc{obj}_\theta$ for "Oblique" *-m*), in a Constructive Case (Nordlinger 1998) fashion:

(21) *-m* N_{case} $(\textsc{obj}_\theta \uparrow)$

(22) *-r* N_{case} $(\textsc{subj} \uparrow)$

This simple system correctly describes the case marking and indexing pattern when both are present. Unfortunately, it has a serious problem: namely, it does not make verbal indexing of Oblique arguments obligatory, licensing ungrammatical examples such as (23b) alongside the grammatical (23a):

(23) a. *č̣'ale-m apč'ə-r* ***ə**-q^wəta-ʁ*
boy-OBL glass-ABS 3SG.ERG-break-PST
'The boy broke the glass.' (Arkadiev et al. 2009, 73)

b. **č̣'ale-m apč'ə-r q^wəta-ʁ*
boy-OBL glass-ABS break-PST

To capture this obligatoriness in the syntax, two options are available. First, a special case-like feature can be introduced by the prefixes and checked by oblique NPs using constraining equations: this will ensure that oblique NPs only occur when there is a corresponding prefix on the verb. An alternative, suggested by an anonymous reviewer, is to use another, already existing feature (such as person or number) in the same way to ensure coindexation.

The latter approach seems preferable, since it avoids stipulating an extra feature solely for the purposes of indexation. We believe that it is person that should be used as a checking feature. In fact, number cannot be used in this function because there are number mismatches with distributive NPs, which are morphologically singular but can occur with a plural prefix (Bagirokova, Lander, and Phelan, n.d.); this was earlier described for the closely related Besleney Kabardian in Arkadiev and Lander (2013). More substantially, examples like (9) suggest that nouns in West Circassian are actually unmarked for person and receive this feature from the verbal prefixes, such that a noun can appear in an argument posi-

[18] Number is also agglutinatively expressed, and number morphemes could be described as occupying their own N_{num} nodes.

tion coindexed for the 1st or 2nd person. This approach is also in line with earlier suggestions in Lander and Bagirokova (2017) that verbal prefixes are functionally akin to determiners. Furthermore, requiring each NP to have a person feature makes sense from a semantic point of view.

Thus, no modifications to indexing prefixes are required; they assign person features as in (16). Nominal stems should be modified to require the presence of a person feature:[19]

(24) *č̣'ale* $\mathrm{N}_{\mathrm{stem}}$ $(\uparrow \textsc{pred}) = \text{'boy'}$ (from ex. 20)
$(\uparrow \textsc{num}) = \textsc{sg}$
$(\uparrow \textsc{pers})$

The Absolutive must be redefined as introducing the third person feature, because it appears without verbal indexation and is only used with third person arguments:

(25) *-r* $\mathrm{N}_{\mathrm{case}}$ $(\textsc{subj} \uparrow)$ (from ex. 22)
$(\uparrow \textsc{pers}) = 3$

Case marking in West Circassian is optional, which means that bare nouns can occur in any core argument position regardless of verbal indexing. In our system, this can be expressed via a disjunction in the sublexical N rule, shown in (26). If the stem is used without a suffix, it includes an optional third person definition[20]

(26) $\mathrm{N} \longrightarrow \left\{ \underset{\uparrow=\downarrow}{\mathrm{N}_{\mathrm{stem}}} \; \underset{\uparrow=\downarrow}{\mathrm{N}_{\mathrm{case}}} \;\middle|\; \underset{\substack{\uparrow=\downarrow \\ ((\uparrow \textsc{pers})=3)}}{\mathrm{N}_{\mathrm{stem}}} \right\}$

Pronouns, somewhat counterintuitively, license their person features via constraining equations rather than define them:

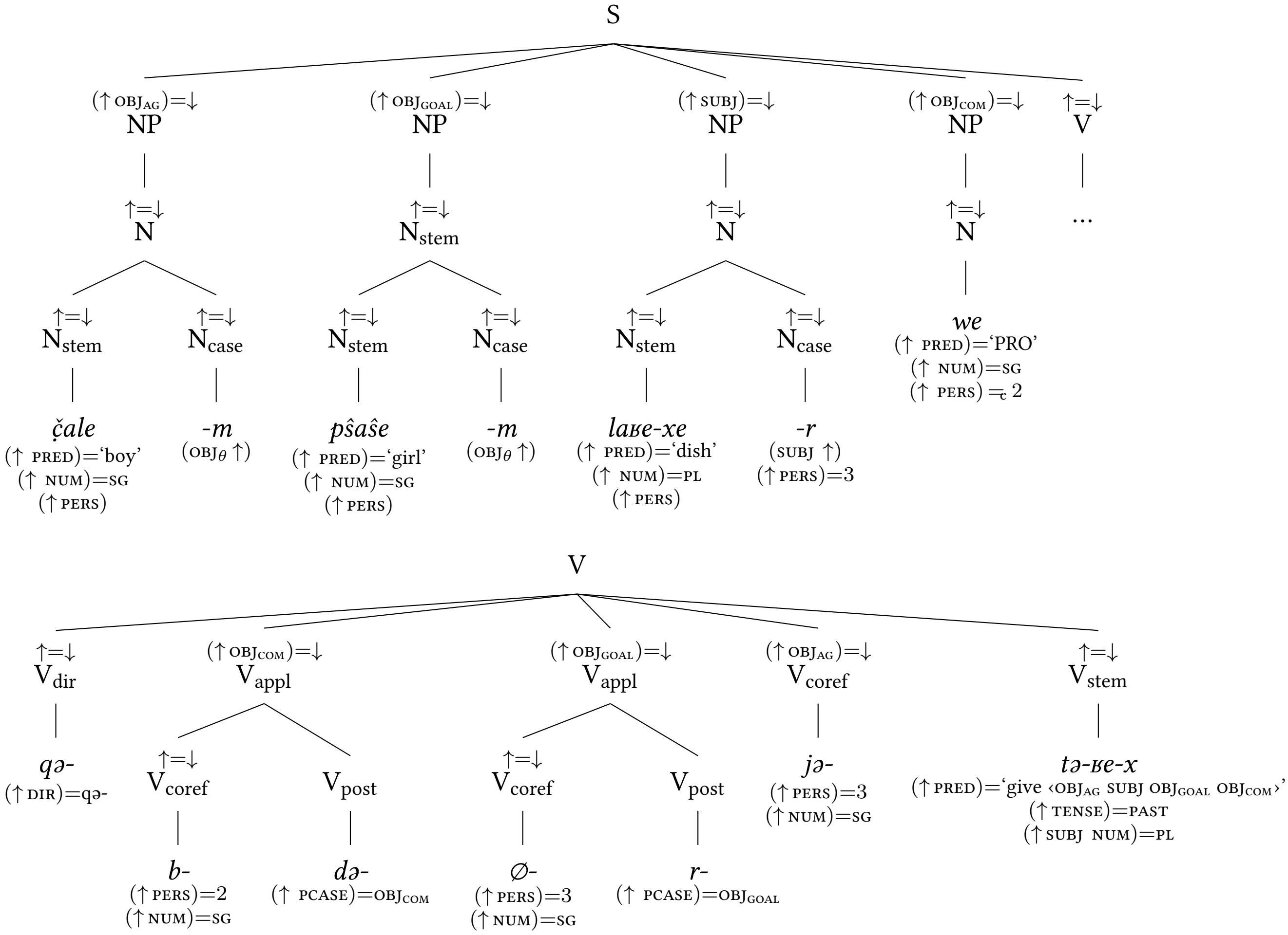
(27) *we* N $(\uparrow\ \textsc{pred}) = \text{'PRO'}$
$(\uparrow\ \textsc{pers}) =_c 2$
$(\uparrow\ \textsc{num}) = \textsc{sg}$

With these definitions, we can finally provide a full analysis for (1). The c-structure tree with annotations is given in Figure 1. It produces the f-structure in (7).

[19]This person constraint can be viewed as a syntactic placeholder for a proper semantic constraint on arguments having person (that is, being defined as speakers, hearers, or neither).

[20]A possibly more elegant way would be to capture this lexically via a morphological module like PFM (Stump 2001) coupled with a morphology-syntax interface (Dalrymple 2015), which would produce three different definitions for unmarked, absolutive and oblique noun forms.

Figure 1. Annotated c-structure for (1)



Curiously, our analysis appears to strike a balance between the pronominal argument hypothesis of polysynthesis as defined in Jelinek (1984) and Baker (1996) and the standard LFG approach where most verbal indexing markers are analyzed in essentially the same way as agreement (Austin and Bresnan 1996), unless there are clear syntactic arguments in favour of dislocation/topicalization of the indexed argument, as for the Chicheŵa verbal object marker Bresnan and Mchombo (1987). On the one hand, we generally follow the latter approach, since pronominal PRED values are only introduced optionally, in the absence of full NPs, just like verbal inflection in *pro*-drop languages. But on the other hand, the licensing of person makes the verbal prefixes more “argument-like” in the sense of their obligatoriness: they are the only elements that define this essential feature, while full NPs only check it via constraining equations. Furthermore, the internal structure of applicative markers closely mimics that of PPs; it is they that define the specific OBJ_θ function that the argument will take. Finally, non-pronominal NPs are undefined for person and only receive this feature from the verbal prefix, which leads to the grammaticality of examples such as (9).[21] If the analysis is augmented with semantics, meaning constructors associated with person (and possibly other features) will be introduced by the prefixes, not by the nouns. Verbal prefixes thus work somewhat akin to determiners in languages that have articles: They do not define the lexical content of NPs, but are obligatory and contribute essential semantic information. In that, they differ strikingly from ordinary agreement morphemes.

5 Conclusion

West Circassian presents an interesting pattern of case marking, indexing and GF assignment that does not quite fit any of the well-known alignment types. While the core system is syntactically ergative ($\mathcal{S}/\mathcal{P}$ has SUBJ status), unlike most syntactically ergative systems, $\mathcal{A}$ does not have any special syntactic role ($\hat{\text{GF}}$ in Falk 2006, CORE in Manning 1996); all arguments indexed on the verb (direct objects, indirect objects, obliques) are OBJ_θ. Furthermore, GF assignment is different for 1st and 2nd person arguments, which never have SUBJ status and are mapped to OBJ_θ ($\mathcal{A}$, $\mathcal{S}/\mathcal{P}$, applicative arguments) or OBL_θ (postpositional phrases). This means that the Subject Condition (Bresnan and Kanerva 1989) is violated, which is not without precedent (Falk 2006; Kibort 2006), but unusual in this case because the violation is not lexically determined, but person-dependent.

The sketch analysis we propose in this paper mainly views verbal prefixes as agreement morphemes, but has certain features that resemble the pronominal argument hypothesis (Jelinek 1984; Baker 1996), in particular the fact that it is the verbal prefixes, not the nominal heads, that define the person of lexical nouns and pronouns; the lexical heads only constrain the person value. In this sense, verbal prefixes may also be said to resemble determiners, as proposed in Lander and Bagirokova (2017). A semantic analysis of West Circassian case marking and indexing will be essential for exploring this resemblance in more detail.

An open question that remains is how this analysis can be reconciled with Lexical Mapping Theory (LMT, Bresnan and Kanerva 1989; Kibort 2014). Syntactic ergativity, understood as inverse mapping (Manning 1996), is rather straightforward to implement: In terms of Kibort (2007), Patients/Themes map to arg_1 [-r], Agents to arg_3 [+o]; arg_2 is not used. However, LMT does not allow OBJ_θ [+o, +r] to be present in the absence of OBJ [+o, -r]: the highest-ranking [+o] should map to OBJ, and the highest-ranking [+r] to OBL_θ, according to the Markedness Hierarchy. A possible solution is to state that the OBJ function is simply unavailable in West

[21] An anonymous reviewer observes that it is counter-intuitive to propose that person is not inherent in nouns, especially in light of constructions where nouns do not appear in the presence of a verb yet receive a third-person interpretation (appositives, answers to questions). But the statement that nouns are unmarked for person only applies to forms that include case marking, which is optional in West Circassian. Forms unmarked for case can appear in any position – both Absolutive and Oblique – and do have an (optional) inherent third person feature, as shown in (26).

Circassian, hence arguments map to the next available slot(s) on the Markedness Hierarchy.[22]

Another problem is the effect of person on semantic role to GF mapping. Such constraints have not been formalized in current versions of LMT. More importantly, the change from SUBJ to OBJ_θ is impossible in the LMT system, as there is no way to transform [-r] to OBJ_θ [+o, +r] due to conflicting features. Only the change to OBJ [+o, -r] is possible, which is clearly not what is required. A possible solution is to state that person in West Circassian influences inherent feature specifications; such a solution, however, seems to be contrary to the general ideas behind LMT.

Finally, West Circassian verb morphology, the semantics of applicative marking, the nature of these "derived" arguments, and differences between Oblique-marked and postposition-marked arguments, will have to be worked out in future analyses.

References

Arkadiev, Peter, and Yury Lander. 2013. "Non-quantificational distributive quantifiers in Besleney Kabardian." *Snippets* 27 (September): 5–7.

Arkadiev, Peter M., Yury A. Lander, Alexander B. Letuchiy, Nina R. Sumbatova, and Yakov G. Testelets. 2009. "Vvedenie: Osnovnye svedenija ob adygejskom jazyke" [Introduction. General information on Adyghe]. In *Aspekty polisintetizma: Očerki po grammatike adygejskogo jazyka,* [Aspects of polysynthesis. Explorations in Adyghe grammar], ed. by Yakov G. Testelets, in collab. with Peter M. Arkadiev, Alexander B. Letuchiy, and Nina R. Sumbatova. Moscow: Izdatel'stvo RGGU.

Arkadiev, Peter M., and Yakov G. Testelets. 2019. "Differential nominal marking in Circassian." *Studies in Language* 43 (3): 715–751.

Austin, Peter, and Joan Bresnan. 1996. "Non-configurationality in Australian aboriginal languages." *Natural Language and Linguistic Theory* 14 (2): 215–268.

Bagirokova, Irina, Yury Lander, and Paul Phelan. n.d. "Number in West Circassian." In *Number in the World's Languages,* ed. by Paolo Acquaviva and Michael Daniel. Toappear, Berlin: Mouton de Gruyter.

Baker, Mark. 1996. *The Polysynthesis Parameter.* M: Blackwell Publishers.

Beliaeva, Anna. 2006. "Konstrukcii s postpozitivnoj utočnjajuščej gruppoj" [Constructions with postposed appositive modifiers]. Handout from a talk at the Adyghe Seminar, July 15, 2006.

[22]Clauses with OBJ_θ in the absence of OBJ are claimed to occur even in languages which generally have the OBJ function, notably in Dalrymple and Nikolaeva (2011) for unmarked direct objects in differential object marking systems. Hence, a more general solution may be required anyway.

Börjars, Kersti, and Nigel Vincent. 2008. "Objects and OBJ." In *Proceedings of the LFG '08 Conference,* ed. by Miriam Butt and Tracy Holloway King, 150–168. S: CSLI Publications.

Bresnan, Joan, Ash Asudeh, Ida Toivonen, and Stephen Wechsler. 2016. *Lexical-functional syntax.* Second edition. Blackwell Textbooks in Linguistics 16. Oxford: Wiley Blackwell.

Bresnan, Joan, and Jonni M. Kanerva. 1989. "Locative inversion in Chichewa: A case study of factorization in grammar." *Linguistic Inquiry* 20 (1): 1–50.

Bresnan, Joan, and Sam A. Mchombo. 1987. "Topic, pronoun, and agreement in Chichewa." *Language* 63 (4): 741–782.

Bresnan, Joan, and Lioba Moshi. 1990. "Object asymmetries in comparative Bantu syntax." *Linguistic Inquiry* 21 (2): 147–185.

Dahlstrom, Amy. 2009. "OBJθ without OBJ: A typology of Meskwaki objects." In *Proceedings of the LFG '09 Conference,* ed. by Miriam Butt and Tracy Holloway King, 222–239. S: CSLI Publications.

Dalrymple, Mary. 2015. "Morphology in the LFG architecture." In *Proceedings of the LFG15 Conference,* ed. by Miriam Butt and Tracy Holloway King, 64–83. Stanford, CA: CSLI Publications.

Dalrymple, Mary, John J. Lowe, and Louise Mycock. 2019. *The Oxford reference guide to Lexical Functional Grammar.* Oxford: Oxford University Press.

Dalrymple, Mary, and Irina Nikolaeva. 2011. *Objects and information structure.* Cambridge Studies in Linguistics 131. Cambridge: Cambridge University Press.

Ershova, Ksenia. 2019. "Syntactic Ergativity in West Circassian." PhD diss., University of Chicago.

Falk, Yehuda N. 2006. *Subjects and Universal Grammar: an explanatory theory.* Cambridge: Cambridge University Press.

Haspelmath, Martin. 2019. "Indexing and flagging, and head and dependent marking." *Te Reo* 62, no. 1 (Special issue in honour of Frantisek Lichtenberk): 93–115.

Jelinek, Eloise. 1984. "Empty categories, case, and configurationality." *Natural Language and Linguistic Theory* 2:39–76.

Kibort, Anna. 2006. "On three different types of subjectlessness and how to model them in LFG." In *Proceedings of the LFG '06 Conference,* ed. by Miriam Butt and Tracy Holloway King. S: CSLI Publications.

Kibort, Anna. 2007. "Extending the applicability of Lexical Mapping Theory." In *Proceedings of the LFG '07 Conference,* ed. by Miriam Butt and Tracy Holloway King, 250–270. Stanford, CA: CSLI Publications.

Kibort, Anna. 2014. "Mapping out a construction inventory with Lexical Mapping Theory," ed. by Miriam Butt and Tracy Holloway King, 262–282. Stanford, CA: CSLI Publications.

Kumakhov, Mukhadin, and Karina Vamling. 2019. *Circassian Clause Structure.* Malmö: Malmö University.

Lander, Yury A. 2005. "«Pronominal'nye argumenty» i «ad"junktnye imennye gruppy» v adygejskom jazyke" ["Pronominal arguments" and "adjunct noun phrases" in West Circassian]. In *Vtoraja konferencija po tipologii i grammatike dlja molodyx issledovatelej. Materialy,* 90–95. Saint Petersburg.

Lander, Yury. 2010. "Subject properties of the Adyghe absolutive: evidence from relatives." Unpublished manuscript.

Lander, Yury A. 2012. "Reljativizacija v polisintetičeskom jazyke: Adygejskie otnosite'lnye konstrukcii v tipologičeskoj perspektive" [Relativization in a polysynthetic language. Adyghe relative clauses in typological perspective]. Cand. thesis, Russian State University for the Humanities.

Lander, Yury A. 2015. "Aktanty i sirkonstanty v morfologii i sintaksise adygejskogo jazyka." [Arguments and adjuncts in the morphology and syntax of Adyghe]. *Vestnik RGGU. Serija: Istorija. Filologija. Kul'turologija. Vostokovedenie: [Russian State University for the Humanities Bulletin]* 1:7–31.

Lander, Yury. 2016. "Adyghe." In *Word-Formation: An International Handbook of the Languages of Europe,* ed. by Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen, and Franz Rainer, vol. Vol. 5, 3508–3527. HSK 40.5. Berlin: Mouton de Gruyter.

Lander, Yury, and Irina Bagirokova. 2017. "Pronominal cross-reference in Circassian: description and theory." Conference presentation, Debrecen Workshop on Pronouns, University of Debrecen, February 25, 2017.

Lander, Yury, and Michael Daniel. 2019. "West Circassian relative prefixes as resumptives." *Linguistics* 57 (6): 1239–1270.

Lander, Yury, and Alexander Letuchiy. 2017. "Valency-decreasing operations in a valency-increasing language?" In *Verb valency changes: Theoretical and typological perspectives,* ed. by Albert Álvarez González and Ìa Navarro, 286–304. Typological Studies in Language 120. Amsterdam: John Benjamins.

Lander, Yury, and Yakov Testelets. 2017. "Adyghe." In *The Oxford Handbook of Polysynthesis,* ed. by Michael Fortescue, Marianne Mithun, and Nicholas Evans, 948–970. Oxford: Oxford University Press.

Letuchiy, Alexander. 2012. "Ergativity in the Adyghe system of valency-changing derivations." In *Ergativity, valency and voice,* ed. by Gilles Authier and Katharina Haude, 323–354. Belin: Mouton de Gruyter.

Manning, Christopher D. 1996. *Ergativity: Argument structure and grammatical relations.* S: CSLI Publications.

Nordlinger, Rachel. 1998. *Constructive case: evidence from Australian languages.* Dissertations in Linguistics. Stanford: CSLI Publications.

Rogava, Georgij V., and Zajnab I. Keraševa. 1966. *Grammatika adygejskogo jazyka* [A grammar of Adyghe]. Krasnodar and Majkop: Krasnodarskoe knižnoe izdatel'stvo.

Stump, Gregory T. 2001. *Inflectional morphology: a theory of paradigm structure.* Cambridge Studies in Linguistics 93. Cambridge: Cambridge.

Testelets, Yakov G. 2009. "Nevyražennye aktanty v polipredikativnoj konstrukcii" [Unexpressed arguments in complex clauses]. In *Aspekty polisintetizma: Očerki po grammatike adygejskogo jazyka,* ed. by Yakov G. Testelec, in collab. with Peter M. Arkadiev, Alexander B. Letuchiy, and Nina R. Sumbatova, 654–711. Moscow: Izdatel'stvo RGGU.

Raising and passive in Sanskrit

John J. Lowe
University of Oxford

Adriana Molina-Muñoz
University of Oxford

Antonia Ruppel
University of Oxford and LMU München

Proceedings of the LFG'21 Conference

On-Line

Miriam Butt, Jamie Y. Findlay, Ida Toivonen (Editors)

2021

CSLI Publications

pages 243–263

http://csli-publications.stanford.edu/LFG/2021

Keywords: control, raising, passive, Sanskrit, subject condition, Visser's Generalization.

Lowe, John J., Molina-Muñoz, Adriana, & Ruppel, Antonia. 2021. Raising and passive in Sanskrit. In Butt, Miriam, Findlay, Jamie Y., & Toivonen, Ida (Eds.), *Proceedings of the LFG'21 Conference, On-Line*, 243–263. Stanford, CA: CSLI Publications.

Abstract

In this paper we present and analyse data for a set of Sanskrit constructions involving the passive of raising / functional control verbs. Our analysis has theoretical consequences for the analysis of control and raising in LFG, and bears on the so-called 'Subject Condition' (Bresnan and Kanerva 1989) and Visser's Generalization (Bresnan 1982).

1 Preliminaries

In this paper we explore the syntax of functional control constructions in Sanskrit, with particular reference to the evidence provided by passive control structures. The type of construction which we focus on in this paper is illustrated in (1), though there are alternative passive constructions to that shown in (1b) which will be introduced fully below.[1]

(1) a. *rājāno* *rāmaṃ* *hantuṃ* *na* *śaknuvanti*
kings.NOM.PL.M R.ACC.SG.M slay.INF not can.3PL
'The kings cannot slay Rāma.'

b. *rāmo* *rājabhir* *hantuṃ* *na* *śakyate*
R.NOM.SG.M kings.INS.PL.M slay.INF not can.PASS.3SG
'Rāma cannot be slain by the kings.'

We begin in this section by introducing the two main morphosyntactic categories relevant for the present paper: the infinitive (§1.1), the morphological category of the predicate of controlled complement clauses; and the passive (§1.2). In §2 we present the data for complement control structures in Sanskrit; in §3 we discuss the LFG analysis. In §4 we conclude.

1.1 The infinitive

The Classical Sanskrit infinitive is a common non-finite verb form, used for the verbal predicates of a) complement clauses of certain, mainly modal, predicates, b) purposive adjunct clauses, and c) clauses dependent on certain nouns/adjectives. In this paper our focus is exclusively on infinitival complement clauses, as in (1a), and their passives as in (1b).

[†]We are grateful to Miriam Butt, Agnieszka Patejuk, Ash Asudeh and the audience at LFG21 for discussions of different aspects of this work. We also thank the reviewers for their comments. This work is part of the project 'Uncovering Sanskrit Syntax', funded as a Research Project Grant (RPG-2018-157) by the Leverhulme Trust.

[1]In this paper we mix constructed examples, as in (1), with examples from corpus searches which served as the basis of our empirical investigations. Corpus examples are attributed to particular texts; constructed examples are unattributed.

Sanskrit distinguishes three voices or diatheses: active, passive, and middle (self-beneficial/reflexive).[2] The passive and middle are often syncretic, but otherwise these voices are morphologically fully distinct in all finite categories and in the most common non-finite category, the participles. The infinitive, however, does not distinguish voice. By default, the infinitive adopts the active voice, and there is no way in Classical Sanskrit to express an explicitly passive sense with an infinitive, with the exception of the constructions discussed in this paper; see table 1.

Table 1: The Sanskrit voice system

	Finite	Participle	Infinitive
Active	*pacati* 'he cooks'	*pacant-* 'cooking'	*paktum* 'to cook'
Middle	*pacate* 'he cooks (for himself)'	*pacamāna-* 'cooking (for oneself)'	*'to cook (for oneself)'
Passive	*pacyate* 'it is cooked'	*pacyamāna-* 'being cooked'	*'to be cooked'

However, a passive reading of the infinitive is obligatory in infinitival clauses which are arguments of morphologically passive verbs, as in (1b).[3] We therefore assume that the active reading of the infinitive is a default, which is overridden in certain syntactic contexts.[4]

1.2 The passive

Finite and participial passives always function in opposition to a corresponding active finite or participial form. But Sanskrit also has an exclusively passive construction, the 'gerundive', a nonfinite (morphologically adjectival) form which has a usually deontic modal sense. For example, beside the forms of *pac* 'cook' given above, a gerundive *paktavya-* '(fit/intended) to be cooked' can be formed. There is no corresponding active or middle formation. Despite being (morphologically) nonfinite, the gerundive is very common as a main clause predicate; the majority of our data below involves gerundival matrix clauses.

[2]Often in Classical Sanskrit the self-beneficial/reflexive sense of the middle is weak, and it is functionally all but equivalent to the active. It is the difference between active and passive which matters for our purposes.

[3]On the passive reading of the infinitive see Oberlies (2003b, 276–278), who cites also Whitney (1896, §988) and Speyer (1896, 65–66) for the same occasional passive use in Vedic.

[4]The passive reading of an infinitive is also optional when an infinitival clause is an adjunct (purposive) to a passive verb. We do not analyse that here. In a few exceptionally rare instances – only a handful recognized in the whole of Classical Sanskrit literature – an infinitive appears to have a passive reading while not under the scope of a morphologically passive matrix verb. Most examples are from the Sanskrit epics, the language of which is less standardized than the majority of Classical Sanskrit literature. Such examples are best treated as sporadic cases of a passive interpretation overriding the default active interpretation forced by the context.

Another morphologically nonfinite verb form which is sometimes considered 'passive' is the 'past participle', often labelled the 'past passive participle' or 'perfect passive participle'. This is not in fact a truly passive formation, showing rather ergative-absolutive alignment: the past participle agrees with the patient/object-like argument of transitive verbs (O), like a standard passive, but with the single subject-like argument (S) in the case of intransitive verbs. This contrasts with true passives (including the gerundive), which are freely formed to intransitive verbs in Sanskrit, resulting in impersonal constructions with default third person singular or neuter singular morphology. Table (2) contrasts the transitive verb *pac* 'cook' with the intransitive *svap* 'sleep'; the finite passive and gerundive illustrate the true passive alignment, and while the past participle mirrors the argument alignment of the true passives in the case of *pac*, it mirrors the active in the case of *svap*.

Table 2: Argument alignment in Sanskrit

	transitive	intransitive
active	*pacati* 'A cooks O'	*svapiti* 'S sleeps'
fin. passive	*pacyate* 'O is cooked (by A)'	*supyate* 'It is slept (by S)'
gerundive	*paktavyam* '(O is) to be cooked (by A)'	*svaptavyam* 'It is to be slept (by S)'
past ptc.	*pakva-* '(O) has/having been cooked (by A)'	*supta-* '(S) has/having slept'

Nevertheless, there are certain complications in distinguishing passive from ergative in Sanskrit. Firstly, the past participle occasionally shows passive alignment; that is, impersonal passives to intransitive verbs are sometimes found even with the past participle. So a construction like *tena suptam* lit. 'it was slept by him' is possible, alongside the standard *sa suptaḥ* 'he (has) slept'. This is likely analogical on the finite passive, but in any case prevents us from entirely excluding the label 'passive' for the past participle. Secondly, most subject tests in Sanskrit target the most agentive argument rather than what we might consider the 'grammatical subject', making it hard to prove, for example, whether the promoted patient or the demoted agent of a finite passive is the grammatical subject. In the case of complement control, past participles of transitive raising/control verbs have exactly the same effect on the voice interpretation of the controlled infinitive as finite passives: the infinitive must be interpreted as passive. At least in this respect, then, the past participle of transitive verbs is functionally passive. In this paper we focus on the verb *śak*, which is intransitive, but the constructions we analyse are perfectly possible with a transitive verb in the past participle. The following example illustrates this with the transitive verb *yuj* 'join', which in the passive can mean 'is fitting, is possible'.

(2) *sa* *te* *daṇḍayituṃ* *yuktaḥ*
he.NOM.SG.M you.GEN punish.INF join.PST.PTC.NOM.SG.M
'He ought to be punished by you.' (*Kathāsaritsāgara* 9.2.114)

2 Raising/control in Sanskrit

2.1 The categories of verbs

Pāṇini, the ancient Indian grammarian whose Sanskrit grammar, the *Aṣṭādhyāyī*, was both a highly sophisticated generative description of late Vedic Sanskrit and a standard for prescriptive use for the Classical language, specifies a number of semantic categories of verb which govern infinitive clauses: verbs of 'desiring' (e.g. *iṣ* 'want, desire'); verbs of 'ability' (e.g. *śak* 'can, be able'); verbs of 'daring' (e.g. *dhṛṣ* 'dare'); verbs of 'knowing' (e.g. *jñā* 'know'); verbs of 'aversion' (e.g. *glai* 'be averse, dislike'); verbs of 'striving' (e.g. *ghaṭ* 'strive, endeavour'); verbs of 'beginning' (e.g. *rabh* 'begin'); verbs of 'success/permission' (e.g. *labh* 'succeed, have permission'); verbs of 'undertaking' (e.g. *kram* 'undertake, set out'); verbs of 'capability' (e.g. *sah* 'have power, be capable'); verbs of 'deserving' (e.g. *arh* 'be worthy, deserve'); verbs of 'being' (e.g. *as* 'be').[5]

Almost all of these verb classes occur in control constructions only as subject control predicates. In our corpus, the only exceptions are certain preverb-verb combinations involving *jñā* 'know': *anu-jñā* 'permit' and *ā-jñā* 'command' show object control; in its simplex form and with other preverbs, *jñā* shows only subject control. In this paper we consider only subject control.

An important distinction must be drawn between 'raising' and 'equi' verbs, that is between verbs which place semantic constraints on their subject argument, i.e. which have thematic subjects, and those which do not, i.e. which have non-thematic subjects.[6] It has not been previously noted that in Sanskrit only raising verbs, i.e. verbs with non-thematic subjects, are at all common in the passive; control predicates with thematic subjects, such as the otherwise highly frequent *iṣ* 'want', are distinctly rare in the passive.[7]

The empirical investigation which served as the foundation of this project was based on an electronic corpus of around 5.5 million words.[8] Our corpus contains

[5] Aṣṭādhyāyī 3.3.158, 3.4.65.

[6] By 'subject' here we mean subject in the active, i.e. in argument structure terms the arg_1, not the grammatical function SUBJ.

[7] When used as control predicates, that is. The passive of *iṣ* 'want' as a simple transitive verb is common.

[8] The corpus comprises texts from a broad variety of genres and periods of Sanskrit. It includes c. 1.3 million words of late Vedic text, c. 1.7 million words of Epic and c. 2.5 million words of various genres of Classical (i.e. post-Pāṇinian) texts dating as late as the 13th century AD. The 'Vedic' texts are restricted to the later Vedic prose texts (Brāhmaṇas, Āraṇyakas and Upaniṣads), which are linguistically much closer to Classical Sanskrit than early Vedic, and represent a form of the language particularly close to that which Pāṇini's grammar set out to describe. The 'Epic' texts are based on oral traditions whose origins predate Pāṇini but that, in their final form, employ a language mostly

1,071 tokens of passive raising constructions: 879 tokens with the passive of *śak* ‘can’; 159 tokens with the passive of *yuj* ‘join’ (always with *na* ‘not’ meaning ‘not fit to, not able to’); 23 tokens with the (gerundive-only) predicate *nyāyya* ‘be proper’; and 10 tokens with the passive of *rabh* ‘begin’. Although there are more equi verbs overall, few occur in the passive: we identified 9 tokens with *iṣ* ‘want’, 4 tokens with *jñā* ‘know’, and one token with *īh* ‘desire’. We are not certain of the status of *labh* ‘have opportunity’ (i.e. whether or not its subject is thematic), of which we identified ten relevant passive tokens.

For reasons of space, in this paper we do not address the analysis of equi/control verbs, restricting ourselves to raising verbs. While the phenomena to be analysed, including the possible passive constructions, are superficially similar, it is worth noting that the analysis we propose for verbs like *śak* ‘can’ depends on the non-thematic status of the active subject, and could not be extended to verbs like *iṣ* ‘want’. We hope to address the latter in future work.

2.2 The passive constructions

When a raising verb is active, there is only one possible construction and interpretation, as in the constructed example (1b) and in (3), from our corpus:

(3) *na bhīṣmaṃ pāṇḍavā aśaknuvan raṇe jetum*
not Bh.ACC P.NOM.PL can.IMPF.3PL battle.LOC conquer.INF
‘The Pāṇḍavas could not conquer Bhīṣma in battle.’
(*Mahābhārata* 6.105.10)

As discussed, subject control is obligatory. As seen in (3), Sanskrit is a non-configurational language, and there is no requirement for the infinitival complement clause to form a single constituent in the c-structure.

In the passive, there are three possible constructions, all apparently semantically equivalent. We begin with examples of the gerundive *śakya-* ‘able to be done’, which most clearly and commonly attests all three variants.[9] The following examples are from our corpus and all involve the same logical object of the infinitive, the first person pronoun.

(4) *na ahaṃ vedair na tapasā na dānena na ca*
not I.NOM Veda.INS.PL not asceticism.INS not generosity.INS not and
ijyayā śakya evaṃvidho draṣṭum
reverence.INS can.GDV.NOM.SG.M such.NOM.SG.M see.INF
‘I cannot be seen in this way, neither through the Vedas, nor asceticism, nor generosity, nor reverence.’ (*Mahābhārata* 6.33.53)

following Pāṇinian rules; the ‘Classical’ corpus covers a range of textual genres (narrative literature, poetry, drama, śāstra (= technical literature in a variety of fields) and religious texts).

[9]Beside the root *śak*, all three constructions are also attested with the similar raising predicates *yuj*.PASS ‘it is fitting’ and *nyāyya-* ‘it is proper’.

(5) *anyathā nahi mām̩ draṣṭum̩ śakyam*
otherwise not I.ACC see.INF can.GDV.SG.NT
'Otherwise no one can see me.' (*Kūrmapurāṇa* 2.10.4)

(6) *na śakyam̩ mānavair draṣṭum r̥te dhyānād aham̩ tv iha*
not can.GDV.NT.SG men.INS see.INF without meditation.ABL I.NOM but here
'But without meditation men cannot see me here.' (*Liṅgapurāṇa* 1.24.8)

In (4), the matrix predicate *śakya* agrees with the pronoun *aham*, which is functionally the arg_2 of the infinitive, but appears here in the nominative as the arg_1 of *śakya*. The argument with which a verb agrees is the SUBJ, in Sanskrit, so we appear to be dealing with a kind of raising to subject of an argument of the complement clause. This is the most common construction with *śakya-*, accounting for 74% of unambiguous instances. We refer to this as the 'agreeing type'.

Alternatively, as in (5), the arg_2 of the infinitive may appear in the accusative case, with the gerundive in the form *śakyam*. This neuter singular form of the gerundive is a default form, used when there is no agreeing subject, e.g. also in impersonal gerundive constructions (i.e. gerundives to intransitive verbs, as *svaptavyam* in table 2). This is relatively rare, constituting only 3% of the unambiguous instances of *śakya-* in our data. Note we will argue below that the infinitival clause is not the subject of the passive matrix verb, meaning that we cannot translate this construction as something like 'To see me cannot be done'; its use and sense are indistinguishable from the types in (4) and (6). We refer to this as the 'accusative type'.

Thirdly, as in (6), the gerundive may apparently occur in the default neuter singular form, but with the object of the infinitive in the nominative. We refer to this as the 'non-agreeing type'. If the gerundive really is showing default neuter singular agreement in this case, it is a highly problematic construction, since agreement between the gerundive and its subject is obligatory, and there is no way to explain the nominative case of the infinitive's object except by treating it as the grammatical subject of the matrix clause. (Infinitives alone can never license nominative arguments, for example.) A simpler alternative here is that *śakyam*, at least in these instances, is an invariant predicate with no agreement properties. That is, rather than being an instance of the gerundive *śakya-* in the neuter singular, it is a separate invariant predicate *śakyam* which, like other invariant predicates in Sanskrit, can appear with a nominative subject with which it shows no agreement.

Fourthly, we may have ambiguous cases. In Sanskrit, nominative and accusative cases are syncretic in the neuter gender. This, and the fact that the default non-agreeing form of the gerundive is neuter singular, means that if the logical object of the infinitive is a neuter singular noun, the three constructions introduced above are indistinguishable. Such ambiguous cases are rather common, making up 30% of all constructions with *śakya-*. The following example is from our corpus.

(7) *na cec chakyam atha utsraṣṭuṃ vairam etat*
not if can.GDV.SG.NT but renounce.INF enmity.SG.NT this.SG.NT
sudāruṇam
terrible.SG.NT
'If this terrible enmity cannot be renounced...' (*Mahābhārata* 6.117.29)

As suggested by Gippert (1995), this ambiguity may be the origin of the existence of multiple constructions. Gippert assumes that what we call the agreeing type is the original pattern, with the accusative and non-agreeing types created on the basis of ambiguous constructions like (7). However, as argued below it is the accusative type which is the theoretically expected passive construction, so we would rather assume that this was the original type, and that the nominative and non-agreeing types were extracted from ambiguous structures (with the nominative type becoming predominant). In any case, the diachronic situation is not relevant for the synchronic analysis which we pursue in this paper.

Fifthly, we may simply lack any logical object. When the infinitival predicate is intransitive, it has no arg_2 to appear in either the nominative or accusative. Necessarily, the gerundive then appears in the default neuter singular. This type makes up 12% of the gerundive data.

(8) *śakyam idānīm āśvāsitum*
can.GDV.SG.NT now breathe.INF
'Now (we) can breathe.' (Lit. 'it can be breathed (by us).') (*Śakuntalā* 4.1)

With all these constructions, any agent of the infinitive is expressed in the instrumental, as exemplified in (6).

The five constructions illustrated above with the gerundive *śakyam* are also attested with other raising and control verbs, and also with the finite passive of *śak*, with the exception of the type in (6), which never occurs with finite passives. Table (3) shows the distribution of passive types with finite and gerundive forms of *śak*. That the non-agreeing type is unattested with finite verbs supports the argument that, where this is found with *śakyam*, *śakyam* is an invariant unagreeing predicate rather than a nt.sg. form of the gerundive; finite verb forms are never used as unagreeing predicates in Sanskrit, so this would explain the gap. Whereas if it were possible for an agreeing neuter singular gerundive to appear with a non-neuter and/or non-singular subject, the same ought in principle to be possible for the 3sg. finite verb.

3 Analysing *śak*

For ease of comparison, in this section we provide analyses for constructed examples. We begin with the active sentence in (9). We assume the f-structure for this in (10):

Table 3: Passive types with *śak*

śak	Agreeing	Accusative	Non-agreeing	Ambig.	Intr.
Finite	86	4	0	66	18
Gerndv. (non-nt.)	365	0	0	0	0
Gerndv. (*śakyam*)	0	16	28	212	84
Total	451	20	28	278	102

(9) *rājāno rāmaṃ hantuṃ na śaknuvanti*
kings.NOM.PL.M R.ACC.SG.M slay.INF not can.3PL
'The kings cannot slay Rāma.'

(10)
$$\left[\begin{array}{ll} \text{PRED} & \text{`}\textit{can}\langle\text{XCOMP}\rangle\text{SUBJ'} \\ \text{NEG} & + \\ \text{SUBJ} & \boxed{1}\left[\text{PRED `}\textit{kings}\text{'}\right] \\ \text{XCOMP} & \left[\begin{array}{ll} \text{PRED} & \text{`}\textit{slay}\langle\text{SUBJ,OBJ}\rangle\text{'} \\ \text{SUBJ} & \boxed{1} \\ \text{OBJ} & \left[\text{PRED `}\textit{Rāma}\text{'}\right]\end{array}\right]\end{array}\right]$$

This f-structure reveals a number of analytical choices, which we justify in the following sections. Given the corpus-based nature of Sanskrit, there are no clear syntactic tests which would enable us to establish these choices purely on the basis of the active. This is why the passive constructions are so crucial, and we justify our analyses below primarily on the basis of the passive constructions.

Firstly, as discussed above, we take the SUBJ argument of *śak* to be non-thematic. The verb *śak* originally had a more lexical sense 'be able, have power' in pre-Classical Sanskrit, with (presumably) semantic selection of its subject argument. Its semantic bleaching was a gradual process, and the earlier sense can sometimes be read into Classical examples. But in the Classical language *śak* can take non-animate subjects, and never needs to be interpreted as taking a thematic subject; the non-thematic status of its subject is further justified below.

Secondly, we assume functional rather than anaphoric control. Functional control by a non-thematic subject of course follows the standard LFG approach to raising vs. equi (Dalrymple et al. 2019, chapter 15). No empirical criteria have been proposed for distinguishing functional from anaphoric control in Sanskrit, however (Sanskrit does not even have expletive arguments); we offer a theoretical argument below.

While active forms of *śak* are necessarily bivalent, taking a SUBJ and XCOMP (in our analysis), it is important to note that *śak* is fundamentally intransitive, in the sense of not selecting for an object argument. This is evident from the past participle, *śakta-*, which patterns in the same way as unambiguously intransitive verbs; see table (4).

Table 4: Alignment patterns in past participle

	Present active	Past participle
Monovalent intrans.: *svap* 'sleep'	*svapiti* '(S) sleeps'	*supta-* '(S) having slept'
Bivalent trans.: *han* 'slay'	*hanti* '(A) slays (O)'	*hata-* '(O) (having been) slain'
Bivalent intrans.: *śak* 'can'	*śaknoti* '(S) can (+inf.)'	*śakta-* '(S) having been able to (+inf.)'

3.1 The accusative construction

We now move on to the passive constructions, beginning with the second type introduced above, the 'accusative type', where the verb appears in the default 3sg. (or neuter singular, in the case of the gerund), and the object of the infinitive remains in the accusative.

(11) *rājabhī rāmaṃ hantuṃ na śakyate*
kings.INS R.ACC slay.INF not can.PASS.3SG
'Rāma cannot be slain by the kings.'

As discussed above, the passive of an intransitive in Sanskrit sees the active subject realised as an instrumental-case oblique argument and no explicit subject argument, the verb appearing in the default 3sg. (or nt.sg.). This passive construction therefore fits exactly with what we would expect for the passive of the intransitive raising verb *śak*.

In our approach to argument structure and the passive we adopt the 'valency template' of Kibort (2007):[10]

(12) $\langle$ $\underset{[-\text{O}/-\text{R}]}{\text{arg}_1}$ $\underset{[-\text{R}]}{\text{arg}_2}$ $\underset{[+\text{O}]}{\text{arg}_3}$ $\underset{[-\text{O}]}{\text{arg}_4}$ ... $\underset{[-\text{O}]}{\text{arg}_\text{n}}$ $\rangle$

In Kibort's (2007) approach, the passive agent is an OBL$_\theta$, rather than an ADJ. The passive is the result of a [+R] specification added to the first argument position in a valency frame which is pre-specified as [−O]. For the passive of *śak*, we require that this does not result in the XCOMP argument being promoted to subject. We therefore take XCOMP with *śak* to represent the realization of a clausal argument in the arg$_3$ position; arg$_3$ is prespecified as [+O], meaning that it can never be realized as SUBJ. To represent the difference between clausal and non-clausal

[10]We assume the formalization of Findlay (2014, 2016) underlying this, though we retain the less technical representation.

arguments, we use a feature [+C].[11] Thus in the active the argument structure of *śak* will resolve as in (13), while in the passive it will resolve as in (14).

(13)					
	'can'	⟨	arg_3	⟩	arg_1
[default]			[+O,+C]		[−O]
			XCOMP		SUBJ

(14)					
	'can'	⟨	arg_3	⟩	arg_1
[default]			[+O,+C]		[−O]
[passive]					[+R]
Mapping:			XCOMP		OBL_θ

The passive therefore produces a subjectless construction, in violation of the supposed 'Subject Condition' (Bresnan and Kanerva 1989, Berman 1999), but in line with the analysis of passives of intransitives proposed by Kibort (2006). Deshpande (1980) takes a different approach, arguing that here the infinitival phrase is the subject of the main verb. In principle this is possible, but there is no evidence for subject properties associated with the infinitival phrase, and as shown above *śak* clearly patterns as an intransitive verb in the past participle, suggesting that it should form an impersonal (subjectless) passive, as assumed here.[12]

A minor problem is the instantiation of the θ in OBL_θ. Given Kibort's approach to the passive, the demoted subject necessarily maps to OBL_θ, but in this case the arg_1 of the predicate is a non-thematic argument and so has no role with which θ can be instantiated.[13] We assume that it is possible for θ to have a null instantiation, that is OBL_0, or more precisely (though less clearly) simply OBL. The only alternative to this would be to say that Kibort's approach to the passive predicts that passives of subject raising verbs are impossible; but that is clearly not the case.

We therefore assume the following f-structure for the sentence in (11):

(15)
$$\begin{bmatrix} \text{PRED} & \textit{`can}\langle\text{XCOMP}\rangle\text{OBL}_0\textit{'} \\ \text{NEG} & + \\ \text{VOICE} & \text{PASS} \\ \text{XCOMP} & \begin{bmatrix} \text{PRED} & \textit{`slay}\langle\text{SUBJ,OBJ}\rangle\textit{'} \\ \text{SUBJ} & \boxed{1} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \textit{`Rāma'} \end{bmatrix} \end{bmatrix} \\ \text{OBL}_0 & \boxed{1}\begin{bmatrix} \text{PRED} & \textit{`kings'} \end{bmatrix} \end{bmatrix}$$

Since *śak* still selects for an XCOMP, we need a controller. The only available argument is the oblique argument, the OBL_0. There are a number of interesting consequences. Firstly, we must assume that the infinitive does not state constraints on

[11] We follow Dalrymple and Lødrup (2000) in assuming the usefulness of distinct grammatical functions for at least some clausal arguments. [+C] would of course be unnecessary if COMP and XCOMP were eliminated in line with e.g. Alsina et al. (2005).

[12] Furthermore, as pointed out to us by Agnieszka Patejuk, if an open clausal argument were to be a subject, we would have to assume control into a subject, a phenomenon not widely admitted (though see Arka and Simpson 1998, Stiebels 2007, Patejuk and Przepiórkowski 2020).

[13] We thank an anonymous reviewer for pointing this out to us.

the case of its subject; this is supported by the rare possibility of infinitives taking accusative case subjects (Oberlies 2003b, 278), alongside the standard nominative case controllers of the active construction discussed above.

Secondly, it will not be sufficient to assume a standard subject control equation such as:

(16) $(\uparrow \text{SUBJ}) = (\uparrow \text{XCOMP SUBJ})$

Such an equation will not account for both active and passive of *śak*; we will therefore require a more nuanced phrasing; this is discussed further below.

We are here considering only raising verbs. Yet in the comparable case of control verbs, (anaphoric) control by a passive agent violates Visser's Generalization, as formulated by Bresnan (1982). Falk (2006, 142) similarly claims that only core arguments, i.e. SUBJ or OBJ, may function as controllers. But as argued by van Urk (2013), Visser's Generalization applies only in the case of personal passives, i.e. where the passive control verb agrees with an explicit subject argument; in impersonal passives, oblique controllers are possible.[14] Van Urk (2013, 170) gives the following example from German:[15]

(17) *Es wurde versucht, Eichhörnchen zu fangen.*
it was tried squirrels to catch.INF
'(Lit.) It was tried to catch squirrels.'

The control relation between the implicit agent of the control verb and the PRO subject of the infinitive is obligatory here, just as in the Sanskrit example above. Thus, the Sanskrit evidence for raising verbs fully parallels the modification of Visser's Generalization proposed by van Urk (2013), suggesting that this may be a more general constraint applicable to both raising and control verbs.

Van Urk (2013) provides a derivational account of the modified Visser's Generalization. For an LFG account, we can begin with the generalization that the presence of a SUBJ argument rules out control by an OBL, but in the absence of a SUBJ, control by OBL is possible. We propose to model this below with reference to Kibort's (2007) theory of argument structure.

3.2 The agreeing type

As we argued in the previous section, the accusative type is in formal terms the 'expected' passive construction, i.e. exactly what we would predict if we applied

[14] On Visser's Generalization see also Boeckx et al. (2010, 125–141).

[15] All of van Urk's examples involve implicit agents, but in German just as in Sanskrit explicit oblique agents in this construction are unproblematic:

(i) *Es wurde von Hans versucht, Eichhörnchen zu fangen.*
it was by Hans tried squirrels to catch.INF
'(Lit.) It was tried by Hans$_i$ (e$_i$) to catch squirrels.'

standard principles of passivization to the standard active control construction. But in frequency terms, it is significantly outnumbered by the agreeing type introduced in (4), where the object of the infinitive appears in the nominative and the matrix verb shows agreement with this argument:

(18) *rāmo rājabhir hantuṃ na śakyate*
R.NOM.SG.M kings.INS.PL.M slay.INF not can.PASS.3SG
'Rāma cannot be slain by the kings.'

This is more problematic to analyze, because it is not immediately obvious how or why the object of the infinitive, which has no direct relation with the raising verb, can become its subject.

Superficially similar constructions have been discussed in an LFG setting by Ørsnes (2006) and Lødrup (2014). Ørsnes (2006) discusses the 'complex passive' in Danish, as in the following example:

(19) *bilen forsøges repareret*
the.car is.tried repaired
'As for the car, an attempt is made to repair it.' (Ørsnes 2006, 388)

Here, the logical object of 'repair' becomes the subject of the passivized control verb, parallel to the Sanskrit construction under discussion. Ørsnes (2006) assumes that passivization involves suppression of the arg_1 in the argument structure, rather than demotion, and that the subject of the (passive) embedded predicate is raised to subject of the matrix predicate in order to fulfil the Subject Condition. In contrast, we assume a demotional account of the passive, and we do not assume the Subject Condition. Moreover, we are not starting with an equi verb showing obligatory anaphoric control, but with a raising verb showing functional control, nor are we starting with an embedded predicate which is marked as passive. Our analysis must therefore differ in a number of ways, and we do not need to assume a kind of last-resort raising where there was no raising before; since we already have a functional control relation in the active, it makes sense that this same relation passes over into the passive.

Lødrup (2014) discusses a superficially similar construction in Norwegian, which he calls the 'long passive':

(20) *viktige stridsspørsmål blir unnlatt å presiseres*
important issues are neglected to clarify.INF.PASS
'They neglect clarifying important issues.' (Lødrup 2014, 368)

Lødrup (2014) shows that the long passives of Norwegian are different in certain important respects from the complex passives discussed by Ørsnes (2006). Lødrup's analysis of the long passive involves a kind of restructuring, where the control and embedded verb merge in the argument structure to form a complex predicate.

The question is now whether the Sanskrit construction should be treated by assuming restructuring; a complex predicate analysis would offer a clear alternative to the control-based analysis pursued here. In fact, Deshpande (1980) and Kiparsky (2002) both refer to the passive construction with *śak* in terms which could be taken to imply a complex predicate analysis. Deshpande (1980, 102) claims that *śak* and its dependent infinitive are "increasingly bracketed" together, "creating a sort of "compound verb" like *kar saknā* ['able to do'] in Hindi." Kiparsky (2002) similarly claims that *śak* and its dependent infinitive are treated as a single predicate, by virtue of a "verb union process". Neither author further expands or justifies these claims, however. In contrast, the descendant of *śak* in Hindi/Urdu, *saknā* 'can', is a standard raising verb which embeds an XCOMP (Bhatt et al. 2011, Butt 2014). There is no light verb version of *saknā* in Hindi/Urdu, and there is no standard path of diachronic development whereby a light verb could develop into a raising verb. Rather, the opposite development is expected. Thus the modern Indo-Aryan situation renders it highly unlikely that a complex predicate analysis should be proposed for Sanskrit *śak*.[16]

Moreover, evidence from ellipsis and negation strengthens the claim that *śak* and infinitive do not form a complex predicate. Restrictions of space prevent a detailed discussion, but most tellingly it is possible to independently negate *śak* or the infinitive, with different readings. The following phrases are both common in Patañjali's *Mahābhāṣya*, often considered a standard of clear prose Sanskrit:

(21) a. *na śakyaṃ kartum*
not can.GDV.NT.SG do.INF
'(This) cannot be done.'

b. *śakyam a-kartum*
can.GDV.NT.SG NEG-do.INF
'(This) does not need to be done.' (Lit. 'can be not done')

A complex predicate analysis is therefore not viable. We propose to analyse this 'agreeing' type by permitting the passive argument structure operations to apply not, in this case, to the matrix verb which carries the morphological marking of the passive, but rather to the infinitival predicate. As discussed above, Classical Sanskrit infinitives have a single invariant form with no voice marking, and outside of this construction show regular active syntax and semantics. Nevertheless, the interpretation of the infinitive is clearly passive in this construction. The f-structural analysis we assume is the following:

[16] We thank Miriam Butt (p.c.) for discussion of the points in this paragraph. See also Butt and Lahiri (2013) on the diachronic tendencies of light verbs.

(22)
$$\begin{bmatrix} \text{PRED} & \text{‘}\textit{can}\langle\text{XCOMP}\rangle\text{SUBJ’} \\ \text{NEG} & + \\ \text{SUBJ} & \boxed{1}\begin{bmatrix}\text{PRED} & \text{‘}\textit{Rāma}\text{’}\end{bmatrix} \\ \text{XCOMP} & \begin{bmatrix} \text{PRED} & \text{‘}\textit{slay}\langle\text{SUBJ,OBL}_{\theta}\rangle\text{’} \\ \text{VOICE} & \text{PASS} \\ \text{SUBJ} & \boxed{1} \\ \text{OBL}_{\theta} & \begin{bmatrix}\text{PRED} & \text{‘}\textit{kings}\text{’}\end{bmatrix} \end{bmatrix} \end{bmatrix}$$

The passive morphology of the raising verb can therefore be associated with functional passivity of its embedded predicate, rather than itself. For simplicity let us assume that the functional passivity, together with its argument structure consequences, is associated with an f-structure feature PASSIVE; we can then capture the variable application of the passive with *śak* very simply, by assuming that the PASSIVE feature is subject to a functional uncertainty in the lexical entry of the morphologically passive form of the raising verb:

(23) $(\uparrow\ (\text{XCOMP})\ \text{VOICE}) = \text{PASSIVE}$

The predicate of whichever f-structure gets the PASSIVE voice feature will necessarily show the associated passive argument structure operations, resulting in either the ‘accusative’ type discussed above, or the ‘agreeing’ type discussed here. Thus both types can be derived from a single point of optionality in an otherwise uniform control construction.

The analysis proposed here offers support for the non-thematic status of the subject position of *śak*: since there is no difference in the selectional properties of the verb between the active and agreeing passive types (e.g. between (10) and (22)), but the subject of the verb does change, the subject position of *śak* must be non-thematic.

In terms of the passive reading of the infinitive, despite the lack of passive morphology and the fact that infinitives cannot freely take a passive reading, we assume that the possibility of an infinitive with passive argument structure is licensed in the lexicon, but can only surface in a construction which specifies a passive reading for the infinitive. Thus infinitives cannot be used freely with a passive sense, but only when embedded under particular predicates, like the passive of *śak*, which are capable of specifying the passive voice feature of their embedded predicate. We assume that the functionally passive version of the infinitive is associated with the following specification:

(24) $\text{VOICE} =_{c} \text{PASS}$

3.3 Intransitive verbs

As illustrated in (8), when the embedded verb is intransitive, there is no embedded object argument to appear in either the nominative, as in the ‘agreeing’ construction, or in the accusative, as in the ‘accusative’ construction. An additional example, constructed for the purposes of analysis (based roughly on *Mahābhārata* 12.314.20), follows:

(25) *na tatra śakyate gantuṃ rāmeṇa*
not there can.PS.3SG go.INF R.INS
'Rāma cannot go there.' (Lit. 'it cannot be gone there by Rāma.')

Of the two analyses proposed so far, the first, the accusative type – in which *śak* undergoes passivization and its OBL$_\theta$ argument controls the embedded subject position – can unproblematically be applied to intransitive embedded verbs as well:

(26)
$$\left[\begin{array}{ll} \text{PRED} & \text{`}\textit{can}\langle\text{XCOMP}\rangle\text{OBL}_\theta\text{'} \\ \text{NEG} & + \\ \text{VOICE} & \text{PASS} \\ \text{XCOMP} & \left[\begin{array}{ll} \text{PRED} & \text{`}\textit{go}\langle\text{SUBJ}\rangle\text{'} \\ \text{SUBJ} & \boxed{1} \\ \text{ADJ} & \left\{\left[\text{PRED}\ \text{`}\textit{there}\text{'}\right]\right\} \end{array}\right] \\ \text{OBL}_\theta & \boxed{1}\left[\text{PRED}\ \text{`}\textit{Rāma}\text{'}\right] \end{array}\right]$$

If we tried to apply the analysis of the agreeing type – where the passive, which is marked morphologically on the matrix verb, applies in fact to the predicate of the embedded infinitive – we would run into problems. The single argument of the infinitive would appear as OBL$_\theta$; this OBL$_\theta$ would be necessarily case marked as instrumental, but such an argument could not then serve as the SUBJ of *śak*, since that must necessarily be nominative.

(27) Illicit structure:

$$\left[\begin{array}{ll} \text{PRED} & \text{`}\textit{can}\langle\text{XCOMP}\rangle\text{SUBJ'} \\ \text{NEG} & + \\ \text{SUBJ} & \boxed{1}\left[\text{CASE}\ \ \text{NOM}\right] \\ \text{XCOMP} & \left[\begin{array}{ll} \text{PRED} & \text{`}\textit{go}\langle\text{OBL}_\theta\rangle\text{'} \\ \text{VOICE} & \text{PASS} \\ \text{OBL}_\theta & \boxed{1}\left[\begin{array}{ll} \text{PRED} & \text{`}\textit{Rāma}\text{'} \\ \text{CASE} & \text{INSTR} \end{array}\right] \\ \text{ADJ} & \left\{\left[\text{PRED}\ \text{`}\textit{there}\text{'}\right]\right\} \end{array}\right] \end{array}\right]$$

Such an analysis is therefore impossible; it is ruled out given our assumption of functional control. If we had assumed anaphoric control – and additionally backward control (which is attested in other control structures in Sanskrit) – then the equivalent of the structure in (27) would be possible. That this should not be the case is a desirable outcome, since it eliminates an analytical ambiguity for sentences like (25). We therefore take this as a theoretical argument in favour of functional control.

3.3.1 Excursus: the active *śakyate*

In fact, we can take this argument further. An intriguing possibility is that we can explain the development of a morphologically passive but functionally *active* form of *śak* by means of an attempted repair of the structure in (27). Particularly in Epic Sanskrit, what is formally the passive of *śak* can sometimes have active sense:[17]

(28) *na tu māṃ śakyase draṣṭum anena eva sva-cakṣuṣā*
not but I.ACC can.2SG see.INF this.INS EMPH own-sight.INS
'But you cannot see me with this sight of yours.' (*Mahābhārata* 6.33.8)

As a functionally active present stem, *śakyate* (or in this case, *śakyase*) would not be morphologically impossible in Sanskrit, since a few verbs do form functionally active present stems which are morphologically like a passive. But this is generally found with verbs which do not regularly form passives, so the ambiguity of active vs. passive *śakyate* is unusual, and in addition *śak* already has a regular active present stem, *śaknoti*. If active *śakyate* could be analysed as somehow derived from the passive *śakyate*, this would therefore be preferable to assuming an independently created present stem which unnecessarily introduces ambiguity into the paradigm.

We propose, therefore, that the active *śakyate* may derive from an attempt to construct the 'agreeing' passive type with intransitive infinitival predicates. The only way to repair the structure in (27) is to put the single argument in the nominative case, to provide a valid subject for the matrix verb. That is, the sentence in (25) would have to be reformulated as follows:

(29) *na tatra śakyate gantuṃ rāmaḥ*
not there can.PS.3SG go.INF R.NOM.SG
'Rāma cannot go there.'

But this is now superficially an active structure. Conceivably, a first attempt to parse (29) might try to force a passive interpretation on the infinitive, but this could only work with anaphoric control of the embedded OBL argument:

(30) Illicit structure:

$$\left[\begin{array}{ll} \text{PRED} & \textit{`can}\langle\text{XCOMP}\rangle\text{SUBJ'} \\ \text{NEG} & + \\ \text{SUBJ} & \left[\begin{array}{ll}\text{PRED} & \textit{`R\=ama'}\\ \text{CASE} & \text{NOM}\end{array}\right] \\ \text{XCOMP} & \left[\begin{array}{ll} \text{PRED} & \textit{`go}\langle\text{OBL}_\theta\rangle\text{'} \\ \text{VOICE} & \text{PASS} \\ \text{OBL}_\theta & \left[\begin{array}{ll}\text{PRED} & \textit{`pro'}\\ \text{CASE} & \textit{instr}\end{array}\right] \\ \text{ADJ} & \left\{\left[\text{PRED}\;\; \textit{`there'}\right]\right\}\end{array}\right]\end{array}\right]$$

[17] See Oberlies (2003a, 198), for whom this "looks like a passive used as active".

But evidence for anaphoric control by *śak* is otherwise lacking. By far the simpler way to interpret (29) is as a simple active structure, by making the assignation of passivity by *śakyate* optional:

(31)
$$\begin{bmatrix} \text{PRED} & \text{‘}\textit{can}\langle\text{XCOMP}\rangle\text{SUBJ’} \\ \text{NEG} & + \\ \text{SUBJ} & \boxed{1}\begin{bmatrix}\text{PRED} & \text{‘}\textit{Rāma}\text{’}\\ \text{CASE} & \text{NOM}\end{bmatrix} \\ \text{XCOMP} & \begin{bmatrix} \text{PRED} & \text{‘}\textit{go}\langle\text{SUBJ}\rangle\text{’} \\ \text{SUBJ} & \boxed{1} \\ \text{ADJ} & \left\{\begin{bmatrix}\text{PRED} & \text{‘}\textit{there}\text{’}\end{bmatrix}\right\} \end{bmatrix} \end{bmatrix}$$

To recapitulate our argument, then: given the analyses proposed above, with intransitive infinitives only the accusative type passive is possible, but with transitive verbs, it is the agreeing type which predominates. This predominance may have led to attempts to construct an agreeing type with intransitive infinitives, but given the case constraints, this could only be realised by effectively reinterpreting the passive *śakyate* as an active. To our knowledge there has been no better explanation proposed for the otherwise unexpected active *śakyate*.

3.4 The non-agreeing type

As discussed above, the non-agreeing type is found only with the gerundive, never with the finite passive.[18]

(32) *na tena śakyaṃ hantuṃ rāmaḥ*
not he.INS can.GDV.SG.NT slay.INF R.NOM
‘Rāma cannot be slain by him.’

As suggested above, the best way to analyse this is to take the matrix predicate here not as the nt.sg. of the gerundive but as an invariant, non-agreeing predicate. The analysis of this type will therefore be entirely parallel to the analysis of the agreeing type, the only exception being that there will be no direct agreement between the form of *śak* and its nominative subject argument.

3.5 The control equation

As discussed above, a simple subject control equation will not suffice to cover all the constructions discussed in this section. In particular, the violation of Visser’s Generalization requires us to license control by an OBL argument, but only in the absence of a SUBJ argument. The controlled argument is always a SUBJ, regardless

[18]In this section we are only considering *śak*. With other verbs such as *yuj*, the non-agreeing type is also found with the past participle. This fits with our proposed analysis, since it is in principle possible for the nt.sg. of past participles, just as of gerundives, to become non-agreeing predicates.

of the voice of the infinitive. We therefore reformulate the control equation with reference to argument structure positions, rather than grammatical functions:

(33) $(\uparrow_\sigma \text{ ARG}_1)_{\sigma^{-1}} = (\uparrow \text{ XCOMP SUBJ})$.

Following Kibort (2007), arg_1 (= s-structure ARG_1, following Findlay 2014, 2016) will be the subject in an active construction, but in the passive will be associated with OBL_θ; since *śak* is intransitive, when arg_1 is realized as OBL_θ, there will be no subject argument, thus capturing the generalization.

4 Conclusion

In this paper we have developed an LFG analysis of raising constructions in Sanskrit, with a particular focus on the verb *śak* 'can', and on interaction of raising with the passive. In passive raising constructions, passive morphology appears on the raising verb, while the form of the infinitive does not change, as there is no morphologically marked passive infinitive. From five superficially distinct passive types (the agreeing type, the accusative type, the non-agreeing type, ambiguous cases and constructions with intransitive infinitives), we distinguished two formal variants, distinguished by a single point of variation in the application of the passive feature.

In the first, the passive operation applies as expected to the argument structure of the raising verb, resulting in a subjectless construction with functional control by the matrix OBL_{theta} of the XCOMP SUBJ. This underlies the accusative type, and the construction with intransitive infinitives.

In the second, the passive operation applies rather to the argument structure of the infinitival predicate, despite being morphologically marked on the raising verb. This gives a standard subject to subject raising construction, but with passive interpretation of the infinitive, meaning that the logical object (the arg_2) of the infinitive can appear as the nominative subject of the matrix verb.

Our analysis provides further evidence against the universal status of the so-called 'Subject Condition'; it also supports the modification of Visser's Generalization proposed by van Urk (2013), and extends its applicability to raising verb. The latter point, which applies beyond Sanskrit, requires control equations to be stated not purely in terms of grammatical functions, as is standard in LFG, but at least partly in terms of argument structure positions.

References

Alsina, Alex, Mohanan, K. P. and Mohanan, Tara. 2005. How to get rid of the COMP. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG05 Conference*, Stanford, CA: CSLI Publications.

Arka, I Wayan and Simpson, Jane. 1998. Control and complex arguments in Balinese. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG98 Conference*, Stanford: CSLI Publications.

Berman, Judith. 1999. Does German satisfy the Subject Condition? In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG99 Conference*, Stanford, CA: CSLI Publications.

Bhatt, Rajesh, Bögel, Tina, Butt, Miriam, Hautli, Annette and Sulger, Sebastian. 2011. Urdu/Hindi modals. In *Proceedings of the LFG11 Conference*, pages 47–67, Stanford, CA: CSLI Publications.

Boeckx, Cedric, Horstein, Norbert and Nunes, Jairo. 2010. *Control as movement*. Cambridge: Cambridge University Press.

Bresnan, Joan. 1982. Control and complementation. *Linguistic Inquiry* 13(3), 343–434, reprinted in Joan Bresnan (ed.), *The mental representation of grammatical relations*, MIT Press, pp. 282–390.

Bresnan, Joan and Kanerva, Jonni M. 1989. Locative inversion in Chicheŵa: A case study of factorization in grammar. *Linguistic Inquiry* 20, 1–50.

Butt, Miriam. 2014. Control vs. complex predication: Identifying non-finite complements. *Natural Language & Linguistic Theory* 32(1), 165–190.

Butt, Miriam and Lahiri, Aditi. 2013. Diachronic pertinacity of light verbs. *Lingua* 135, 7–29.

Dalrymple, Mary and Lødrup, Helge. 2000. The grammatical functions of complement clauses. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG00 Conference*, Stanford, CA: CSLI Publications.

Dalrymple, Mary, Lowe, John J. and Mycock, Louise. 2019. *The Oxford reference guide to Lexical Functional Grammar*. Oxford: Oxford University Press.

Deshpande, Madhav M. 1980. *Evolution of syntactic theory in Sanskrit grammar: Syntax of the Sanskrit infinitive* -tumUN. Ann Arbor: Karoma.

Falk, Yehuda N. 2006. *Subjects and Universal Grammar: An explanatory theory*. Cambridge: Cambridge University Press.

Findlay, Jamie Y. 2014. The prepositional passive: A lexical functional account. MPhil. thesis, University of Oxford.

Findlay, Jamie Y. 2016. Mapping Theory without argument structure. *Journal of Language Modelling* 4(2), 293–338.

Gippert, Jost. 1995. Zur Syntax des Infinitivs auf *-tum* im Altindischen. In Wojciech Smoczyński (ed.), *Kuryłowicz Memorial Volume* (Vol. 1), pages 255–277, Cracow: Universitas.

Kibort, Anna. 2006. On three different types of subjectlessness and how to model them in LFG. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG06 Conference*, Stanford, CA: CSLI Publications.

Kibort, Anna. 2007. Extending the applicability of Lexical Mapping Theory. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG07 Conference*, pages 250–270, Stanford, CA: CSLI Publications.

Kiparsky, Paul. 2002. On the architecture of Pāṇini's grammar. Three lectures delivered at the Hyderabad Conference on the Architecture of Grammar, January 2002 and at UCLA, March 2002. Revised version 2008. http://www.stanford.edu/~kiparsky/ last accessed 01/01/2017.

Lødrup, Helge. 2014. Long passives in Norwegian: Evidence for complex predicates. *Nordic Journal of Linguistics* 37(3), 367–391.

Oberlies, Thomas. 2003a. Aśokan Prakrit and Pāli. In George Cardona and Dhanesh Jain (eds.), *The Indo-Aryan Languages*, pages 161–203, London: Routledge.

Oberlies, Thomas. 2003b. *A Grammar of Epic Sanskrit*. Berlin: de Gruyter.

Ørsnes, Bjarne. 2006. Creating raising verbs: An LFG-analysis of the complex passive in Danish. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG06 Conference*, pages 386–405, Stanford, CA: CSLI Publications.

Patejuk, Agnieszka and Przepiórkowski, Adam. 2020. Predicative adverbs: Evidence from Polish. *Linguistic Inquiry* pages 1–17, early Access Corrected Proof.

Speyer, Jacob Samuel. 1896. *Vedische und Sanskrit-Syntax*. Strassburg: K. J. Trübner.

Stiebels, Barbara. 2007. Towards a typology of complement control. In Barbara Stiebels (ed.), *Studies in complement control*, pages 1–80, Berlin: Universitätsbibliothek Johann Christian Senckenberg.

van Urk, Coppe. 2013. Visser's Generalization: The syntax of control and the passive. *Linguistic Inquiry* 44(1), 168–178.

Whitney, William Dwight. 1896. *A Sanskrit Grammar: Including both the Classical language, and the older dialects, of Veda and Brahmana*. Leipzig: Breitkopf & Härtel, third edition.

Adjuncts at the syntax-prosody interface in nominal structures in Dela

Thersia M. Tamelan
Artha Wacana Christian University

I Wayan Arka
Australian National University/Universitas Udayana

Proceedings of the LFG'21 Conference

On-Line

Miriam Butt, Jamie Y. Findlay, Ida Toivonen (Editors)

2021

CSLI Publications

pages 264–284

http://csli-publications.stanford.edu/LFG/2021

Keywords: nominal alternation, NP head-adjunct relation, light/heavy syllable, syntax-prosody interface, modular model.

Tamelan, Thersia M., & Arka, I Wayan. 2021. Adjuncts at the syntax-prosody interface in nominal structures in Dela. In Butt, Miriam, Findlay, Jamie Y., & Toivonen, Ida (Eds.), *Proceedings of the LFG'21 Conference, On-Line*, 264–284. Stanford, CA: CSLI Publications.

Abstract

This paper discusses an intriguing syntax-prosody interface phenomenon in Dela, an Austronesian language spoken on Rote Island, eastern Indonesia. Typologically, Dela is an SVO language where the NP's head-adjunct relation is marked by a light/heavy syllable alternation resulting in C(onsonant) deletion/insertion. Our study contributes to the typological and theoretical research on the nature and function of prosody in grammar. We demonstrate that LFG's modular model nicely captures the syntax-prosody phenomenon in Dela.

1 Introduction

This paper describes the NP head-adjunct relation at the syntax-phonology interface in Dela.[1] An adjunct in this language is marked by a light/heavy syllable alternation, as illustrated by the examples in (1). The consonant-final word *anin* 'wind' retains its final consonant in (1a) because it is not followed by an adjunct.[2] However, in (1b), it is followed by an adjunct (*ɓarat* 'west') and the word-final consonant of *anin* is elided. The final word of the NP in Dela receives (phrasal) prosodic prominence.[3] Thus, the NP [*a<u>ni</u> ɓarat*] in (1b) has its phrasal prosodic peak, represented by H* (of the melodic H*L), associated with the word at the right boundary (RB) of the intonational unit/prosodic phrase. The prosodic peak typically coincides with the penultimate syllable, which is also the bearer of primary stress in a word.

(1) a. The underlying word-final C is retained

[[***a**<u>nin</u>*]$_{NP}$=*a*]$_{DP}$	*tao*	*maŋgarauʔ=ra*	*ra-ta-mbele*
wind=DEF	make	rubbish=PL	3PL-VBLZ-fly

'*The wind* blew the rubbish.' [YN2.23][4]

[1] We thank the Dela speakers who provided us with data for this paper, the anonymous reviewers of the abstract, and the LFG 2021 audience members, in particular, Louise Mycock, Chen Xie, and Stephen Wechsler, for their detailed and valuable comments which helped improve this paper significantly. We also thank Charbel El-Khaissi for proofreading our paper. All remaining errors and inconsistencies are our own.

[2] Abbreviations are as follows: 1, 2, 3=first, second, third person; VBLZ=verbaliser; DEF=definite; DEM=demonstrative; DISC=discourse; DIST=distal; GEN=genitive; NOM=nominative; NMLZ= nominaliser; Nʔ=nominal ʔ; ORD=ordinal; PL=plural; PROX=proximal; REL=relativiser; SG=singular; STAT=stative.

[3] The syllable that shows a heavy/light alternation in adjunct relation in the relevant NP under discussion is underlined. The syllable that bears the word stress and/or phrasal prosodic peak/prominence is in bold.

[4] This code in square brackets [ ] following the free translation is the citation for the recorded texts indicating the source of the example in the Dela corpus.

b. The underlying word-final C is elided

kalo	*fula*	*ka-sanahulu-n*	*na*	[*a<u>ni</u>*	***ɓa**rat*]$_{NP}$
					H*L
if	month	ORD-ten-ORD	DISC	wind	west

'If it's October, (it's) *west wind*.' [NA1.30]

Based on the data for the NP head-adjunct relation, we argue for two key empirical points:

1) Syllable marking: a noun's word-final syllable alternation (light vs. heavy) encodes the presence/absence of an adjunct close to the head of the NP.
2) Unit alignment: the syntactic unit of NP is aligned with the prosodic unit of Phonological Phrase (PhP), characterised by the melodic tone/prosodic peak of H*L at the right edge of the NP/PhP.

Our analysis builds on previous LFG approaches to the prosody-syntax interface. In this approach, the prosodic structure is a layer with its own properties and distinct from other layers such as c- and f-structures (Butt and King 1998, Dalrymple and Mycock 2011, Mycock and Lowe 2013, Dalrymple, Lowe, and Mycock 2019, Bögel 2015). The analysis consists of the lexical entry specification with prosodic information and the proper grammatical function (GF) annotation to account for the syntax-prosody interface.

This paper is structured as follows: in §2, we provide an overview of Dela phonology and morphosyntax. Then, we describe the NP head-adjunct relations in §3, followed by an analysis of these relations within the LFG framework in §4. We contextualise this phenomenon in Dela by providing an areal comparison with neighbouring languages that exhibit a similar pattern (§5). Finally, in §6, we conclude with the implications of this study and questions for future investigation.

2 An overview of Dela phonology and morphosyntax

A number of phonological and morphosyntactic structures of Dela that are related to the topic of this paper are presented here. These include word stress, syllable structure, basic clause structure, nominal structure and morphology of Dela nouns. The description is mainly based on the work of Tamelan (2021).

2.1 *Phonology*

In Dela, word stress falls on the penultimate syllable of the root and each vowel is a syllable nucleus. The root or stem is typically disyllabic, forming a trochee (metrical foot consisting of a stressed syllable followed by an unstressed syllable). This is illustrated in words with the root ***ma**ta* 'eye' and stem ***na**a* '3SG.eat' in (2) with stress on the first (boldened) syllables. Prefixes, genitive and plural enclitics are extrametrical—they do not influence word stress, or in other words, stress does not shift. Notice that when the genitive enclitic *=na* attaches to a word, the nominal suffix *-ʔ* is omitted (i.e. *mata=na* is underlyingly *mata-ʔ=na*).[5] The plural enclitic *=ra* can, however, co-occur with the nominal suffix.

(2) Stress pattern for disyllabic roots

```
              x
    (x)       x   x         (x)     (x)
 (C)(V-)(C)V(C)V(C)(-C)(V)(-C)(V)
          m a t a   -ʔ                      'eye-Nʔ (unassociated)'
          m a t a   =n  a                   'eye=3SG.GEN'
          m a t a   =n  a   =r  a           'eye=3SG.GEN=PL'
          m a t a   -ʔ      =r  a           'eye-Nʔ=PL'
          n-a   a                           '3SG-eat'
  n  a-   n-a   a   -t                      'food'
  n  a-   n-a   a   =n  a   =r  a           'food=3SG.GEN=PL'
```

* (X): the syllables are extrametrical

Double vowels are sequences of two (like or unlike) vowels with each vowel being the head of a syllable nucleus. This is shown by the fact that the stress remains on the penultimate syllable (i.e. vowel) of the root, as in (3), for both like vowels, which results in a phonetically long vowel (e.g. *fuu*), and unlike vowels (e.g. *fui*). In reduplicated words with like vowels, as in (4), the double vowels are not reduplicated as a unit. Instead, the syllable that is copied and prefixed to the stem is a short vowel (e.g. *fu-* in *fu~fuu)*.

(3) Stress remains in the penultimate syllable

fui [ˈ**fu**ɪ] 'wild' *fuu* [ˈ**fu**ː] 'blow'
lilii? [liˈ**li**ːʔ] 'forget' *la~lai* [laˈ**la**ɪ] 'slicing'

(4) CV reduplication does not reduplicate two like vowels as a unit

fuu [ˈ**fu**ː] 'blow' → *fu~fuu* [fuˈ**fu**ː] 'blowing'
laa [ˈ**la**ː] 'float' → *la~laa* [laˈ**la**ː] 'floating'

In Dela, the medial C-slots in monomorphemic roots are ambisyllabic—they occur as both the coda of the syllable to their left and the onset of the syllable to their right. A template for disyllabic root structure is given in (5). The evidence for this analysis is shown in reduplication such as *mali* 'laugh' → *ma**l**~mali* 'laugh intensively' and *esa* 'one' → *e**s**~esa* 'each one'. In these examples, the root medial consonants are /l/ and /s/ respectively, and they serve as the coda of the syllable to their left and the onset of the syllable to their right. Hence, they are included in the reduplicants.

---

[5] In Dela, genitive enclitics usually replace any final consonant of the root they attach to except when the omission causes lexical ambiguity (Tamelan 2021:52).

(5) Template for disyllabic root structure

2.2 *Morphosyntax*

Dela has a basic SVO order, as shown in (6) and (7). It has a nominative-accusative alignment system where S and A are treated the same, as opposed to O (or patient), as seen in pronoun sets.[6] See Tamelan (2021:97–106) for a more detailed discussion on Dela's pronominal system.

(6) [*ana*]$_{\text{SUBJ}}$ [*n-ita*]$_{\text{V}}$ [*e*]$_{\text{OBJ}}$
3SG.NOM 3SG-see 3SG.ACC
'She saw him.' [YB6.57]

(7) [*ana*]$_{\text{SUBJ}}$ [*lao*]$_{\text{V}}$
3SG.NOM leave
'She left…' [YNHN1.9]

Dela is predominantly left-headed and, as expected, shows post-head modifiers despite being a predominantly prepositional language, as in (8). The maximal nominal unit with D is DP. Like several other languages in the region, Dela does not have a separate grammatical class of adjectives. Words expressing property concepts such as 'old' behave as either a noun (obligatorily appearing with the nominal suffix -*ʔ*), as in (8), or a verb (obligatorily taking the subject and stative prefixes), as in (9).

(8) *tou* *lasi-ʔ* *naa*
male old-Nʔ DIST
'that old man'

(9) *ana* *na-ma-lasi*
3SG.NOM 3SG-STAT-old
'He is old'/ 'He becomes old'

Nouns in Dela can be morphosyntactically derived or free (i.e. non-derived), as in (10) and (11), and they have either a root-final light syllable ('light noun') or root-final heavy syllable ('heavy noun'). Free nouns are typically light nouns (i.e. with a final syllable without a C coda), as in (10a). Heavy non-derived nouns such as *anin* 'wind' in (10b) are rare. Heavy syllables are typically syllables with

[6] Whether a verb takes a subject prefix or not is lexically determined. The prefixes consist of two paradigms: syllabic and non-syllabic prefixes. The non-syllabic prefixes consist of the initial consonant of the syllabic prefixes (Tamelan 2021:130).

a consonant coda. Note that a syllable with a phonetically long vowel is analysed as a sequence of two vowels, with each vowel being the head of a syllable nucleus (Tamelan 2021:23). Syllables with diphthongs are only found in loan words. Underlyingly or by default, all derived nouns are, by definition, heavy nouns. They are derived from precategorial roots or subcategorised roots.[7]

(10) Morphosyntactically free nouns:

a. 'Light' noun: having a light root-final syllable (i.e. without a coda)
e.g. *o<u>e</u>* 'water', *mo<u>ko</u>* 'big'.

b. 'Heavy' noun: having a heavy root-final syllable
e.g. *a<u>nin</u>* 'wind'

(11) Derived nouns (typically from a light root-final syllable) primarily by means of nominaliser *-ʔ, -s* or *–t*

a. From 'precategorial' roots*;* e.g. *la<u>si-ʔ</u>* 'old', *he<u>di</u>*<u>-s</u> 'illness'

b. From subcategorised roots;
e.g. *ŋgo<u>e</u>* 'to lock' → *ŋgo<u>e-ʔ</u>* 'a key',
n-a<u>a</u> '3SG eat' → *na~n-a<u>a-t</u>* 'food'
o<u>e</u> 'water' → *o<u>e-ʔ</u>* 'liquid'

Some free nouns with a root-final light syllable such as *oe* 'water' and *moko* 'big' may behave differently in terms of nominal marking when they appear attributively. Both *oe* 'water' and *moko* 'big' can function as head nouns like in (12a) and (13a). The noun *oe* 'water', however, requires the general/default nominal marking *-ʔ* when it appears as an adjunct, appearing as *oeʔ* in (12b). It appears that the nominaliser *-ʔ* functionally turns 'entity' nouns with referential meaning to a 'property/quality' noun. By contrast, the noun *moko* requires no *-ʔ* marking when it functions attributively as in (13b), apparently due to existing 'property/quality' signification in the noun. That is, it is a property (or 'adjectival') noun, and it does not need a nominaliser to function attributively in Dela.

(12) a.

ara	*nasu*	[***oe***	*hanas*]$_{NP}$
3PL.NOM	boil	water	hot

'They boiled hot *water*.' [LK2.28]

b.

ina-ʔ=ra	*mana*	*lemba*	[*tasi*	***oe-ʔ***]
female-Nʔ=PL	REL	carry.with.pole	sea	water-Nʔ

'The women are the ones who carried sea *water*.' [HL4.7]

[7] Precategorial roots are those with no clear evidence that one derived form is more basic than the other/s, while subcategorised roots are those that clearly belong to one particular morphological or syntactic category (Tamelan 2021:67).

(13) a.

au	*teŋga*	*ʔ-ala*	[***moko***=*ra*]
1SG	pick	1SG-get	big=PL

‘I picked the *big* ones.’ [YNML6.10]

b.

ana	*hasa*	*n-ala*	[*ɓaruu*	***moko***]
3SG.NOM	buy	3SG-get	pants	big

‘He bought *big* pants.’ [Elicited]

3 NP head-adjunct relation

There are two patterns of heavy/light syllable alternation involved in an NP-adjunct relation. The first pattern is the heavy-to-light alternation which happens to the noun head or a noun adjunct preceding another adjunct. This alternation is regular and applicable (or imposed) across Dela nouns (cf. the second pattern discussed below). Further, the alternation is syntactically motivated by the presence of an adjunct that follows a noun with a lexically specified word-final heavy syllable. The presence of an adjunct is marked by a word-final light syllable as seen in the deletion of a word-final consonant in a noun. Hence, *anin* ‘wind’ becomes *ani* as in the previous examples (1a, b), repeated below in (14a, b). The noun retains its final consonant in (14a) because there is no adjunct following it, but it lacks a coda in (14b) to mark the presence of the following adjunct, *ɓarat* ‘west’. Notice that in (14c), the noun adjunct *ɓarat* ‘west’ also lacks a coda to mark the presence of the second adjunct, *monaeʔ* ‘big’.

(14) a. The underlying word-final C is retained

[[***anin***]$_{NP}$=*a*]$_{DP}$	*tao*	*maŋgarauʔ=ra*	*ra-ta-mbele*
wind=DEF	make	rubbish=PL	3PL-VBLZ-fly

‘*The wind* blew the rubbish.’ [YN2.23]

b. The underlying word-final C is elided

kalo	*fula*	*ka-sanahulu-n*	*na*	[*ani*	***ɓa**rat*]$_{NP}$
if	month	ORD-ten-ORD	DISC	wind	west

‘If it’s October, (it’s) *west wind*.’ [NA1.30]

c.

ʔoi	*fula*	*ka-esa-ʔ*	*ia*	[*ani*	*ɓara*	***mo**nae-ʔ*]$_{NP}$
say	month	ORD-one-Nʔ	PROX	wind	west	big-Nʔ

‘(It is) reported that in this January, (there will be) *strong west wind*.’ [YB8.28]

There is no heavy-to-light syllable alternation when marking an adjunct relation if the final syllable of the preceding noun is already lexically light. This is because the requirement of a word-final light syllable is respected (see example [15a] below). However, a syllable alternation pattern still exists with these noun types despite being of a different kind to the first pattern in (14). This second pattern is a light-to-heavy syllable alternation. This marking takes place when an adjunct relation is associated with certain nouns containing a word-final light syllable, such as *oe* ‘water’, as seen in (15). Unlike the highly regular pattern exemplified in (14), a light-to-heavy syllable alternation is semantically constrained and only applicable to certain adjunct nouns. The semantic constraint appears to be associated with the typical nature of an adjunct as a ‘property/quality’ word. That is, since light nouns like *oe̲* ‘water’ are entity nouns, they appear in their basic form when heading an NP with or without an adjunct (15a). When an entity noun carries a modifying function, as in (15b), it has to be turned into ‘property’ noun subclass by means of the suffix -*ʔ̲*: *oe̲-ʔ̲*. Hence, the formation of *oe̲-ʔ̲* is derivational at the level of morphology and provides the syntactic level with the right semantic (sub-)type in order to appear in an attributive slot.

(15) a.

ara	*nasu*	[*oe̲*	(***ha**nas)*]NP	
3PL.NOM	boil	water	hot	

‘They boiled (*hot) water*.’ [LK2.28]

b.

ina-ʔ=ra	*mana*	*lemba*	[*tas̲i̲*	***o**e̲-ʔ̲*]NP
female-Nʔ=PL	REL	carry.with.pole	sea	water-Nʔ

‘The women are the ones who carried *seawater*.’ [HL4.7]

c.

ia	[*tua*	*oe̲*	*ma-**ʔe**i-ʔ*]NP
PROX	palm	water	STAT-sour-Nʔ

‘This is *sour palm juice*.’ [Elicited]

The light adjunct noun in (15c) deserves additional commentary. Unlike (15b), the noun *oe* ‘water’ in (15c) differs in its final light syllable. This is because it is followed by another adjunct, *maʔeiʔ* ‘sour’. The surface form *oe̲* in (15c) is underlyingly *oe̲-ʔ̲*, which is identical to the derived form in (15b). That is, it is a property noun, but its final -*ʔ* has been elided to satisfy the requirement of phonological (light-syllable) marking in nouns followed by an adjunct.

Data points in (16)-(17) provide further evidence that the alternation to a light syllable is associated with adjunct marking. The quantifiers *naeʔ* ‘much’ and numeral *rua* ‘two’ in (16) do not trigger such an alternation. The final C of *anin* cannot be elided in (16a) because the following segment is not an adjunct. Likewise, the numeral *rua* does not trigger the C deletion in (16b): the V-final noun *tou* ‘male’ obligatorily appears with a heavy syllable (i.e. final glottal C).

(16) a.

afis=a	*ani̲n̲/*a̲n̲i̲*	*naeʔ.*
yesterday	wind	much

‘Yesterday, it was *windy*.’ [YN4.16]

b.

*tou̲-ʔ̲/*tou*	*rua*	*kerja*	*sa*	*ʔofaʔ=a*
male-Nʔ	two	work	LOC.IPFV	boat=DEF

‘Two men work in the boat.’ [CH1.4]

In addition to final -*ʔ*, there are other nominalising suffixes in Dela whose distribution is lexically determined. Certain nouns, such as ‘food’ in (17), are derived by means of -*t*. Given it is a consonantal suffix, the derivation unsurprisingly gives rise to a light/heavy syllable alternation that is subject to the constraints discussed so far. Thus, in examples (17b-c), the word-final C nominaliser *-t* (needed by the morphology for lexical derivation) is retained because there is no adjunct following the derived noun. The clitic pluraliser, *=(a)ra,* that follows it does not trigger light syllable marking, as shown in (17c).

(17) a.	*n-aa*	b. *na~n-aa-t*	c. *na~n-aa-t=ara*
	3SG-eat	RDP~3SG-eat-NMLZ	RDP~3SG-eat-NMLZ=PL
	'She eats'	'(different kinds of)food'	'the different kinds of food'

As expected, the word's light/heavy final syllable, together with the NP's prosody, disambiguates syntax. For example, it differentiates equative clauses or possessive NPs from attributive phrases (Tamelan 2021:243). As seen in their translation differences, (18a) is an NP showing an internal head-adjunct relation, whereas (18b) is a nominal clause consisting of two NPs. Similarly, (19a) shows an NP with an internal head-adjunct relation while (19b) is a genitive construction.

(18) a. Attributive relation within a single NP

na~n-aa	*ma-laɗa-ʔ*
RDP~3SG-eat	STAT-tasty-Nʔ

'tasty food'

b. Equative clausal relation involving two NPs

na~n-aa-t	*ma-laɗa-ʔ*
RDP~3SG-eat-NMLZ	STAT-tasty-Nʔ

'The food is tasty.'

(19) a. Attributive relation

mbela	*ɗeke-ʔ*
corn	seed-Nʔ

'corn seed(s)' (i.e. corn seeds that are no longer attached to the cob)

b. Genitive construction

mbelaʔ	*ɗeke=n*
corn	seed=3SG.GEN
POSSR$_{NP}$	N$_{POSSED}$

'The seeds of a corn' (i.e. seeds that are part of a corn plant)

Since a relative clause (RC) is syntactically an adjunct, it also triggers the light syllable alternation in Dela, as seen in (20). The first noun, *kokis,* appears in its original form and by itself as an NP with a heavy final syllable. However, its second occurrence appears with the word-final light syllable, *koki*, because a relative clause adjunct follows it. Note that the RC's final word, *tunu-ʔ,* also has a heavy syllable, but its appearance is derivational and semantically motivated (i.e. light-to-heavy syllable alternation as discussed earlier). That is, the noun *tunu-ʔ* is a property noun as depicted by the free translation, and it is functionally the predicate of the RC.

(20) [ko<u>kis</u>]NP, [ko<u>ki</u>]N [*mana* *tu<u>nu-ʔ</u>*]RC
cake cake REL roast-Nʔ
'Cake, cake [which are *baked*].' [YNHN1.25]

However, a complication arises when another adjunct exists in an NP string with an RC. Dela deals with this issue by splitting the NP into two prosodic phrases and postposing the RC's heavy unit outside the NP. Consider the pair in (21). The NP (21a) is pronounced as a single prosodic phrase with both the noun head (*tou*) and the first adjunct (*lasi*) appearing with light syllables. This light syllable marking provides support that the two adjuncts belong to the same/single NP. Furthermore, the prosodic property provides empirical support for clear correspondence of units in the phonology-syntax interface (i.e. NP=PhP). In contrast, (21b) consists of two prosodic phrases. A heavy syllable marks the first one at its right edge (*la<u>siʔ</u>*); that is, no deletion of the syllable coda-nominaliser *-<u>ʔ</u>* occurs. Syntactically, this heavy syllable marks the following RC as a non-adjunct in the same NP unit. The noun head *tou* remains with a word-final light syllable because it is the head noun followed by an adjunct. The retention of the suffix *-<u>ʔ</u>* in the adjunct *la<u>si-ʔ</u>* (together with the H*L prosody) marks the absence of the following adjunct and the right boundary of NP. In other words, the RC is not a syntactic adjunct within the NP, and (21b) therefore consists of two NPs, with the RC being a headless appositive RC (captured by the PS rule; see [23c.i]).

(21) a. [*to<u>u</u>*]N [*la<u>si</u>*]N [*mana* *ŋgo~ŋgo<u>o-ʔ</u>*]RC =*a*
male old REL RDP~senile-Nʔ =DEF
'The senile old man,' [YB9.92E]

b. [*to<u>u</u>* *la<u>si-ʔ</u>*]NP, [*mana* *seo* *ʔuʔu*]RC =*a*, *ana* *mate* *ena*
male old-Nʔ REL sell fish =DEF 3SG.NOM die PFV
'The old man who sells fish, he has passed away.'

4 LFG analysis

Our LFG analysis consists of three components: (i) lexical entry specification, (ii) phrase structure and prosodic structure rules regulating syntactic (s-) string and prosodic (p-) string respectively, and (iii) the alignment mechanism of s- and p-strings. Each of these components of analysis is discussed in order.

4.1 *Lexical entry*

Building on previous works in LFG (Dalrymple and Mycock 2011, Mycock and Lowe 2013, Dalrymple, Lowe, and Mycock 2019, Bögel 2015), we include *p(rosodic) form* information relevant in the lexical entry, in addition to the *s(yntactic)-form* information relevant for a syntactic string (c-structure). However, our simplified approach does not strictly follow theirs, and points of difference will be briefly outlined as necessary. For example, the (simplified) lexical entries for *anin* 'wind' and *oe* 'water' are given in (22i-ii), representing the pairing of FORM and MEANING. The form side consists of a string of segments, which are organised (and labelled) differently.[8] Its s-form (22.i.a) says that it is a morphosyntactic word, precisely a noun (N) root. This grammatical information is relevant for morphosyntactic string manipulation both within morphology (e.g., word formation) and in the morphology-syntax interface. Its p-form (22.i.a) says that it is also a phonological word (PW), with syllable properties (in this case, two syllables with syllable boundaries indicated by a dot [.]). This information is relevant for p-string manipulation. Both properties are essential in the lexical phonology-morphology interface and the post-lexical prosody-syntax interface when accounting for the word-final C deletion and insertion (or retention) in the head-adjunct nominal structure in Dela.[9]

(22) (i) FORM: anin oe

a. s-form: $[\text{anin}]_{\text{N.ROOT}}$ $[\text{oe}]_{\text{N.ROOT}}$

b. p-form: $[\text{a.nin}]_{\text{PW}}$ $[\text{o.e}]_{\text{PW}}$

(ii) MEANING (f-info): (↑PRED)= 'wind' (↑PRED)= 'water'

(iii) s-string:

$$\begin{bmatrix} \text{FM} & [\text{anin}] \\ \text{L} & \{\text{N.ROOT}\} \\ \text{R} & \{\text{N.ROOT}\} \end{bmatrix}$$

p-string:

$$\begin{bmatrix} \text{FM} & [[\text{a}]_{\text{L}.\sigma}\ [\text{nin}]_{\text{R}.\sigma}] \\ \text{STRESS} & \text{L}\sigma \\ \text{L} & \{\text{PW, F}\} \\ \text{R} & \{\text{PW, F}\} \end{bmatrix}$$

The same information can be alternatively represented in an attribute-value matrix (AVM) as in (22.iii). The advantage of an AVM representation is that it explicitly captures the left (L)/right (R) element in the relevant hierarchical s- and p-strings. For example, at the most basic level of the morphosyntactic string

---

[8] The segments are phonological in spoken language or graphical in written language. The dot in (22b) indicates the syllable boundary.

[9] The specification of prosodic information in the lexical entry as seen in (22.i.b) highlights the difference between our approach and the approach adopted by Dalrymple, Lowe and Mycock. In our analysis, the status of PW is not wholly inherited from the p-structure after the word is inserted into it. This is just like the availability of the categorial information of N in s-form and c-structure, which, for example, allows for a proper lexical item's insertion to c-structure. Thus, relevant PW information (e.g., syllabification or stress) is available at the levels of the lexical entry and p-structure. We assume a hierarchical p-structure as captured by the p-structure rule in (24) (cf. Selkirk 1986). The p-information coming from the lexical entry interacts in a dynamic way with the prosodic information from other PWs in the p-structure, and is also subject to phonology-syntax interface constraints, which is captured by the rules in (25). This will result in the final prosodic (PW/PhP) properties (e.g., whether the PW also carries the phrasal prosodic peak, as in the rule 25.c.ii, and as further discussed in section 4.3). The specification of PW information in the lexical entry is also motivated by the fact that words must have their proper prosodic properties even when they are pronounced in isolation (i.e., without a larger context of p-structure or c-structure).

registered in the lexical entry, the form (FM) *anin* is a N root. Hence, its L and R value is 'N.Root'. Its corresponding p-string at this basic level is a prosodic word (PW) that is also a foot (F) consisting of two syllables, with the stress falling to its (L) syllable.[10] We demonstrate the significance of this explicit information in sections 4.3-4.4.

4.2 *Phrase structure and prosodic structure rules*

Phrase structure (PS) rules that capture Dela's internal nominal structure are given in (23). The nominal is analysed as a DP (23a), which can have a quantifier phrase (QP) before or after the NP as in (23b). Crucially, it can have multiple adjunct XPs (where XP={PP|NP|VP}) in one of two positions: outside the NP and structurally adjoined to the NP (23c.i) or within the NP and immediately following the head noun (23c.ii). The two adjuncts are called 'NP-external' and 'NP-internal' adjuncts, respectively. The NP-external adjunct is the position of the appositive RC, as in example (21b).

(23) a. DP → QP D

b. QP → Q , NP

c. $$\text{NP} \rightarrow \left\{ \begin{array}{ll} \text{i. NP} & \underset{\downarrow\in(\uparrow\text{ADJUNCT})}{\text{XP*}} \\ \text{ii. N} & \underset{\downarrow\in(\uparrow\text{ADJUNCT})}{\text{XP*}} \end{array} \right\}$$

The important point to note is that the prosodic word-final marking involved in the C-deletion/retention alternation only applies to the NP-internal adjunct relation domain, and is captured in (23c.ii) (cf. [15]). To capture the word-final C deletion/retention involved in NP-adjunct marking at the (morpho)syntax-phonology interface, we also need prosodic phonological rules, given in (24).

Recall in section 3 that the word-final C-deletion alternation applies to the N head in the presence of a following adjunct. It also applies to a non-final adjunct in NPs with multiple adjuncts, which is captured by the notation XP* in the rule (23c.ii) above. The addition and retention of a word-final consonant nominaliser (e.g. -ʔ in example 15b) applies to the rightmost adjunct, or the right edge of the NP (23c.ii). This is a complex outcome of a constraint at the morphology-phonology-syntax interface: the suffixation of a stem that results in a property N (i.e. morphology) is structurally required in an attributive position within the NP

10 Note that we introduce the feature STRESS in our analysis. This information is handled by a TONE feature and a separate SYLLSTRESS, as in Dalrymple et al. (2019). We opt for a simple approach to stress, which is a trochee in Dela. Hence, its value is Lσ (i.e., a metrical foot consisting of a stressed syllable, L, followed by an unstressed syllable, R).

(i.e. syntax). Crucially, in the phonology-syntax interface, the C-coda/suffix is retained because it is part of the N rightmost adjunct, in which case the light syllable marking does not apply.

The simplified and informal prosodic phonological rules in (24) regulate p-string from the internal structure of a phonological word (PW) to the higher units of a phonological phrase (PhP) and intonational phrase (IntP).[11] The notations of IntP$^+$ and PW$^+$ mean that IntP$^+$ and PW$^+$ consist of at least one PhP and one PW, respectively.

(24) a. IntP → PhP$^+$

(RB_TONE=[**H***L]$_F$)

b. PhP → PW$^+$

c. PW → (σ) [$\boldsymbol{\sigma_L}$ $\boldsymbol{\sigma_R}$]$_F$ (σ)

(L=**H**)(R=L)

Crucially in the PhP, the rightmost PW word must carry the PhP's prosodic peak of H*L tone melody (with H* marking the prosodic peak, or nuclear tone). This right boundary tone melody (RB_TONE for short) is informally represented as having the value [**H***L]$_F$ in (24a). The notation [**H***L]$_F$ refers to the prosodic property of PW and is similar to the one shown in (24c) where its prosodic prominence (i.e., stress) falls on the left syllable of the foot (F), which is also the morphological root (cf. section 2.1). The only difference between (24a) and (24c) is that (24a) specifies that the syllable stress is the most prominent syllable (i.e. the peak) at the level of PhP. This rule captures the empirical point that PhP in Dela is characterised by this salient prosodic melodic feature of the rightmost PW in PhP, and that the syllable carrying H* is also the one associated with the PW's stress.

4.3 *Aligning c-structure and p-structure*

The basic idea of the c-string and p-string alignment in Dela NP-internal adjunct marking is to formulate a mechanism to capture two properties: (i) unit-alignment of NP (syntax) with PhP (prosody), and (ii) prosodic/phonological marking that gives rise to the heavy/light alternations discussed in section 3. These properties can be informally schematised in (25a, b), and are equivalent to Selkirk's Match Theory (Selkirk 2011; Bögel forthcoming). The s-string (25a) is the syntactic adjunct-head domain and applies the associated phonological

[11] We do not discuss this complexity and its representation in this paper. Nonetheless, it will suffice to say that a more precise representation will need relevant arrows, and possibly more rules, in line with the ideas discussed by Dalrymple et al. (2019:422). This would ensure that the prosodic melody at the right edge is [H*L]$_F$, and that the syllable being H* is also the one that is lexically assigned stress.

marking constraints/properties in (25b). The prosodic constraints/markings can be represented as conditional 'if-then' prosodic rules as shown in (25c), and incorporate L (left) and R (right) edge features (Dalrymple and Mycock 2011; Mycock and Lowe 2013; Dalrymple et al. 2019).

(25) The prosodic adjunct marking constraints of NP in Dela:

a. s-string: [[$]N^{+}$ [NP] $]_{\mathbf{NP}}$

HEAD (ADJUNCT)

b. p-string [[σ **light.$\boldsymbol{\sigma}$]**$_{\mathrm{PW}}$ [PW]* $]_{\mathbf{PhP}}$

(RB_TONE=[**H***L]$_{\mathrm{F}}$)

c. **Prosodic NP ADJ marking constraints:**
   (i) **If** {NP, PhP} is associated with an ADJUNCT,
   then it must come with a daughter's node with R values of
   {**N, PW, light.syll**}
   (ii) If {NP, PhP}, then it must come with a daughter's node
   with R values of {**N, PW, [H*L]$_{\mathbf{F}}$**}

The prosodic rule in (25c.i) is the light syllable constraint marking in the phonology-morphosyntax interface. It is the NP-internal adjunct rule that results in the final light syllable of the head N—this constraint dictates that if there is an adjunct in the NP, the preceding N must be a light N (i.e. with a word-final light syllable). The prosodic rule in (25c.ii) specifies the Right Boundary H*L melodic tones, which dictates that the right edge of an NP is also the right edge of the PhP with the prosodic word carrying a prosodic contour of H*L. Note that while this marking is not exclusive to adjunct nouns, the specification in (25c) is intended for marking the prosodic NP adjunct noun—hence we include the categorial information of N in the set values in (25c.ii).[12]

The proposed constraints in (25c) regulate the possibilities of prosodic chunking that are needed to ease cognitive processing in long and complex NPs with multiple adjuncts (recall example [21]). The constraints correctly capture the interconnection between p-structure (i.e. Phonological/prosodic phrase) and c-structure (i.e. NP). This interconnection is regulated by the Interface Harmony principle (Dalrymple, Lowe, and Mycock 2019, Dalrymple and Mycock 2011, Mycock and Lowe 2013).

---

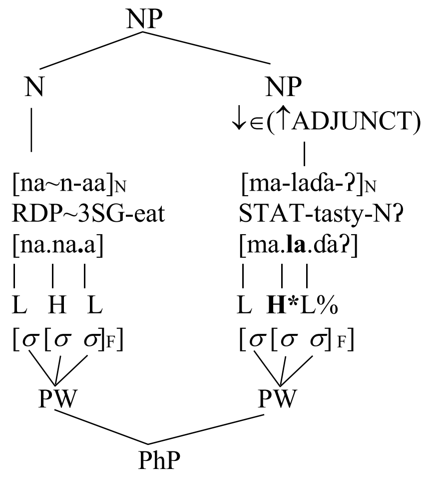
[12] Alternatively, the rule (25.c.ii) can be formulated as the default or elsewhere rule in which case the categorial N(P) information needs to be replaced by XP as it does not exclusively apply to NP.

4.4 *Demonstration of the analysis*

Now that the relevant properties have been outlined, we are in a position to demonstrate our analysis and account for typologically unusual cases at the (morpho)syntax-phonology interface in Dela. In particular, we consider how the syntactic marking of NP adjuncts accesses lexical phonology via the PW's internal structure. Consideration is given to two types of morphemic material involved in the removal of the consonant coda of word-final syllables: morphemic material (e.g. the nominaliser =*ʔ*) and non-morphemic segmental material (e.g. *anin* →*ani* 'wind'). We also demonstrate how syntax and post-lexical phonology interact with one another via the PhP's right edge and H*L prosodic contour (25c.ii). We illustrate these facts with reference to the examples in (18a, b) since they provide crucial evidence on how different prosodic properties disambiguate syntax.

The s-string in (18a) is interpreted as an NP with a c- and p-structure analysis represented in (26) below. This example involves a single NP s-constituent whose top NP node corresponds to a single prosodic phrase (PhP). The node of the NP's/PhP's right daughter is occupied by $[\text{ma.}\mathbf{la}\text{.ɗaʔ}]_{N=PW}$. Its internal unit is comprised of [F(oot)=Root] and its prosody is characterised by the most prominent L syllable at the PW and PhP levels. Put differently, the R values of the AVMs, as shown in (26b.ii), correspond to s- and p-string units and contain the set {NP, PhP, N, PW, $[1][\text{H*L}]_F$}, which satisfies the constraint in (25c.i). Tag [1] in (26b.ii) means that the same melodic tone value of [H*L] at the foot level also marks RB tone melody for the NP=PhP alignment. Thus, $[\text{ma.}\mathbf{la}\text{.ɗaʔ}]_{N=PW}$ becomes the rightmost word in the aligned NP=PhP unit. Note that the head N [na.na.a] does not carry the prosodic peak LH*L because it is a PW that is not at the right edge of PhP.

(26) a.



'tasty food'

b. AVMs:

(i)

s-string

FM	*nanaa*
L	{NP, N}
R	{N }

p-string:

FM	[na.[na]$_{L.\sigma}$[a]$_{R.\sigma}$]$_F$]
STRESS	P$_{[na]L\sigma}$
L	{PhP, PW, F}
R	{PW, F, light.syl}

(ii)

s-string

FM	*ma**la**ɗaʔ*
L	{N}
R	{IP, NP, N }

p-string:

FM	[ma][[**la**][ɗaʔ] $_F$]
RB_TONE	[1][H*L]$_F$
STRESS	P$_{[la]L\sigma}$
L	{PhP, PW, F}
R	{IntP, PhP, PW, [1][H*L]$_F$ }

The left daughter's word [na.**na.**a]$_{N=PW}$ contains a light syllable (i.e., no C-coda)[13] and has the underlying form-meaning representation of *nanaat* 'food' (not V 'eat'). The surface form *nanaa* signals the deletion of a final coda suffix nominaliser (-*t*) and flags the presence of ADJUNCT in the NP via a heavy-to-light syllable alternation. This satisfies the ADJUNCT constraint stated in (25c.i).

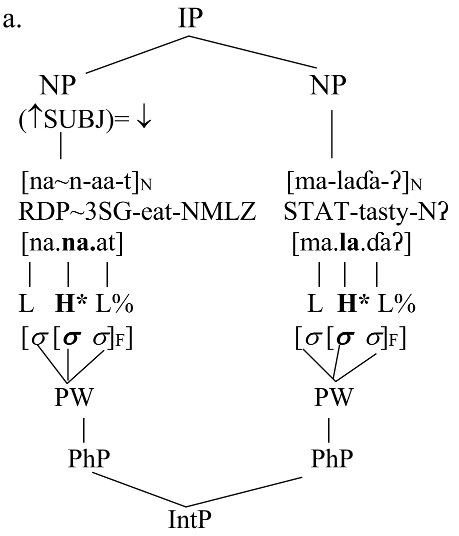
In contrast to (18a), the string in (18b) is parsed as two NPs. This is represented by the c- and p-structures in (27). There are two prosodic features that are critical to note. First, the left word N [na.**na.**at] 'food' contains a heavy syllable via a C final coda/nominaliser. Consequently, the following N [ma.**la**.ɗaʔ] cannot be an adjunct otherwise it would violate the NP Adjunct constraint in Dela (25.c.i). Second, the string contains a melodic H*L in addition to its word-final heavy syllable. This makes it a single PW/PhP that is aligned with NP. The resulting AVM values include R with the value of {NP, PhP, N, PW, F, H*L} as in (27b.i). Likewise, the right N [ma.**la**.ɗaʔ] is a single NP/PhP as seen in the L values of (27b.ii). In short, the top node is a syntactic unit of (sentential) IP and consists of two NPs that correlate with two PhPs.

We have demonstrated how p-structure properties serve as a marker in Dela syntax in the same way as agreement in morphosyntax (e.g. a prefix on the verb marks the presence of SUBJ). Thus, Dela data points captured by (25c) support the idea of a direct connection between p-structure and c-structure, which is consistent with the idea of 'transfer of structure' in Bögel (2015).

---

[13] Note that we also adopt an approach where syllables are units of p-string as seen in (26.a). The notation of 'light.syl' in (26b.i) is an informal shorthand of a feature-value pair [SYLL light].

(27) a.



‘The food is tasty’

b. AVM:

(i)

s-string

FM	*nanaat*
L	{IP, NP, N}
R	{NP, N}

p-string:

FM	[na.[na]$_{L.\sigma}$[at]$_{R.\sigma}$]$_F$]
RB_TONE	[1][H*L]$_F$
STRESS	P$_{[na]L\sigma}$
L	{IntP, PhP, PW, F}
R	{PhP, PW, [1][H*L]$_F$, heavy.syl}

(ii)

s-string

FM	*ma**la**ɗaʔ*
L	{N}
R	{IP, NP, N }

p-string:

FM	[ma][[**la**][ɗaʔ] $_F$]
RB_TONE	[1][H*L]$_F$
STRESS	P$_{[la]L\sigma}$
L	{PhP, PW, F}
R	{IntP, PhP, PW, [1][H*L]$_F$ }

5 Similar patterns from neighbouring languages

Phrase-medial and phrase-final alternations in NPs are a feature of both Austronesian and non-Austronesian languages in the Timor region (Culhane 2018:82; Tamelan 2021:241–246). However, the alternation is not a unitary phenomenon as it is marked differently across languages. Parallel forms and functions to Dela nominal alternations are found in a number of languages in the region. Some languages, including Amarasi (Edwards 2020), Helong (Balle 2017), Leti (van Engelenhoven 2004) and Mambae (Grimes et al. 2014) mark the nominal alternation by metathesis. Examples are given from Amarasi in (28). These examples show that non-final nominals in NPs undergo metathesis (open syllable → closed syllable, e.g. <u>afu</u> → <u>auf</u>), whereas final nominals do not. This distribution of unmetathesised and metathesised nouns is similar to that of consonant-final and vowel-final alternation in Dela.

(28) Attributive nominal phrases in Amarasi (Edwards 2020:272)

Noun (citation)		Modifier			NP	Gloss
afu	'earth'	*meʔe*	'read'	→	*auf meʔe*	'red earth'
fatu	'stone'	*mutiʔ*	'white'	→	*faut mutiʔ*	'white stone'
bare	'place'	*koʔu*	'big'	→	*baer koʔu*	'big place'
kase	'foreign'	*mutiʔ*	'white'	→	*kaes mutiʔ*	'European'
rasi	'matter'	*reʔuf*	'bad'	→	*rais reʔuf*	'evil matter'

Other languages, such as Amfo'an (Culhane 2018), Buru (Grimes 1991), Central Lembata (Fricke 2019) and Sawila (Kratochvíl 2014), mark nominal phrase-medial and phrase-final alternations by vowel and consonant final forms, respectively (i.e. phrase-medial=vowel, phrase-final=consonant), and are similar to the nominal alternation in Dela. Some examples from Amfo'an and Buru are given in (29) and (30). In these examples, the nouns with consonant and vowel alternation are underlined. Examples in (29) show that the consonant-final forms undergo consonant deletion before an attributive modifier. All nominals in Amfo'an have vowel-final and consonant-final forms.

(29) Consonant-final nominals in Amfo'an NPs (Culhane 2018:35)

Citation form	+	Modifier			Phrase	Gloss
sisidʒ	'meat'	*metoʔ*	'dried'	→	*sisi metoʔ*	'dried meat'
asuŋ	'dog'	*anaʔ*	'small'	→	*asu anaʔ*	'small dog'
kasel	'foreigner'	*mutiʔ*	'white'	→	*kase mutiʔ*	'white person'
muʔit	'animal'	*fuidʒ*	'wild'	→	*muʔi fuidʒ*	'wild animal'
kuan	'village'	*tuaf*	'person'	→	*kua tuaf*	'village person'

Similarly in Buru, nouns can have consonant-final and vowel-final alternation through truncation of the head noun roots before attributive modifiers.

(30) Truncation of roots (Grimes 1991)

Noun		Modifier			NP	Gloss
huma	'house'	*fatu*	'stone'	→	*hum fatu*	'stone house'
huma	'house'	*hawa*	'garden'	→	*hum hawa*	'garden house'
geba	'person'	*fehu-t*	'young'	→	*geb fehut*	'young person'

Comparatively, nominal alternation is marked differently across the languages in the region, however they all mark a head-adjunct relation. A summary of the different marking is presented in (31). For languages that have V-final and C-final alternation, such as Dela, Amfo'an and Buru, V-final nouns usually mark NP-internal adjunct relations, and C-final nouns mark NP-external adjunct relations. On the other hand, languages like Amarasi that have metathesised and unmetathesised alternations usually mark NP-internal adjunct relations via metathesised nouns, and NP-external adjunct relations via unmetathesised nouns.

(31) Summary of nominal alternation in some languages of Timor

Language	Adjunct marking (word-final syllable)
Dela, Amfo'an and Buru	(a) light, no coda NP-internal
	(b) Heavy syll, NP-external adjunct related
Amarasi	Metathesis: CV->VC, heavy syll

A detailed LFG analysis for the patterns in (31) is beyond the scope of the present paper. However, we believe that our proposed Dela analysis can be straightforwardly extended to cases in Amfo'an and Buru. The analysis for Amarasi would be slightly more complex as it requires a non-segmental mechanism to deal with metathesis in the morphosyntax-phonology interface.

6 Conclusion

In this paper, we have described a syntax-prosody interface phenomenon as seen in the NP head-adjunct structure in Dela. We have argued for two key empirical points to account for its prosodically marked NP head-adjunct relation: 1) word-final syllable alternation (light vs. heavy) encodes the presence/absence of an adjunct close to the NP head; 2) unit alignment of NP and Phonological Phrase (PhP) with the prosodic peak at the right edge the NP/PhP is marked by H*L.

Our study contributes to the theoretical and typological research on the nature and function of prosody in grammar. We have demonstrated that LFG's modular model nicely captures the syntax-prosody phenomenon in Dela. LFG's modular architecture provides a natural framework to account for the lexical and post-lexical phenomena exhibited by the alternation of word-final C deletion/insertion in the NP's head-adjunct relation.

We proposed two conditional 'if-then' phonology-morphosyntactic interface rules in LFG, making use of left (L)/right (R) edge features to account for the prosodic head-adjunct marking in Dela. We have demonstrated how the proposed LFG analysis can capture intricacies of phonology-morphosyntax in Dela, in particular the role of prosody for correctly parsing and disambiguating the syntax of (almost) identical s-strings.

Similar phenomena exploiting phonological resources to mark NP adjuncts (e.g. final C-deletion/insertion and metathesis) were also encountered in other languages in the eastern Indonesian region. We believe our analysis can be straightforwardly extended to these languages. Further research is needed to answer the following questions: 1) why are phonological resources only exploited between nominal units of an NP?; 2) what is special about an adjunct relation in contrast to other elements such as Q(uantifier) and D(eterminer) within the nominal?; 3) how common is this cross-linguistically? Since the phenomenon in Dela reveals that prosodic marking, such as word-final C-deletion, involves relational units closer to the head, we expect that much can be gained from further investigation into the mechanism and resources exploited in aligning lower equivalent units across domains in the hierarchical structure of phonology (prosodic word, prosodic phrase) and morphosyntax (morphological word and syntactic phrase).

7 References

Balle, Misriani. 2017. *Types of reduplication in Helong, an Austronesian language in eastern Indonesia.* Master's thesis, Payap University

Bögel, Tina. 2015. *The syntax-prosody interface in Lexical Functional Grammar.* PhD thesis, University of Konstanz.

Bögel, Tina fortcoming. "Prosody and its interfaces." In Mary Dalrymple (ed.), *The Handbook of Lexical Functional Grammar (Empirically Oriented Theoretical Morphology and Syntax)*, Berlin: Language Science Press.

Butt, Miriam and King, Tracy H. 1998. Interfacing Phonology with LFG. In Miriam Butt and Tracy H. King (eds.), *Proceedings of the LFG' 98 Conference*, Stanford: CSLI publications. http://csli-publications.stanford.edu/LFG/3/butt-king/butt-king.html.

Culhane, Kirsten. 2018. *Consonant insertions: A synchronic and diachronic account of Amfo'an*. Honours thesis, Australian National University. http://hdl.handle.net/1885/160794.

Dalrymple, Mary, Lowe, John J. and Mycock, Louise. 2019. *The Oxford Reference Guide to Lexical Functional Grammar*. Oxford: Oxford University Press.

Dalrymple, Mary and Mycock, Louise. 2011. The Prosody-Semantics Interface. In Miriam Butt and Tracy H. King (eds.), *Proceedings of the LFG' 11 Conference*, pages 173–193, Stanford: CSLI Publications.

Edwards, Owen. 2020. Metathesis and unmetathesis in Amarasi. *Studies in Diversity Linguistics 29*. Berlin: Language Science Press. https://langsci-press.org/catalog/book/228.

van Engelenhoven, Aone. 2004. *Leti, a language of southwest Maluku*. Leiden: KITLV press.

Fricke, Hanna L.A. 2019. *Traces of language contact: The Flores-Lembata languages in eastern Indonesia*. Ph. D.thesis, University of Leiden. http://hdl.handle.net/1887/80399.

Grimes, Charles E. 1991. *The Buru language of eastern Indonesia*. PhD thesis,, The Australian National University. http://hdl.handle.net/1885/10925.

Grimes, Charles E., Marçal, Carlos and Fereira, Paolino. 2014. *Introductory dictionary of Mambae (Same): Mambae—English, English—Mambae, Mambae—Indonesia— Tetun Dili, Indonesia—Mambae, Tetun Dili—Mambae.* Darwin: Australian Society for Indigenous Languages.

Kratochvíl, František. 2014. Sawila. In Antoinette Schapper (ed.), *The Papuan languages of Timor, Alor and Pantar Vol. 1*, pages 351–438. Berlin: De Gruyter Mouton.

Mycock, Louise and Lowe, John J. 2013. The prosodic marking of discourse functions. In Miriam Butt and Tracy H. King (eds.), *Proceedings of the LFG' 13 Conference*, pages 440–460, Standford: CSLI Publications. http://csli-publications.stanford.edu/.

Selkirk, Elisabeth O. 1986. On derived domains in sentence phonology. *Phonology Yearbook 3,* pages 371–405. DOI: 10.1017/s0952675700000695.

Selkirk, Elisabeth O. 2011. The syntax-phonology interface. In John A. Goldsmith, Jason Riggle and Alan C. L. Yu (eds.), *The handbook of phonological theory*, pages 435–484, Malden, MA: Blackwell. DOI: 10.1002/9781444343069.ch14.

Tamelan, Thersia M. 2021. *A grammar of Dela: an Austronesian language of Rote, eastern Indonesia.* PhD thesis, The Australian National University. http://hdl.handle.net/1885/250953.

Grammatical functions in the (Old English) Noun Phrase

Roxanne Taylor
University of Manchester

Proceedings of the LFG'21 Conference

On-Line

Miriam Butt, Jamie Y. Findlay, Ida Toivonen (Editors)

2021

CSLI Publications

pages 285–305

http://csli-publications.stanford.edu/LFG/2021

Keywords: grammatical functions, possessors, action nominal constructions, noun phrases, internal syntax

Taylor, Roxanne. 2021. Grammatical functions in the (Old English) Noun Phrase. In Butt, Miriam, Findlay, Jamie Y., & Toivonen, Ida (Eds.), *Proceedings of the LFG'21 Conference, On-Line*, 285–305. Stanford, CA: CSLI Publications.

Abstract

Noun phrase grammatical functions and the internal syntax of the noun phrase more generally have taken a back seat in Lexical Functional Grammar compared to work on grammatical functions in the verbal domain, and there remains no consensus as to the number and nature of grammatical functions postulated within the nominal domain. Outstanding issues include the validity and appeal of using traditionally verbal grammatical functions within the noun phrase, the characteristics of some distinctly nominal grammatical functions, and the diagnostic criteria used to identify grammatical functions in the noun phrase. This paper explores questions surrounding the identity and characteristics of noun-phrase internal grammatical functions, using newly collected empirical data from Old English to highlight the successes and pitfalls of previous accounts. The paper also makes tentative suggestions for two grammatical functions for the Old English noun phrase: a primary unrestricted function POSS, accounting for low valency in the noun phrase and instantiated not only by possessors but also by prepositional phrases and clausal complements, and a highly marginal oblique grammatical function.

1 Introduction

Work in Lexical Functional Grammar (LFG) on the grammatical functions (GFs) within the noun phrase (Markantonatou 1995; Sadler 2000; Laczkó 2000; Falk 2001; Kelling 2003; Chisarik & Payne 2003) has been relatively limited in comparison to work on GFs at the level of the clause and on argument mapping in the verb phrase. There is no consensus as to the number and identity of nominal GFs, nor as to whether nominal GFs are a distinct set from verbal GFs. This article assesses work on nominal GFs within LFG in light of new corpus data on action nominal constructions (ANCs) from Old English. The Old English data gives evidence for various phenomena which previous proposals do, and do not, account for, like reduced valency and diversity in surface forms. Building on and altering pre-existing formulations for nominal GFs in LFG, a tentative proposal is made for two GFs in the Old English noun phrase, a modified POSS no longer associated with possessor constructions, and OBL. Although POSS alone, with properties to match low valency in the noun phrase, is sufficient for most Old English ANCs, the presence of multiple arguments in some ANCs requires two GFs to be posited.

[†] I thank those who attended my poster virtually at LFG2021 for their fruitful comments and discussion, as well as the two reviewers whose detailed comments on a previous draft have greatly improved this paper and helped my thinking. All errors are of course my own.

The Old English data support the preference in existing studies (Markantonatou 1995; Sadler 2000; Falk 2001) for a subject-like GF within the noun phrase. However, the Old English data also lead to a rejection of the commonly-encountered association (Sadler 2000:97–99; Laczkó 2000:218; Dalrymple, Lowe & Mycock 2019:35) between the nominal GF POSS and morphosyntactic markers of possession, and the reliance on prepositional phrases at c-structure to identify OBL_θ (Kelling 2003).

This paper begins with a review of pre-existing proposals for nominal GFs from different studies in the LFG, and a brief introduction to the Old English material. Four sections of empirically-grounded analysis follow, addressing different patterns of how arguments are realised in Old English ANCs. These sections use the prevalence of possessor forms in the Old English data set, low valency in the noun phrase, the marginal presence of non-possessor forms, and finally patterns of co-occurrence as the basis for comparison with and criticism of previous theoretical suggestions. Section 8 briefly considers the arguments for and against distinguishing POSS from SUBJ

2 Previous work on noun phrase grammatical functions

As a general characterisation, work in LFG on nominal GFs dates to the turn of the millennium, and focuses on nominal GFs in the context of argument structure and argument-structural inheritance in deverbal nominalisation (Falk 2001; Kelling 2003). As such, in common with work on nominal syntax beyond LFG (Grimshaw 1990), the focus is on ANCs, event-denoting noun phrases, rather than on the canonical noun phrase, without eventive semantics. For the purposes of this article, following Comrie & Thomson (1985:358), an ANC is defined as noun phrase headed by a derived eventive nominal, ‘with one or more reflexes of a proposition or a predicate’, and containing one or more “reflex(es)” or expressions of the participants in this proposition.

(1) the enemy’s destruction of the city

(2) my horse’s winning of the race was no surprise

Laczkó (2000) and Falk (2001) focus exclusively on ANCs, addressing argument mapping in Hungarian and Hebrew ANCs respectively. Markantonatou (1995) and Kelling (2003) focus on psych-verb ANCs in Modern Greek and French respectively, namely ANCs with nominal heads derived from psychological predicates.[1]

[1](3) is Markantonatou’s (1995) (54).

(3) o misos tis Marias yia ton Yiani epi tosa hronia ine
the hatred the-GEN Maria-GEN for the Yianis for so-many years is
paralogo
unreasonable
Maria's hate for Yianis for so many years is unreasonable

Sadler (2000) and Chisarik & Payne (2003) are exceptions; both studies consider the full gamut of noun phrases in Welsh, present-day English, and Hungarian.

Much of the scholarship mentioned thus far (Markantonatou 1995; Laczkó 2000; Sadler 2000; Falk 2001) adopts a nominal GF, POSS. POSS is consistently understood as being available only in the noun phrase, not in the clause (Laczkó 2000; Dalrymple, Lowe & Mycock 2019:35). In addition, all proposals for POSS work with an understanding that the grammatical function is [-o(bjective)] and [-r(estricted)] (Markantonatou 1995:283; Sadler 2000:97; Laczkó 2000:211). POSS has a close connection to possessors: the GF has been elucidated with explicit reference both to the morphosyntactic sense of possessors and to the semantic role of possessors (Sadler 2000:97-101), with examples like (4) used to identify the grammatical function. (4) shows both a semantic role of canonical possession— ownership— and a morphosyntactic possessor— the clitic *s* (Dalrymple, Lowe & Mycock 2019:35).[2]

(4) Chris' book

POSS, understood as [-o, -r], is featurally identical to SUBJ. The exact nature of the relationship between POSS and SUBJ has been an important point of debate. Positions range from the total separation of POSS from SUBJ (Laczkó) to the eradication of POSS in favour of SUBJ (Chisarik & Payne 2003:185). Ambivalence on the question is evident in Markantonatou (1995:284), where reference is made to an unrestricted function. Others view POSS as a nominally oriented subset of a single function SUBJ: Sadler (2000:97) describes POSS as "SUBJective and discourse-oriented". A similar view of POSS as a kind of subtype of SUBJ is also evident in Falk (2001:96). In Falk's analysis, POSS and SUBJ are distinct attributes at f-structure, although since SUBJ and POSS share a single f-structure as their value, POSS has the role of a nominal 'proxy' for SUBJ, part of the SUBJ grammatical function.

The number of other GFs postulated for the noun phrase alongside POSS varies (none, one, two, or more). Several studies (Laczkó 2000:212; Falk 2001) suggest that OBL_θ can appear alongside POSS in the noun phrase, for instance in

[2](4) is Dalrymple, Lowe & Mycock's (2019) (67).

so-called “passive” ANCs which feature a realisation of the agent argument by an oblique, as seen with passive verbs. Markantonatou (1995:283,287) finds that Modern Greek deverbal psych nominals can contain only a single instantiation of the [-r] function POSS/SUBJ, but unlimited instantiations of OBL_θ. Kelling’s proposal (2003:175) for psychological ANCs in French takes a rather different approach; in these noun phrases, Kelling determines that OBL is the sole GF, taking the part filled by POSS in other studies. Multiple instantiations of OBL, specified by form (OBL_{de}, OBL_{pour}) can co-occur in the French psych noun phrase.

Aside from OBL_θ and POSS, one other GF has been proposed for the noun phrase: the entirely novel ADNOM postulated by Chisarik & Payne (2003:185–186). ADNOM is proposed to account for a small group of typologically unusual languages, including present-day English and Hungarian, in which there are two default possessor constructions in variation (i.e. two default markers of adnominal dependency which can both mark possessor semantic relations). As well as being restricted in its applicability to the languages of the world, there are difficulties with the the reliance in Chisarik & Payne (2003) on an ad-hoc feature [±d(iscourse oriented)] to distinguish ADNOM ([-d]) from POSS/SUBJ ([+d]).

There is no common consensus as to how many GFs might be needed within the noun phrase and whether or not, and how, these GFs might differ from those assumed for the verbal domain.

3 Old English Action Nominal Constructions

In line with the prevailing trend in work on nominal GFs (Markantonatou 1995; Falk 2001; Kelling 2003), the empirical focus in this study is not on noun phrases generally, but rather on a specific set of noun phrases: ANCs. It is assumed that eventive nouns, which head ANCs can take arguments, just as can verbal predicates. According to Needham & Toivonen’s criteria for argumenthood (2011:404–405), an argument is any participant necessary for the event described by the predicate but also specific to the predicate in question. Since this definition is formulated in essentially semantic terms of events and participants, it is as appropriate for nouns denoting events as it is for verbal predicates; a criterion referring to the specificity of a participant to an event can be applied equally well to nominal predicates. The adnominal dependents in the ANC to which this paper makes reference are therefore assumed to be arguments.

Old English ANCs are headed by deverbal nominal predicates in *-ung* and -

ness. [3] Old English ANCs resemble canonical noun phrases in their external syntactic distribution. Aside from the eventive semantics of the head nouns, there are no grounds for adopting a mixed category analysis along the lines of that used for present-day English gerunds or seen in Bresnan & Mugane (2006). All Old English examples are drawn from the York-Toronto-Helsinki Parsed Corpus of Old English Prose (2003) (henceforth, *YCOE*) and are referred to with *YCOE* token IDs.

(5) se apostol Paulus spræc be
DET.NOM.SG apostle-NOM.SG Paul-NOM.SG speak.PST.3SG PREP
ðære getimbrunge **þære geleaffullan**
DET.DAT.SG building-DAT.SG DET.GEN.SG faithful-GEN.SG
gelaðunge
congregation-GEN.SG
Paul the apostle spoke about the construction of the faithful congregation
(cocathom2, ÆCHom_II,_45:342.223.7667)

(6) ic cyðe eow ætforan eallum
1SG.NOM make-known-1SG.NPST. 2PL.DAT PREP all-DAT.SG
folce **eower** unrihtwisan
people-DAT.SG 2PL.POSS.ACC.SG unrighteous-ACC.SG
ehtnysse **ofer ða cristenan**
persecution-ACC.SG PREP DET.ACC.PL christian-ACC.PL
I must make known to you, in front of all the people, your unrighteous persecution of the Christians
(coaelive, ÆLS_[Sebastian]:451.1485)

Old English ANCs were retrieved from *YCOE*, a 1.5 million word corpus with part of speech annotation. ANCs were identified in the corpus as those noun phrases headed by a deverbative noun with eventive semantics and including some realisation of at least one argument of the nominal eventive predicate. The noun phrases were retrieved from the corpus by way of head morphology and syntactic structure. All noun phrases headed by a noun suffixed with *-ung* or *-ness* and containing some adnominal dependent were retrieved using *CorpusSearch2* (Randall 2003).[4] Noun phrases were annotated automatically

[3] Although present-day English *-ness* only denotes abstract qualities, it can form nouns with eventive semantics in Old English; present-day English *-ing* forms verbal and nominal gerunds, as well as deverbal nouns, but in Old English there are no gerunds like this; verbal participles are not formally identical with deverbal suffixed nouns.

[4] The corpus was interrogated for noun phrases headed by nouns containing the strings U-N-G, I-N-G, Y-N-G, U-N-C-G, I-N-C-G, Y-N-C-G, N-E-S, N-I-S, N-Y-S, N-Æ-S, and N-U-S.

(using *CorpusSearch2*), and manually for the semantic relation between head and dependent, the type of dependent and its position relative to the head noun. The resulting data set consists of 3472 noun phrases. Null hypothesis statistical testing and binomial and multinomial logistic regressions were carried out using *R* (*R* Core Team 2021).

Old English ANCs mostly include a single genitive case marked noun phrase (henceforth "genitive noun phrase") as a realisation of an argument of the nominal predicate (5).[5]

(7) þa he in æghwæðerum mynstre
CONJ 3SG.NOM.MASC PREP either-DAT.SG monastery-DAT.SG
Hilde þære abbudissan geornlice **his**
Hilde-GEN DET.GEN.SG abbess-GEN.SG eagerly 3SG.GEN.MASC
leornunge ætfealh
learning-ACC.SG adhere.PST.3SG
when he was in either monastery of the abbess Hilde, he eagerly stuck to his learning (cobede, Bede_4:24.334.30.3363)

As well as argument-realising genitive noun phrases, Old English ANCs also include prepositional phrases (8) and clausal complements (9) as forms of argument realisation.[6]

(8) ond æfter Cristes upastignesse heo
and PREP Christ-GEN.SG ascension-OBLIQ.SG 3SG.NOM.FEM
wæs on swa micelre longunge **æfter**
be.PST.3SG PREP so great-OBLIQ.SG desire-OBLIQ.SG PREP
him
3SG.DAT.MASC
and following Christ's ascension she was in a state of great desire for him
(comart3, Mart_5_[Kotzor]:Jy22,A.16.1232)

[5]In Old English, unlike in present-day English, there is only a single marker of adnominal dependency, the morphological genitive, *of* at this stage in the history of English remains a lexical preposition (Allen 2008:72–73).

[6]To avoid confusion with the GF OBL$_\theta$, indeterminate accusative/dative/genitive case marking in Old English is glossed as OBLIQ.

(9) þam deofle wæs micel twynung
DET.DAT.SG devil-DAT.SG be.PST.3SG great.NOM.SG doubt.NOM.SG
hwæt Crist wære
COMP Christ.NOM.SG be.SUBJ.3SG
there was in the devil great doubt what Christ was
(cocathom1, ÆCHom_I,_11:267.37.2013)

Old English ANCs can also contain multiple means of argument realisation, as in (10).

(10) þurh **Godes** foresceawunge **þæt heo symle**
PREP God-GEN.SG foresight-OBLIQ.SG COMP 3SG.NOM.FEM ever
on anre stowe ne wunige
PREP one.OBLIQ.SG place-OBLIQ.SG NEG dwell-SUBJ.3SG
through God's prediction that she would never dwell in a single place
(cotempo, ÆTemp:4.42.165)

Table 1 details of numbers of adnominal dependents realising arguments in the ANCs of the data set.

NUMBER OF ARGUMENT-REALISING ADNOMINAL DEPENDENT(S) IN THE ANC	NUMBER OF OBSERVATIONS
one adnominal dependent	3443
two adnominal dependents	29
more than two adnominal dependents	0
TOTAL	3472

Table 2 shows the distribution of types of adnominal dependents realising arguments in the data set.

TYPE(S) OF ARGUMENT-REALISING ADNOMINAL DEPENDENT(S) IN THE ANC	NUMBER OF OBSERVATIONS
one genitive case noun phrase	3379
one prepositional phrase	47
one clausal complement	17
one genitive case noun phrase + one prepositional phrase	25
one genitive case noun phrase + one clausal complement	3
other combination	1
TOTAL	3472

4 Genitive noun phrases in the ANC

In the data set of Old English ANCs, the great majority of observations (97%, *n*= 3379) include as the sole argument-realising adnominal dependent a genitive noun phrase.

(11) þurh **ðæs** **apostoles** mungunge þe ðus cwæþ
PREP DET.GEN.SG apostle-GEN.SG admonishing-OBLIQ.SG REL thus say.PST.3SG
through the apostle's admonishing, who spoke thus
(cobenrul, BenR:28.52.18.648)

Genitive noun phrase arguments in the ANC are identical in terms of morphological form to genitive possessor noun phrases in non-ANC noun phrases. Two canonical possessors (*hire*, *Zacharian*) and a genitive argument of the eventive noun *bodung* are illustrated in (12).

(12) Maria ferde æfter **þæs** **engles** bodunge to **hire** magan Elisabeð. Seo wæs **Zacharian** wif
Mary.NOM.SG go.PST.3SG PREP DET.GEN.SG angel-GEN.SG instructing-OBLIQ.SG PREP 3SG.POSS.FEM.OBLIQ kinswoman.OBLIQ.SG Elizabeth.OBLIQ.SG REL.3SG.FEM.NOM be.PST.3SG Zachariah-GEN.SG wife.NOM.SG
Mary went, after the instruction of the angel, to her kinswoman Elizabeth, who was the wife of Zachariah
(cocathom1, ÆCHom_I,_13:286.160.2492)

Genitive noun phrases in ANCs, which realise some argument of the eventive head noun, are also found to show behaviour similar to that established for genitive noun phrases in non-ANC noun phrases in previous studies of Old English nominal syntax. Both Koike (2006:50) and Allen (2008:114) find from their corpus-based studies that GENITIVE—HEAD (seen in *hire magan* and *Zacharian wif* in (12)) is the preferred order across the period 750–1100CE. Quantitative investigation finds that this general preference for pre-head genitive noun phrases is observed also in the ANCs. According to a chi-square goodness of fit test, pre-head position for adnominal argument-realising genitives is highly significantly more frequent than post-head position (χ^2 =982.22 df =1, $p_{\text{two-tailed}} < 0.0001$).

Table 3 details of numbers of pre-head and post-head genitives realising arguments in the ANCs of the data set.

ORDER OF HEAD AND GENITIVE DEPENDENT	NUMBER OF OBSERVATIONS
genitive—head	2651
head—genitive	756
TOTAL	3407

The most common adnominal dependent in an Old English ANC resembles a canonical Old English possessor both in its form and in its interactions with the head noun. In respect of the long-standing connection in the literature between POSS and possessor constructions POSS would seem to be an appropriate GF to handle Old English ANCs like (5). However, the close association between POSS and possessor constructions is highly problematic. In the verbal domain, although in a given language there will be some association(s) between surface forms and GFs, the proposal for or creation of a GFs is not based in language-specific surface forms (barring functions like OBL_{on} for expressions like *rely on*). This is not an objection to the GF POSS in itself, but rather to the reliance on possessor forms to motivate and define POSS. We need to consider the characterisation of POSS in other ways, and it is to this which we now turn.

5 Number of arguments in the ANC and valency

Quite regardless of any connection to possessor forms, Sadler's (2000:97) proposal for POSS featurally identifies it as [-r(estricted), -o(bjective)], hence, in featural terms, identical to SUBJ. Setting aside for the moment the issue of whether distinct syntactic categories need distinct GFs, what is important about the association of POSS and SUBJ is the comment it makes on the hierarchy, interdependencies and competitiveness of GFs. As Findlay (2020:137) notes, although other GFs are in competition for argument slots, SUBJ stands outside of these competitions and dependencies at the top of the GF hierarchy. SUBJ is not reliant on other GFs, in for instance the way that the presence of OBJ requires SUBJ. Consequently, SUBJ can be the sole GF instantiated in a given context. Identifying POSS with SUBJ similarly implies that POSS can be the sole GF instantiated in a given noun phrase. This is bourne out in the specific analyses provided by Sadler (2000:99–100) for Welsh noun phrases, and those of Markantonatou (1995:287) and Chisarik & Payne (2003:187,189).

99% (*n*= 3443) of ANCs in the data set feature only a single adnominal dependent, realising a single argument of the nominal head. A single argument is the norm even when, at a conceptual level, the event denoted by the head involves two or three participants. This is evident in (5), where the event of building conceptually requires both builder and thing built but only the latter is expressed; in (8), where only the thing desired is expressed although a state of longing requires a desirer too; in (13) a confessing agent, what is confessed, and a person who receives the confession are conceptually necessary but only the latter is expressed.

(13) to Gode gecyrran nellað þurh
PREP god-DAT.SG turn-INF NEG-WANT-NPST.PL PREP
soðe andetnesse **mæssepreosta**
true-OBLIQ.SG confession-OBLIQ.SG priest-GEN.PL
they do not want to turn to God through true confession to priests
(coverhom, HomS_4_[ScraggVerc_9]:18.1248)

The verbs from which the eventive nominalisations derive may be monotransitive or ditransitive, but the overwhelming preference is nevertheless for only a single argument within the noun phrase.

There is substantial evidence to indicate that reduction in valency is a pervasive characteristic of the Old English ANC. 57% of those deverbal nominal heads deemed to have multiple arguments with them take part in a variation as to which argument is realised within the ANC. That is to say that for these heads some ANCs in the data set show one of their arguments realised, whilst other observations show a different argument realised within the noun phrase. Such variation, demonstrated in (14) and (15) implies that the distribution of particular arguments appearing in ANCs is not reflective of a particularly strong limitation on which arguments roles are preferentially realised in the ANC— for instance, it is not that arguments corresponding to objects in the noun phrase are favoured. If there is argument realisation in the ANC, the prevailing tendency is for only one argument realised per noun phrase.[7]

(14) ac ic **his** giomrunga gehyrde
but 1SG.NOM 3SG.GEN.MASC lamentation-OBLIQ.SG hear-PST.1SG
but I heard his lamentation
(coverhom, HomS_40.3_[ScraggVerc_10]:134.1466)

(15) on ðisum dagum we forlætað on
PREP DEM.DAT.PL day-DAT.PL 1PL.NOM relinquish-NPST.3PL PREP
urum repsum Gloria patri for
1PL.POSS.DAT.PL response-DAT.PL Gloria patri PREP

[7]Only around a fifth of noun phrases headed by deverbatives in the corpus appear with any form of argument realisation.

geomerunge **þære** **halgan** **ðrowunge**
lamentation-OBLIQ.SG DET.GEN.SG holy-GEN.SG suffering-GEN.SG
on these days, we put aside the Glory Be *in our liturgical responses, because of the lamenting of the Holy Passion*
(cocathom2, ÆCHom_II,_13:127.8.2776)

A second piece of evidence for reduced valency in the ANC comes from the importance of the lexical identity of the head as a factor in the realisation or non-realisation of different arguments in the ANC. Binary logistic regression modelling was used to identify which of a range of predictors (for instance, weight, animacy, event class of predicate), and interactions of such predictors, gives the highest chance of correctly predicting whether it is the subject-like or object-like argument of a monotransitive or ditransitive nominal predicate which is realised in a particular noun phrase. Models including different predictors and their interactions were compared for success, where success equates to better-than-chance correct prediction of which argument appears in a noun phrase. This statistical analysis indicates that by far the most successful model with a single predictor is one with the predictor lexical identity of the head noun in an ANC (whether the head noun is the lexeme TIMBRUNG, 'building', or EHTNESS, 'persecution', or some other lexeme): Nagelkerke's $R^2 = 0.604$, $C = 0.916$.[8] The impact of lexical identity on argument variation indicates that a reduction in valency is common to all deverbative heads; it is being nominal which gives these deverbative predicates reduced valency, whilst the specific identity of the noun determines which argument preferentially gets realised in the ANC.

ANCs generally exhibit reduced valency. The GF SUBJ (or a noun phrase equivalent POSS) is most appropriate to capture this reduced valency, since SUBJ can be the only GF instantiated in a given context. As the highest GF in a hierarchy based on markedness, SUBJ is not dependent on any other GF for instantiation nor does it compete with other GFs in mapping. These properties make SUBJ a good match for the behaviour of arguments in the Old English ANC; there is no need to posit a dependent GF lower in the hierarchy which participates in competition with other GFs. That said SUBJ is not always the only GF in a given context, nor does the presence of SUBJ preclude the instantiation of other GFs. Although SUBJ is the most appropriate to account for low valency, it does not guarantee or motivate this property of the ANC: in other words, SUBJ is descriptively adequate but offers no explanatory gain. Accounting for the arguments in ANCs with SUBJ/POSS in this way has an advantage

[8]To avoid false reporting of the impact of the head's lexical identity, the data-set which was used to test the impact of the predictor "head lexeme" included only those observations headed by nouns with frequency ≥ 6, 102 heads, 2342 ANCs.

over previous proposals since it requires no reference to form in general nor to possessor constructions specifically.

6 Prepositions in the ANC

In the Old English data, not only are genitive noun phrases observed as the sole means of argument realisation (5, 7, 11, 13) , this is also true of prepositional phrases which likewise can appear as the only form of argument realisation in an ANC (8, 16).

(16) nu hæbbe we ða alysednysse **þurh**
now have.NPST 1PL.NOM DET.ACC.SG salvation-ACC.SG PREP
ðone leofan Drihten
DET.ACC.SG beloved-ACC.SG God.ACC.SG
now we have salvation through the beloved Lord God
(coaelhom, ÆHom_6:262.1005)

It is true that prepositional phrases are in a considerable minority as sole means of argument realisation in the data set, compared with genitive noun phrases. However, ANCs resembling (8) and (16) are not rare or marginal in the data set (*n*= 47). These ANCs illustrate a wide range of prepositional heads drawn from different semantic fields, which have varying core and extended uses and occur with different degrees of frequency in the Old English corpus.[9] Importantly for the identification of a [-r] GF, there is no restriction on the semantic roles of the arguments realised by prepositional phrases in the ANC: prepositional phrases as the sole means of argument realisation realise agents, experiencers, themes, patients, and stimuli.

Clausal complements can also occur as the sole means of argument realisation in the ANC (9), and represent an even smaller minority (*n*=17). These clausal complements do not evidence semantic unrestrictedness, only realising the stimuli, and themes of speech act predicates and predicates of mental consideration. In addition, clausal complements as the sole means of realising arguments are only observed with a small set of nominal heads, whose corresponding verbal predicates also take clausal arguments. For these reasons, clausal complements as the only means of argument realisation in the ANC are assumed to instantiate OBL_θ and are not considered further.

POSS ought to be appropriate for the prepositional phrase arguments in the 47 ANCs like (16). There is no semantic restrictiveness evident as to which arguments can be realised by prepositional phrases, and no sense in which these

[9]These prepositions are all understood as lexical prepositions, in contrast to the functional status of present-day English *of*.

prepositional phrases are dependent on the instantiation of another GF. POSS or SUBJ would be an appropriate GF to descriptively account for low valency in the ANC, seen in (8) and (16) just as in (13) and other ANCs with single genitive noun phrase arguments. Nevertheless, the prepositional phrase realisation of arguments causes problems under the commonly-found view of POSS which draws a close association between semantic possession, morphosyntactic possessors and the nominal grammatical function (Sadler 2000:97; Dalrymple, Lowe & Mycock 2019:35). Prepositions like *æfter* and *þurh* are not possessors in Old English (although of course, prepositions can be possessors, as in French, and can therefore presumably instantiate POSS in French). The solution presented by Old English ANC examples like (8) and (16) is to disassociate POSS from possessor constructions, taking POSS seriously as a GF divorced from a particular surface form. The claim is therefore that a prepositional phrase, headed by a lexical preposition can instantiate POSS. More to the point, a form which is not a possessor construction, and is not used to mark any possessive semantic relations, can instantiate POSS. POSS can remain as a [-o, -r] GF, since these featural specifications allow a descriptive account of low valency in the ANC. However, POSS is divorced from possessor forms.

7 Multiple dependents in the ANC

In a small number of instances, there are multiple argument-realising dependents in the ANC (*n*= 28).

(17) þu goda cyning ne understentst
2SG.NOM good.NOM.SG king.NOM.SG NEG understand.NPST.2SG
þu **þysra** **twegra** **manna** gereonunge
2SG.NOM DEM.GEN.PL two-GEN.PL man-GEN.PL plotting-ACC.SG
ongean me
PREP 1SG.OBLIQ
do you, good king, not understand these men's plotting against me?
(cocathom1, ÆCHom_I,_26:396.226.5159)

(18) he cydde syððan **his**
3SG.NOM.MASC make-known-PST.3SG afterwards 3SG.GEN.MASC
facenfullan syrewunge **hu** **he** **embe**
deceitful-ACC.SG plotting-ACC.SG COMP 3SG.NOM.MASC ADV
wolde
wish-PST.3SG
but afterwards he made known his plotting how he would act on the matter
(cocathom1, ÆCHom_I,_5:219.79.963)

These ANCs indicate the need for two distinct nominal GFs. A single nominal GF cannot account for the distinct realisations of two different arguments; specifically POSS can account for only one of the two arguments in (17) and (18). Moreover, the fact that these ANCs constitute a minority in the data set indicates that a second nominal GF has the status of an optional extra in the Old English noun phrase, being subordinate in frequency and range of use to POSS.

From the LFG literature, there emerge two possibilities for a second nominal GF to accompany POSS (however POSS is understood). The first is Chisarik & Payne's (2003) ADNOM, the second the more recognisable OBL_θ. ADNOM (Chisarik & Payne 2003) has already been put aside for the Old English ANC on the grounds that Old English has only a single possessive construction unlike PDE. It remains only to observe that the prepositions in ANCs with multiple arguments realised, including the preposition *of* (which means 'out of' in this period), are lexical prepositions. The remaining possibility for a second GF is, on the basis of pre-existing proposals OBL_θ. In Falk (2001:97), Laczkó (2000:212), and Markantonatou (1995:289), OBL_θ is employed in the same way as would be appropriate for the Old English noun phrase: to account for the 'optional' or less-frequently observed extra argument in the ANC, although for Markantonatou (1995:289), and Falk (2001:97) there is an association between OBL_θ and agentive prepositions resembling the agents in passive verb phrases which is not applicable in the Old English data set.

In the present data set, ANCs with two adnominal dependents are a tiny minority (*n*= 28). OBL_θ only needs to be invoked in a small number of instances where POSS alone is insufficient to account for realisation of multiple arguments. In the standard understanding, OBL_θ is featurally specified as [+r(estricted), -o(bjective)], and is characterised by way of optionality and semantic restrictiveness. Both these characteristics are a good fit with the Old English data: two arguments realised in the ANC is a rarity, meeting the criterion of optionality.

There is also evidence to meet the criterion of semantic restrictedness. In ANCs like (17) and (18) with two realised arguments, the genitive noun phrase is always a realisation of the argument with the greater number of proto-agent entailments (adopting Dowty's 1991 proto-roles rather than thematic roles). The prepositional phrase or clausal complement in turn realises the argument with the greater number of proto-patient entailments. Insofar as these prepositional phrases (and clausal complements) realise arguments with proto-patient entailments, corresponding to the object or indirect object of the verb phrase, there is a semantic or thematic restriction operative in the Old English data which dovetails appropriately with our expectations of the semantic restrictiveness of OBL_{θ}. Note that this is only true if the prepositional phrase appears alongside another realised argument; when prepositional phrases appear as sole means of argument realisation there is no similar semantic constraint, as is evident from examples with agentive prepositions like (16). (19) demonstrates the pattern whereby a genitive and another adnominal dependent realise subject-like and object-like arguments respectively.

(19) se God þonne þe is ure
DET.NOM.SG God.NOM.SG then REL be.NPST.3SG 1PL.GEN
ealra gemæne gefylle mildelice
all-GEN.SG in common fulfil-NPST.3SG graciously
eowre gewilnunge **to** **his**
2PL.POSS.ACC.SG desire-ACC.SG PREP 3SG.GEN.MASC
wuldre 7 **to** **haligre** **lare**
glory-OBLIQ.SG and PREP holy-OBLIQ.SG teaching-OBLIQ.SG
eowres **lifes**
2PL.POSS.GEN.SG life-GEN.SG
God, then, who belongs to us all in common, graciously fulfils your desire for his glory and for holy teaching for your life
(cochdrul, ChrodR_1:79.6.946)

In (19), it is the desirer, with semantic entailments like volition, animacy, and instigation, of a proto-agent, which is realised by a genitive noun phrase, whilst the *to*-phrases realise the thing desired, with the semantic entailments of the proto-patient, like inanimacy, abstractedness, and non-volition. Semantic restrictedness is evident in the pattern, visible in (19), (17), and (18), whereby two arguments realised in the ANC have a hierarchical relationship, genitives realising higher arguments and prepositional phrases and clausal complements restricted to realising lower arguments. The analysis of (19) is therefore that the *to*-phrases instantiate OBL_{θ}, whilst *eowre* instantiates POSS. It is assumed that the least marked argument, the experiencing desirer maps to POSS, being like SUBJ the most prominent GF free of dependencies on other GFs. There is a descriptive association with possessor form only insofar as most ANCs with two arguments realised feature a combination of a genitive and either a prepositional phrase or a clausal complement; the hierarchical relationship between these forms (genitive realises the higher argument) falls out exclusively from the [-r, -o] status of POSS and the [+r] status of OBL_θ.

There are two dimensions to OBL_θ which have a particular prominence in the literature on nominal GFs. The first is the notion of semantic restrictiveness, already considered for the Old English ANC. The second is an association with prepositional phrases, parallel to the frequently-seen association of POSS and possessor constructions. Kelling (2003) is the most conspicuous proponent of the view that a prepositional phrase within an ANC represents an instantiation of OBL_θ. OBL_θ is selected by Kelling (2003:175) as the relevant GF for French psych nominal ANCs, on the grounds that the experiencer and stimulus participants are expressed by prepositional phrase headed by *de*, *a*, and less frequently *pour*. *a* and *de* are generally considered functional prepositions, and might therefore contradict the [+r] status of OBL_θ. These prepositions mark various arguments of psych nominal ANCs, as well as arguments in other French ANCs, also contradicting the restricted status OBL_θ. With these contradictions between the properties of OBL_θ and the relevant French prepositions in mind, it seems that it is precisely the prepositional nature of the argument realisation, in other words, a question of form, which motivates the proposal for OBL_θ.

Prepositional phrases have so far played a prominent, albeit not exclusive, role in the discussion of OBL_θ as a nominal GF for the Old English ANC. However, the close connection between OBL_θ and prepositional phrases in the preexisting LFG literature (Kelling 2003) proves problematic in the face of variation of form in the Old English data. Prepositional phrases vary with clausal complements as an additional means of argument realisation alongside a genitive noun phrase (18). The conclusion that neither POSS nor OBL_θ is bound by an association to a particular morphosyntactic form, contrary to the perspectives expressed in Sadler (2000:97), Falk (2001:96), and Kelling (2003) leads to the prediction that any combination of two adnominal dependents ought to be a possibility in the Old English ANC. This prediction holds: there is one noun phrase in the data set observed with two prepositional phrases dependent on the same deverbal head. The working analysis is that the higher argument, realised with a *betwux*-phrase maps to POSS whilst the lower argument, realised with a *be*-phrase maps to OBL_θ.

(20) þa wearð micel twynung **betwux**
then become.PST.3SG great.NOM.SG doubt.NOM.SG PREP

þære **burhware** ***be*** ***ðære*** ***cyrcan***
DET.DAT.SG community-DAT.SG PREP DET.DAT.SG church-DAT.SG
hwæðer hi ineodon oððe hi halgian
COMP 3PL.NOM enter-PST.PL or 3SG.ACC.FEM hallow-INF
sceoldon
should-PST.PL

then there arose a great doubt amongst the community concerning the church, whether they ought to go in or hallow it

(cocathom1, ÆCHom_I,_34:467.71.6734)

8 The question of syntactic categories and GFs

Some previous studies have sought a segregation of POSS and SUBJ, others the identity of the two GFs. This section briefly reviews the evidence from Old English ANCs for each position.

ANCs in Old English consistently display a tendency towards monovalency in spite of the transitivity of the base verb from which an eventive nominal is derived. The lexical identity of the nominal predicate strongly influences which argument gets realised in the ANC. These are properties peculiar to the noun phrase. Reduction in valency must therefore be viewed as a characteristic differentiating noun phrase argument structure from argument structure at the level of the clause. Such a consideration might be used to support the view that different syntactic categories require different GFs. As we have seen, either SUBJ or its noun phrase equivalent POSS can descriptively account for low valency in the ANC, neither dependent for instantiation on any other GF; it is not possible to adjudicate between SUBJ and POSS on these grounds since both GFs are appropriate for low valency in the ANC. Neither as it stands offers a motivation for low valency.[10]

POSS is sometimes argued to be distinct from SUBJ on the grounds that there is greater diversity of semantic relations operative between a nominal head and adnominal dependents, than between a verbal head and its subject. This is the argument made by Sadler (2000:97), where non-ANC noun phrases are included in the analysis to demonstrate that POSS incorporates canonical possession, and kinship. The present investigation must reject the conclusion that POSS is more diverse than SUBJ, on the grounds that an association between

[10]To differentiate SUBJ from POSS and to motivate low valency in the ANC, an additional characterisation would need to be made of POSS, circumscribing the instantiation of other GFs alongside POSS, something which is not a characterisation of SUBJ. But it would also be possible to handle this elsewhere in the LFG architecture, i.e. at s-structure or a-structure.

POSS and possessors, which do indeed mark a great range of semantic relations in the non-ANC noun phrase, is not accepted. Accordingly, the range of semantic relations available to possessors has nothing to do with the semantic unrestrictedness of POSS. It is true that the genitive noun phrase arguments in ANCs closely resemble genitives beyond the ANC in Old English, which mark a wide raft of semantic relations (kinship, ownership, part-wholes). For the arguments in the ANC, however, there is no evidence for a notable diversity of semantic roles which would support POSS distinct from SUBJ. The working conclusion drawn is that there does not need to be a GF POSS distinct from SUBJ to account for the Old English ANC: the reduced valency of ANCs can be described by either GF, but not explained by way of POSS as it is currently understood. Likewise there is no evidence from the Old English data set for a greater degree of semantic unrestrictedness to motivate a distinct POSS.

9 Concluding remarks

We are in a position to make certain positive and negative claims about nominal GFs in light of the newly collected Old English evidence. In the first instance, suggestions for a GF POSS/SUBJ successfully account for the low valency of nominal predicates, at least in descriptive terms. However, the association between POSS and possessor forms does not hold for a minority of the Old English data; rather possessors and non-possessors (prepositional phrases and clausal complements) alike are able to instantiate a semantically unrestricted GF POSS. A very small number of Old English ANCs require a second GF. Evidence in favour of OBL_θ comes from the rarity of noun phrases with two realised arguments and the semantic restrictions evident when two arguments co-occur. As with the relationship between POSS and possessors, an assumption that a given form, specifically a prepositional phrase, is closely associated with OBL_θ is challenged by the variation observed in the data set between different forms of argument realisation, in other words between prepositional phrases and clausal complements. The Old English data speaks against a GF POSS distinct from SUBJ, since the arguments for different degrees of semantic unrestrictedness demarcating the two GFs are founded on the association between POSS and possessors, rejected here. Moreover, the particular valency characteristics of ANCs can be reflected elsewhere than through a distinct nominal GF. The assessment given for the nominal GFs in the Old English noun phrase is similar to the proposals of Markantonatou (1995) for Modern Greek and Laczkó (2000) for Hungarian, insofar as a combination of a semantically unrestricted function POSSSUBJ and an infrequently instantiated, semantically more restricted function OBL_θ are used to account for all relevant noun phrases. However, the proposal for Old English is detached from formal realisation and both POSS/SUBJ and OBL_θ are freed from associations with possessors and prepositional phrases respectively. In this way, the account of nominal GFs falls into line with discussions of GFs at the level of the clause, where associations between GFs and specific form have had less dominance in the literature.

References

Allen, Cynthia L. 2008. *Genitives in Early English: Typology and Evidence*. Oxford: OUP.

Bresnan, Joan & Mugane, John. 2006. Agentive Nominalizations in Gĩkũyũ and the Theory of Mixed Categories. In Miriam Butt, Mary Dalrymple & Tracy H. King (eds.) *Intelligent Linguistic Architectures: Variations on themes by Ronald M. Kaplan*, 201–234. Stanford: CSLI Publications.

Chisarik, Erika & Payne, John. 2003. Modelling Possessor Constructions in LFG: English and Hungarian. In Miriam Butt & Tracy H. King (eds.), *Nominals Inside and Out*, 181–200. Stanford: CSLI Publications.

Comrie, Bernard & Thompson, Sandra A. 1985. Lexical Nominalization. In Timothy Shopen (ed.), *Language Typology and Syntactic Description*, 349–398. Cambridge: CUP.

Dalrymple, Mary, Lowe, John & Mycock, Louise. 2019. *The Oxford Reference Guide to Lexical Functional Grammar*. Oxford: OUP.

Dowty, David. 1991. Thematic Proto-Roles and Argument Selection. *Language* 67(3), 547–619.

Falk, Yehuda. 2001. Constituent Structure and Grammatical Functions in the Hebrew Action Nominal. In Miriam Butt & Tracy H. King (eds.), *Proceedings of the LFG01 Conference*, 181–200. Stanford: CSLI Publications.

Findlay, Jamie Y. 2020. Mapping Theory and the Anatomy of a Verbal Lexical Entry. In Miriam Butt & Ida Toivonen (eds.), *Proceedings of the LFG20 Conference, On-Line*, 127–147. CSLI Publications.

Grimshaw, Jane. *Argument Structure*. Cambridge, Mass.; London: MIT Press, 1990.

Kelling, Carmen. 2003. French Psych Verbs and Derived Nouns. In Miriam Butt & Tracy H. King (eds.), *Nominals Inside and Out*, 151–180. Stanford, CSLI Publications.

Koike, Takeshi. 2006. The History of the Genitive Case from Old English onwards.*English Language and Linguistics* 10(1), 49–75.

Laczkó, Tibor. 2000. Derived Nominals, Possessors, and Lexical Mapping Theory in Hungarian DPs. In Miriam Butt and Tracy H. King (eds.), *Argument Realization*, 189–227. Stanford: CSLI Publications.

Markantonatou, Stella. 1995. Modern Greek Deverbal Nominals: An LMT Approach. *Linguistics*, 31(1), 67–299.

Needham, Stephanie & Ida Toivonen. 2011. Derived Arguments. In Miriam Butt & Tracy H. King (eds.), *Proceedings of the LFG11 Conference*, 401–421. Stanford: CSLI Publications.

Randall, Beth. 2003. *CorpusSearch 2*. Philadelphia, PA: University of Pennsylvania.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Sadler, Louisa 2000. Noun Phrase Structure in Welsh. In Miriam Butt & Tracy H. King (eds.), *Argument Realization*, 73–109. Stanford: CSLI Publications.

Taylor, Ann, Anthony Warner, Susan Pintzuk & Frank Beths. 2003. *The York-Toronto-Helsinki Parsed Corpus of Old English Prose*. Available from the Oxford Text Archive.

Arguments and adjuncts across levels

Ida Toivonen
Carleton University

Proceedings of the LFG'21 Conference

On-Line

Miriam Butt, Jamie Y. Findlay, Ida Toivonen (Editors)

2021

CSLI Publications

pages 306–331

http://csli-publications.stanford.edu/LFG/2021

Keywords: argument-adjunct distinction, prototypical characteristics, instruments

Toivonen, Ida. 2021. Arguments and adjuncts across levels. In Butt, Miriam, Findlay, Jamie Y., & Toivonen, Ida (Eds.), *Proceedings of the LFG'21 Conference, On-Line*, 306–331. Stanford, CA: CSLI Publications.

Abstract

The distinction between arguments and adjuncts is useful and widely adopted. It is foundational to many formal approaches to grammar, including LFG. However, it is not always obvious whether a phrase should be classified as an argument or an adjunct. I propose that the multifaceted nature of language can explain why some elements seem to fall in between arguments and adjuncts. Arguments have a prototypical realization at each level of grammar and they are also typically core event participants of their predicate. However, there can be mismatches between levels, and arguments can display atypical characteristics at each level. The specifics of the proposal are formulated with reference to the different structures in LFG's parallel projection architecture.

1 Introduction

The distinction between arguments and adjuncts is fundamental to syntactic and semantic analysis. However, it has proven difficult to pinpoint an exact definition of argumenthood, and it is sometimes difficult to classify a phrase as a clear argument or a clear adjunct. I propose in this paper that the complications stem from mismatches between levels of analysis. Prototypical arguments are core event participants conceptually, occupy specifier or complement positions in the c-structure, carry core grammatical functions at f-structure, compose directly with the verb semantically rather than being predicates of events, are not marked with oblique cases or prepositions, etc. However, these characteristics do not always align, and this complicates the identification of arguments and adjuncts.

This paper is organized as follows: Section 2 reviews some of the reasons why linguists across theoretical frameworks have adopted the argument-adjunct distinction. After that, section 3 lists a number of well-documented problems with the argument-adjunct distinction. Section 4 proposes that the reason why it is sometimes hard to determine whether something is an argument is that there can be mismatches between levels of information. The proposal specifically makes use of the LFG parallel projection architecture. Uncontroversial arguments are "close" to the predicate at all levels of grammar and also conceptually, but conceptual core participants are not necessarily linguistic arguments. Also, there can be mismatches between grammatical levels which may lead to a situation where something is an argument (close to the predicate) at some levels of grammar but not others. Section 5 discusses some previous proposals on how to deal with problematic cases. The section is mainly devoted to the proposals of Arka (2014) and Rákosi (2006, 2012). Finally, section 6 offers some concluding remarks.

2 In defense of the argument-adjunct distinction

Arguments are selected by the verb, but adjuncts are not. Arguments have a closer relationship with the verb syntactically and semantically. In many cases, it is not difficult to identify the arguments and adjuncts in a sentence. Consider, for example, the following two sentences (from Condoravdi 2021), which contain both arguments and adjuncts:

(1) Last year in Rome on 15th March, *Brutus* stabbed *Caesar* in the forum with a knife at midday in front of a large crowd of onlookers.

(2) Last year in Germany, *one or two people* were mugged every couple of hours in a few hidden corners of campus every weekday in some of the more dangerous university towns.

The phrases in italics are uncontroversial arguments of the verbs *stab* and *mug*. The other, more peripheral, dependents are adjuncts, except possibly the instrument *with a knife* in (1), whose status is less clear. Example (2) also contains a passive verb. Should the unexpressed agent of *mug* count as an argument? If the passive agent had been expressed as a *by*-phrase, would it then be an argument or an adjunct? We return to instruments and passive agents later.

The examples illustrate that while some phrases may be difficult to classify categorically as arguments or adjuncts, many (I think most) phrases are in fact easy to classify. As linguists, we can quite freely talk about verbs as intransitive, transitive, or ditransitive without worrying too much about possible complications or misunderstandings: it is generally clear how many arguments a verb (or other predicate) takes. Similarly, when a verb is used in a sentence, it is typically clear which dependents are arguments and which are adjuncts, and I will not argue for a rejection of the argument-adjunct distinction in this paper.

2.1 Predicate arguments and predicate adjuncts

This section reviews some data that will serve as a reminder of the value of the argument-adjunct distinction. First we consider the contrast between predicate arguments (3) and seemingly similar predicate adjuncts (4). The examples in (3–4) are from Bresnan et al. (2016, 286):

(3) a. Mary didn't sound *ashamed of herself*.
b. Louise struck me *as a fool*.
c. Jogging keeps Susan *in a bad mood*.

(4) a. Mary looked down, *ashamed of herself*.
b. Louise enjoyed sports, naturally, *as a Southern Californian*.
c. Susan arrived for lunch, *in a bad mood as usual*.

Bresnan et al. (2016, 286–288) show that predicate complements differ from adjuncts in a number of ways (the discussion is also included in the first edition, Bresnan 2001). For example, omission of the argument results in ungrammaticality or a shift in meaning of the main verb (Bresnan et al., 2016, 287):

(5) a. ??Mary didn't sound.
b. Louise struck me. (different meaning than 3b)

However, the adjunct can be omitted freely, as the reader can test by omitting the predicate adjuncts in (4).

Another difference concerns predication. When a verb takes a predicate argument, it dictates what the subordinate predicate is predicated of. For example, the complement of *strike* is predicated of the subject (6a) and the complement of *regard* is predicated of the object (6b):

(6) a. Mary struck Fred as proud of herself/*himself.
b. Mary regards Fred as proud of himself/*herself.

Verbs do not impose such predication restrictions on adjuncts. Predicate adjuncts differ from complements in that they can in some cases be predicated of the subject (7a) or the object (7b):

(7) a. Mary struck Fred, proud of herself for doing so.
b. Mary struck Fred, so proud of himself for insulting her.

The examples in (7) make use of a reading of the verb *strike* that is different from the reading in (6a), and the subordinate predicate is an adjunct. Adjuncts are often predicated of the matrix subject regardless of what the matrix verb is, but it is also sometimes, like in (7b), possible for predicate adjuncts to be predicated of a non-subject. In sum, the matrix verb determines the interpretation of the subject of its predicate argument but not the interpretation of the subject of a predicate adjunct.

There are also other differences pointed out in Bresnan et al. (2016, 286–288): A predicate argument can host a negative polarity item but a predicate adjunct cannot. The ordering of arguments is fixed compared to the ordering of adjuncts. Each verb takes a unique predicate argument, while it is possible to include multiple predicate adjuncts with a similar role. Predicate arguments allow extraction more easily than predicate adjuncts.

In sum, traditional argumenthood tests yield stark contrasts in acceptability between predicate arguments and predicate adjuncts.

2.2 The adjunct condition

Adjuncts are *islands* in the sense of Ross (1967): they disallow certain kinds of linguistic material such as negative polarity items controlled from the matrix clause. They also disallow gaps, which is what we will focus on here: arguments permit extraction gaps more easily than adjuncts (Huang, 1982; Chomsky, 1986; Johnson, 2003). This generalization has been called *the adjunct condition* (for discussion of the adjunct condition in LFG, see Dalrymple et al. 2019, Ch. 17). The adjunct condition is one of the traditional argumenthood tests mentioned above. It will be considered in some detail here.

The examples in (8–9) illustrate that the adjunct condition governs extractability out of finite subordinate clauses in English. The subordinate clauses in (8) are arguments of *promise* and *hope*, respectively, and they contain gaps. The subordinate clauses in (9), on the other hand, are adjuncts, and the gaps render the examples ungrammatical.

(8) a. Which plants did you say Maria liked _?
    b. Who did Farrah hope that Kevin would marry _?

(9) a. *Who did you stay quiet so that Kevin would marry _?
    b. *Which cousin did Bill cry after he annoyed _?

There is strong support for the adjunct condition, but it is not completely unproblematic. Previous scholars have pointed to some examples where it is in fact possible to extract out of adjuncts (I present a few of those below). However, the counterexamples that have been identified constitute restricted subclasses of adjuncts and the condition otherwise holds. In other words, it seems that the adjunct condition predicts the majority of cases, but individual languages or dialects allow violations of the condition in specific constructions.

Counterexamples to the condition can be found in English non-finite clasuses. While extraction out of nonfinite subordinate adjunct clauses is typically blocked (e.g., (10a)), Borgonovo and Neeleman (2000), Truswell (2007, 2011), and others have pointed out that there are exceptions (e.g., (10b)):[1]

(10) a. *What did John appear whistling?
     b. What did John come home whistling?

Truswell (2007) shows that extraction out of nonfinite adjunct clauses is restricted to a small subset of cases. Specifically, he argues that extraction is only possible if the event denoted by the subordinate predicate is identified with an event position in the semantic representation of the matrix predicate.

Huhmarniemi's (2009, 2012) careful investigation of non-finite forms in Finnish shows that the adjunct condition generally holds in Finnish as well. This is illustrated by the contrast in grammaticality between (11) and (12) from Huhmarniemi (2009):[2]

(11) a. Pekka näki Merjan kirjoittamassa runoja.
     P.NOM saw.3SG M.ACC write.MA.INE poems.PART
     'Pekka saw Merja writing poems.'

[1]Example (10b) is from Borgonovo and Neeleman (2000) and (10a) is from Truswell (2007).

[2]Abbreviations used in glosses: ACC accusative, INE inessive, F feminine, M masculine, MA the third infinitive in Finnish, NOM nominative, OBJ objective case, OBV obviative, OM_2 non-affected object marker, PART partitive, PERF perfective, POSS possessive pronominal marker, REL relational, TI transitive inanimate, TS theme sign.

b. Mitä Pekka näki Merjan kirjoittamassa?
what.PART P.NOM saw.3SG M.ACC write.MA.INE
'What did Pekka see Merja write?'

(12) a. Pekka yllätti Merjan kirjoittamalla runoja.
P.NOM surprised.3SG M.ACC write.ADE poems.PART
'Pekka surprised Merja by writing poems.'

b. * Mitä Pekka yllätti Merjan kirjoittamalla?
what.PART P.NOM suprised.3SG M.ACC write.ADE

In (11), the non-finite verb *kirjoittaa* 'to write' heads an argument of the matrix verb, and an object gap is possible. By contrast, the non-finite *kirjoittaa* in (12) heads an adjunct, and the gap is not permitted.

Huhmarniemi (2009, 2012) discusses the A-infinitive, VA-infinitive, five kinds of MA-infinitives (two of which are illustrated in (11–12) above), rationale and temporal infinitives in Finnish. She concludes that "... when it can be established independently that the phrase occupies an adjunct position, then it is an extraction island" (Huhmarniemi, 2012, 236). The argument-adjunct distinction accounts for most of the Finnish infinitive data, but there are a few potential counterexamples. For example, about 30% of the participants in an experiment allowed extraction of objects (but not subjects or adjuncts) out of the non-finite -ESSA temporal construction "in specific contexts" (182). Like in English, the adjunct condition is a solid starting point for the exploration of gap permissibility in Finnish. The condition alone covers the vast majority of the relevant data, and the potential counterexamples belong to specific grammatical subclasses of adjuncts.

The adjunct condition governs extraction also beyond English and Finnish. For example, adjunct clauses are islands to *wh*-extraction in Norwegian (Kush et al., 2018), Italian (Sprouse et al., 2016), and Jordanian Arabic (Al-Aqarbeh and Sprouse, 2021). Stepanov (2007) presents a cross-linguistic review of the adjunct condition, and he concludes that no languages allow extractions out of adjuncts. Peripheral finite clauses seem to be strong islands in all languages that have been carefully investigated, but there is variation with respect to central adjuncts and non-finite adjuncts. We considered some violation examples from English and Finnish above, and more examples are provided by Müller (2019), who investigates the adjunct condition in Swedish (and other Scandinavian languages), where island effects generally are not as strong as in many other languages.

The brief review of findings provided here has focused on gaps in *clausal* arguments or adjuncts, but non-clausal dependents have also been investigated. Prepositional phrases, for example, are quite permissive in many languages including English. A fuller review will not be attempted here, but see, for example Falk (2009, 2011) for relevant discussion within LFG. Falk proposes that in order to explain island effects, it is necessary to take into account pragmatics and processing in addition to syntax. Hofmeister and Sag (2010) and Hofmeister et al. (2012a,b) explore the possibility that island constraints can be completely reduced to processing constraints related to discourse linking and cognitive complexity. However, the results of a growing number of studies indicate that island constraints cannot be reduced solely to processing (Sprouse et al., 2012a,b; Aldosari, 2015; Goodluck et al., 2017; Müller, 2019; Pham et al., 2020).

Taken together, the findings reviewed above indicate that the argument-adjunct distinction is a strong predictor of the permissibility of gaps. Careful investigation has pointed to circumscribed classes of counterexamples, which shows that the linguistic reality is complex, as is of course the case with all broad grammatical postulates. It is also important to keep in mind that the adjunct condition is not the only constraint on gaps (see Ross 1967 for more). Despite the complexities, the adjunct condition strongly supports the argument-adjunct distinction: the adjunct condition covers an impressive amount of data; data that would be left unexplained if the notion of argumenthood were abolished from grammatical theory.

2.3 Interim conclusion: Arguments differ from adjuncts

Section 2 is included here to serve as a reminder that there is strong support for the argument-adjunct distinction. First, the distinction deserves serious consideration because of its heritage. The idea that arguments have a distinct status has been assumed and argued for across scholarly traditions, sometimes independently. Grammarians and linguists in different time periods and endorsing a variety of theoretical perspectives adopt a distinction between arguments and adjuncts (see Barbu and Toivonen 2016a for a cross-theoretical overview). The intuition of argumenthood builds on centuries of work on language: the notion of direct dependents of the verb is implicitly assumed already in the works of Pānini (Dowty, 1991; Barbu, 2015).

Second, the classification of phrases as arguments and adjuncts is in many cases not at all controversial, as illustrated by the following example:

(13) In the evening, *the lively zebra* peacefully enjoyed *the sunset* in the valley.

In (13), *the lively zebras* and *the sunset* are uncontroversial arguments and the other phrases are not arguments.

Third, systematic comparison between a specific class of arguments and a similar class of adjuncts reveals that the groups differ from each other strikingly in a number of predictable ways. This was illustrated by the comparison of predicate arguments and adjuncts in section 2.1.

Fourth, it is possible to identify specific ways in which arguments and adjuncts differ cross-linguistically. The adjunct condition is an example (section 2.2). The argumenthood diagnostics are not necessarily universal, and the generalizations are often implicational: if a language has characteristic X, and X yields distinctions in grammaticality, then arguments will display one pattern and adjuncts another. These characteristics are used as argumenthood tests.

Fifth, there is ample psycholinguistic and neurolinguistic evidence for the distinction. For example, Di Giovanni (2016) performed an EEG study on well-formed and nonsensical arguments and adjuncts. The nonsense arguments correlated with an early left anterior negativity and an N400 signal, whereas the nonsense adjunct condition triggered a P600. Di Giovanni further found that the processing of arguments correlated with a strong decrease in alpha activity, whereas there was instead a slight increase in alpha-band power with adjuncts. Other psycho- and neurolinguistic studies supporting the argument-adjunct distinction include Shapiro et al. (1989); Britt (1994); Boland (2005); Boland and Blodgett (2006); Tutunjian and Boland (2008); Frisch et al. (2004); Thompson et al. (2007, 2010); and Lee and Thompson (2011).[3]

I conclude that the argument-adjunct distinction rests on solid ground.

3 Problematic aspects

There are strong reasons to adopt the argument-adjunct distinction in linguistic theory, but a number of problematic aspects need to be addressed. One problem is that there is no straightforward and universally agreed-upon definition of *argument*. Textbooks provide definitions that are good enough to convey the intuition behind the concept, but they also tend to point out that the definitions are not foolproof. The definitions also vary between textbooks. Haegeman (1994, 44) offers the following: "The arguments are the participants minimally involved in the activity or state expressed by the predicate." This definition is not identical to the one provided by Carnie (2006, 51): "The entities (which can be abstract) participating in the [predicate] relation are called arguments." Tallerman (2005, 98) includes several relevant characteristics in her definition: "Adjuncts are always optional, whereas complements are frequently obligatory. The difference between them is that a complement is a phrase which is *selected* by the head, and therefore has an especially close relationship with the head; adjuncts, on the other hand, are more like 'bolt-on' extra pieces of information and don't have a particularly close relationship with the head." These characterizations are useful, but they don't always serve to clearly isolate arguments. For example, it is not clear where these definitions leave the instrument and unexpressed passive agent of examples (1–2) at the beginning of this chapter.

Textbook authors often themselves point out that the the issue is complex. Kroeger (2004, 10), for example, remarks that "[t]his distinction between arguments and adjuncts is important, but not always easy to make."

3.1 Tricky cases

As remarked in section 2, many examples of arguments and adjuncts are uncontroversial. However, some cases are less straightforward. For example, numerous

[3] Some psycholinguistic studies specifically indicate that certain speakers judge certain types of phrases as argumentlike in some ways but adjunctlike in others. A few such studies are presented in section 3.1. This, I will argue, is in line with the general proposal of this paper: there can be mismatches between levels.

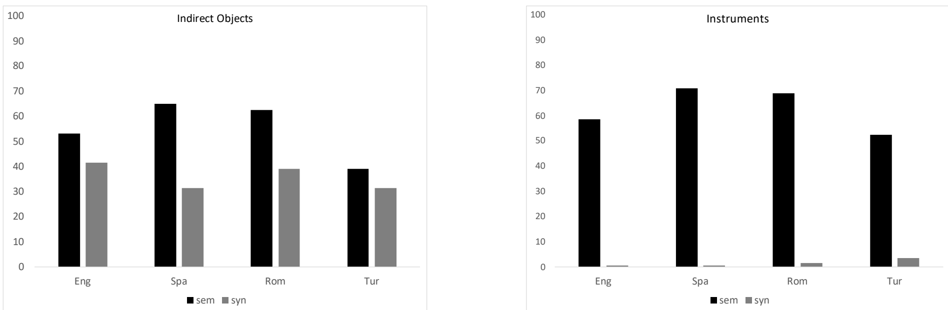
studies show that instrument phrases display characteristics of both arguments and adjuncts (Koenig et al., 2003; Donohue and Donohue, 2004; Tutunjian and Boland, 2008; Needham and Toivonen, 2011; Kifle, 2011; Rissman, 2013; Rissman et al., 2015; Barbu, 2015, 2020; Russo, 2021, a.o.). Example (1) above contains the instrument phrase *with a knife*. Another example is provided in (14) below:

(14) Frank wiped the table *with an old t-shirt*.

Roxana Barbu has conducted a series of studies designed to gauge intuitions on instruments (Barbu, 2015; Barbu and Toivonen, 2016b,a; Barbu, 2020). Barbu investigated English, Spanish, Romanian and Turkish, and her experiments involved two tasks. One task was designed to elicit intuitions about what event participants were necessary based on the meaning of verbs. The participants were provided with a list of verbs, and for each verb they were asked to specify the participants that were necessary in order for the event to take place. The other task was a sentence completion task, designed to elicit judgments about what phrases were necessary in the linguistic string. Specifically, participants specified what phrases had to be expressed in order for the sentence to sound complete. Barbu calls the first task *the semantic task* and the second one *the syntactic task*. For more details on her method, see Barbu (2020, Ch. 4).

The tasks in Barbu's web-based and anonymous studies were quite open-ended, and a certain amount of noise in the data is therefore expected. The results are nevertheless informative. Barbu included regular ditransitive verbs such as *send* and *deliver* in order to be able to compare indirect objects (uncontroversial arguments) to instruments. Figure 1 provides an overview of Barbu's results.

Figure 1: Mentions of indirect objects and instruments in Barbu's (2020) study



The chart on the left displays the proportion of times participants mentioned the indirect object when probed by a ditransitive verb. The chart on the right displays the proportion of times participants mentioned the instrument when probed by a verb that has been claimed to require an instrument (e.g., *draw, sweep, stab*).[4]

[4]Barbu also investigated verbs that have been claimed to allow but not require instruments. Those

Figure 1 separates the results by language. The results of the syntactic task are illustrated with black columns and the results of the semantic task are illustrated with grey columns.

The semantic task elicited more mentions than the syntactic task in general, but the difference between the two is much greater for instruments than indirect objects. In each language, participants mentioned instruments more than half of the time in the semantic task, but instruments were almost not mentioned at all in the syntactic task, even though the same verbs were included in both tasks. Barbu's results indicate that instruments are viewed as core participants of certain verbs, but they nevertheless do not need to be overtly expressed.

Russo (2021) applies standard argumenthood diagnostics to instrument phrases in English and Turkish. The results are summarized in Table 1, adapted from Russo (2021, 33).

Table 1: Argumenthood tests for English and Turkish (Russo, 2021)

Test	English	Turkish
Core participant	ARG/ADJ	ARG/ADJ
Iterativity	ARG/ADJ	ARG/ADJ
Alternation	ARG	ARG
Verb specificity	ARG	ARG
Optionality	ADJ	ADJ
VP anaphora	ADJ	ADJ
Pseudocleft	ADJ	ADJ

Russo's results are mixed: instruments are argument-like in some ways and adjunct-like in others. The notation ARG/ADJ indicates that instruments of some verbs (typically verbs that require instruments) pattern with arguments, and instruments of other verbs (typically verbs that allow but do not require instruments) pattern with adjuncts. It is interesting to note that the results are the same for English and Turkish.

The evidence from Barbu (2020), Russo (2021), and others shows that it is not obvious whether instruments should be classified as arguments or adjuncts. A number of other classes of phrases are also difficult to classify. Some examples from English include passive *by*-phrases, benefactive NPs, personal datives, result phrases, *with*-themes, and telic directional PPs. An example of each of these along with a reference to relevant work is given in (15–20):[5]

(15) The event was stopped *by the police*. (Kibort, 2004)

results are not included here.

[5]The constructions are discussed in the sources provided, but the examples are my own, except for (17) which is from Conroy (2007). Note that personal datives are not accepted in all varieties of English; see Wood and Zanuttini (2018).

Table 2: Argumenthood tests

optionality	core participant	word-order dependent meaning
alternations	verb specificity	weak island extraction
iterativity	VP anaphora	wh-word conjunction
VP ellipsis	fixed preposition	prepositional content
VP-preposing	relative ordering	VP-focussed pseudoclefts
“that happened”	the Adjunct Condition	

(16) Flory roasted *us* a chicken. (Toivonen, 2013)

(17) I’m gonna write *me* a letter to the president. (Wood and Zanuttini, 2018)

(18) Claudine beat the metal *flat*. (Christie, 2015)

(19) The garden swarmed *with bees*. (Lewis, 2004)

(20) Sandeep jumped *onto the platform*. (Van Luven, 2018)

3.2 Diagnostics

A large number of argumenthood diagnostics or tests have been proposed in the literature. Table 3 provides a list of many of them; see Van Luven and Toivonen (2018) for references and examples. The tests are useful, as illustrated in the discussion of the adjunct condition in section 2.2, but they have also been criticized. This section reviews some problematic aspects of a few of the tests.

Each argumenthood test is connected to characteristics that have been noted to align with arguments or adjuncts. For example, it has been observed that arguments tend to be obligatory while adjuncts are optional. This observation lies behind the *optionality* test: Phrases that can be omitted without rendering an example unacceptable are adjuncts, and phrases that cannot be omitted are arguments. However, this test does not work perfectly. Many verbs (e.g., *eat, write, drive*) take optional objects, for example, even though those objects are clearly arguments. Furthermore, many languages (e.g., Turkish, Vietnamese) allow the dropping of all or almost all arguments, given the right discourse context.

It has also been argued that not all adjuncts are optional (Jackendoff, 1990; Grimshaw and Vikner, 1993; Goldberg and Ackerman, 2001). For example, English middle constructions need adverbial modification to be acceptable:

(21) a. Cotton shirts iron *(easily).

Since arguments are not necessarily obligatory and adjuncts are not necessarily optional, the optionality test is problematic.

Another common test is the core participants test. Arguments are core participants of the verb, and adjuncts are more peripheral participants. This test captures the basic intuition behind argumenthood. However, some participants are core participants even though they seem to be adjuncts in other respects. Instruments, discussed above, constitute an example. Price phrases are similarly conceptually necessary for verbs like *buy, sell* and *rent*, even though they are not clear arguments. A buying event must involve a price otherwise it is a taking, trading or bartering event (Apresjan 1992). Conversely, expletives display many argument characteristics, but they are not core participants conceptually.

According to the VP-anaphora test, adjuncts may be added to 'do so' clauses, but arguments may not (Lakoff and Ross, 1966; Baker, 1978; Whaley, 1993):

(22) a. Nalini published a book in January and Joanne did so in February.
     b. *Nalini published a book and Joanne did so an article.

In this construction, *do* is a main verb, (Hankamer and Sag, 1976, fn. 27) and anything that can modify 'do so' is acceptable in the clause. In other words, this test is a test of what can modify 'do so' rather than a test of what arguments the main verb takes. The VP-focussed pseudocleft test and the 'do something' test similarly involve the main verb 'do'.

The argumenthood tests accurately distinguish between arguments and adjuncts in many cases. However, several of the tests are problematic, and every test needs to be applied with care. The adjunct condition serves as an example of this: the condition was presented in section 2.2 as a phenomenon that shows genuine sensitivity to the argument-adjunct distinction. However, it was also pointed out in that section that certain subclasses of phrases seem to escape the condition, and there are crosslinguistic differences. Almost all tests have been criticized by previous scholars, many of them referenced below in sections 3.3 and 5.

3.3 Section summary

Section 3.1 pointed out that some elements seem to display characteristics of both arguments and adjuncts. Section 3.2 showed that argumenthood tests often give unclear results. Such complications are widely acknowledged; see, e.g., Cennamo and Lenci (2019); Moura and Miliorini (2018); Andrason (2018); Ackema (2015); Hole (2015); Williams (2015); Forker (2014); Bosse et al. (2012); Hedberg and DeArmond (2009); Ágel and Fischer (2009); Koenig et al. (2003); Dowty (2003); Vater (1978). Difficulties with the argument-adjunct distinction have also been noted within the LFG literature: Rákosi (2006, 2012); Zaenen and Crouch (2009); Needham and Toivonen (2011); Kifle (2011); Arka (2014); Przepiórkowski (2016).

The complications have led some authors to conclude that the argument-adjunct distinction should be abolished from linguistic theory. Przepiórkowski (2016, 575), for example, calls the distinction "just another linguistic hoax". Other scholars have argued that the distinction between arguments and adjuncts is real and useful, but gradient (e.g., Forker 2014; Arka 2014). I return to a few previous proposals of how to deal with the tricky cases after I sketch my own proposal in section 4.

4 Arguments at different grammatical levels

Language is not monolithic. A sentence may be insightfully analyzed with focus on one or more of the following aspects: truth-conditional meaning, participant roles, grammatical functions, word classes, prosody, illocutionary force, etc. I propose that the multi-faceted nature of grammar explains why certain phrases are difficult to categorize as arguments or adjuncts.

In LFG, different facets of language are analyzed at distinct grammatical levels: c-structure, f-structure, a-structure, s-structure, and so on. This division of labour will be useful for modelling elements that do not seem to be clear arguments or adjuncts. An element can be argument-like at one level even though it is adjunct-like at another. This section goes through the notion of argumenthood at some of the relevant levels.

4.1 Conceptual event participants

Predicates correspond to events and states in the world, and speakers form mental representations of those events and states. This representation includes intuitions about the number and type of participants events require and allow. However, as pointed out by Levin and Rappaport Hovav (2005, 168) and Jackendoff (1990, 156), a participant can be associated with an event denoted by a verb without being a linguistic argument of that verb.

The maximum number of possible linguistic arguments is more restricted than the maximum number of conceptual event participants. The maximum number of linguistic arguments of a given predicate is typically assumed to be three or four, but the number of possible event participants can be higher.

Apresjan (1992) provides examples of verbs denoting events that take many participants ("actants"). For example, he lists the following five actants for the event denoted by the verb *lease*: he who leases, that which is leased, he from whom it is leased, that in exchange for which it is leased (i.e. the pay), and the period of time (for which it is leased). Apresjan concludes that "these actants are sufficient and necessary" (117). However, *lease* takes less than five linguistic arguments. This is an example of a mismatch between linguistic arguments and event participants. The price phrase of *lease*, for example, is a necessary conceptual event participant but not a linguistic argument. Other verbs that take price as an event participant but not an argument include *buy* and *rent*. On the other hand, *pay* and *cost* take price as both an event participant and a linguistic argument.

All (or almost all) events and states can be modified by location, time, and manner phrases. These are also not linguistic arguments, but they differ from price phrases in that they are not associated with the meaning of specific verbs. These general descriptors are not considered core conceptual event participants and are therefore excluded from the discussion here. This follows Koenig et al.'s (2003) "verb specificity" constraint for what is lexically encoded information.

Like price phrases, instrument phrases are necessary conceptual event participants but not linguistic arguments of many verbs. This was proposed in section 3.1 and is also argued by Rissman et al. (2015). A verb like *slice* denotes an event that cannot take place without an instrument. However, the verb does not take a linguistic instrument argument. Some verbs of course allow instruments as arguments. For example, in *the key opened the door*, the instrument is the subject argument of *open*. Also, in some languages instrument arguments can be added through applicativization. The Tigrinya examples in (23) (from Kifle 2011, 68–69) illustrate the applicativization of an instrument:

(23) a. Yonas bɨ-manka-y bäliʕ-u
Yonas spoon-POSS.1SG PERFS.eat-SM.3MSG
'Yonas ate with my spoon.'

b. Yonas n-ät-a manka-y
Yonas OBJ-DET-3FSG spoon-POSS.1SG
bäliʕ-u-la
PERFS.eat-SM.3MSG-OM$_2$.3FSG
'Yonas ate with my spoon.'

In (23a), the instrument 'spoon' is marked with the preposition *bɨ-*. In the applied version (23b), 'spoon' is an applied direct object and obligatorily indexed on the verb (Kifle, 2011, 11).

In sum, prices and instruments can in principle be arguments. However, they are often linguistic adjuncts, even when they appear with verbs that require them as necessary conceptual event participants.

4.2 Argument structure

The number and ranking of arguments of individual predicates are modelled at argument structure (a-s) in LFG. A-s is therefore the level which determines what the actual linguistic arguments are. If an element is listed on the a-s of a predicate, then it is an argument of that predicate. However, there might be disagreements among linguists about how to best analyze the a-s of a given predicate. It is also important to take into account that certain operations operate at a-s. For example, the standard LFG analysis of passives is formulated at a-s: the passive form of a verb is linked to one less argument than the active form. The highest argument of an active verb does not correspond to an argument of the passive form, but it can be expressed as a *by*-phrase, which is syntactically an adjunct.

A-s lists can also be augmented. This is the case for causatives and applicatives, for example: they are accompanied by one more argument than the basic form.

Applicative-like operations whereby arguments are added in a regular fashion are not always accompanied by special morphology. For example, English benefactive NPs (like in (16)) can be analyzed as optionally added arguments that correspond to a restricted class of benefactive *for*-adjuncts. The restriction seems to be that the added argument is interpreted as a recipient, and not just a benefactor in the broader sense; see Toivonen (2013) for references and discussion.

Similarly, personal datives (like in (17)) are added arguments. Personal datives are pronouns which are co-referential with the subject, but nevertheless do not appear in the reflexive form. Personal datives are restricted by grammatical constraints that differ between dialects. For example, some dialects allow second and third person in addition to first person personal datives. It also seems that some dialects allow PP personal datives.[6] Southern US dialects are generally more permissive than other English dialects with respect to personal datives. These generalizations are all from Wood and Zanuttini (2018), who list many attested examples.

The personal datives are similar to *affected experiencers* in German, which are discussed in Bosse et al. (2012) and given an LFG analysis in Arnold and Sadler (2012), where example (24) is from:

(24) Alex zerbrach mir Bens Vase.
Alex broke me Ben's vase
'Alex broke Ben's vase 'on me'.'

Arnold and Sadler (2012) provides an analysis of the interesting semantics of these elements. Syntactically, the affected experiencers are dative objects at f-structure and complements within the VP at c-structure (a regular object position). An affected experiencer is thus a syntactic argument, even though it is not a member of the basic a-s list of the verb. However, it can be viewed as an added a-s argument: a product of a regular a-s operation similar to applicativization.

4.3 Functional structure

Functional structure (f-s) functions are divided into argument functions (SUBJECT, OBJECT, OBJECT$_\theta$, OBLIQUE, COMP, XCOMP) and adjunct functions (ADJ, XADJ). However, an argument function of a verb does not necessarily correspond to elements that are arguments or argument-like at all levels. For example, raising-to-subject verbs like *seem* and raising-to-object verbs like *expect* have a SUBJECT and an OBJECT, respectively, that are not event participants or semantic arguments of the verbs. The embedded verb whose SUBJ or OBJ has raised shares that function with the raising verb at f-s, but the SUBJ/OBJ does not correspond to any elements in a c-structural subject or object position. Raising thus results in a mismatch between conceptual structure, s-s and f-s for the raising verbs and a mismatch between conceptual structure, s-s, c-s, and f-s for the embedded verbs.

East Cree relational morphology provides a striking example of an f-s argument that does not correspond in any obvious way to elements at other grammatical levels. The examples here are drawn from East Cree, but relational morphology is widespread in Algonquian. The examples in (25), from Junker and Toivonen (2015), illustrate the phenomenon:

[6] Wood and Zanuttini (2018) cite *I'm gonna go and play with me and cat* and other examples of PP personal datives.

(25) a. ni-wâpahte-n mistikw.
1-see.TI-1 wood
'I see a stick.'

b. ni-wâpahtam-**w**-â-n mistiku-yû.
1-see.TI-REL-TS-1 wood-OBV
'I see a stick (but she does not)/(over at her place).'

The relational morpheme *-w-* adds a third person animate participant to the interpretation of the sentence. The participant is often a possessor, but it can also be some other participant who is salient from the context, like in (25). Curiously, the participant cannot be expressed with an NP as a dependent of the verb. It can be expressed as a possessor embedded within an NP, but not as an NP dependent of the verb.

East Cree morphosyntax offers strong evidence that the introduced participant is an f-structure argument, specifically an OBJECT. First, when the verb carries relational morphology, the NPs in the sentence must be obviative. Cree has a requirement that at most one third person participant can be proximate, all others must be obviative. The fact that no NP can be proximate when relational morphology is present thus suggests that the introduced relational participant holds an argument function, and it further suggests that this participant is interpreted as being proximate (in the foreground). Second, relational verbs all have a transitive animate theme sign (TS). This theme sign indicates that the clause has an animate OBJECT.

East Cree thus allows for a certain class of clauses to have an f-structure OBJECT, even though that object is not a basic argument or core participant of the verb, and it cannot be overtly expressed as a dependent of the verb in the c-structure. The introduced element is an f-structure argument, but it is not an argument at c- or a-structure.

4.4 Constituent structure

Constituent structure (c-s) is key for the analysis of some elements that seem difficult to classify with respect to argumenthood. For example, expletives are not semantic arguments, but they occupy argument positions at c-s.

Prototypical arguments have specific realizations at c-structure: they are nominal, and they occupy specifier or complement positions. However, not all arguments are NPs. The sentence *Kamala lives in Toronto* contains a PP argument. It is also possible for adjuncts to be expressed as NPs:

(26) Frankie was surprised by his family *that evening*.

NPs are often arguments and PPs are often adjuncts, but this is a tendency and not a rule. It is possible to state generalizations about how arguments and adjuncts are typically realized in terms of c-structure categories (word classes), but mismatches are not uncommon.

In configurational languages, arguments are specifiers or complements at c-structure whereas adjuncts are adjoined (their mother and sister nodes are of the same category). However, non-configurational languages do not conform to these rules. In configurational language as well arguments can be expressed in other positions as dislocated topics or focussed phrases.

4.5 S-structure

The argument-adjunct distinction is foundational in many theoretical approaches to semantics, classical Montague Grammar, for example. Semantic structure is therefore likely to be crucial to the understanding of the argument-adjunct distinction, but exactly how depends in part on what semantic formalism is adopted. The LFG architecture is compatible with a variety of formalizations of meaning, which may require distinct conceptions of s-s, and the LFG community has not decided on one such formalization as the LFG standard. However, LFG+GLUE is emerging a common platform for semantic analysis in LFG (see Asudeh (To appear) for a recent overview). The argument-adjunct distinction has recently been tackled within LFG+GLUE by Asudeh and Giorgolo (2012); Asudeh et al. (2014); Lowe (2015); and Findlay (2016, 2020), who incorporate a-s into s-s and thereby eliminate a-s as a separate level of grammar. If this approach proves successful, then the analyses that were originally cast within Lexical-Mapping Theory at a-s will need to be revisited at s-s.

Several empirical puzzles that concern the argument-adjunct distinction have already been addressed from this perspective. Asudeh et al. (2014) discuss cognate objects such as *(sleep) a great sleep* and *(laugh) a terrible laugh*. Cognate objects are interesting because they are semantically like modifiers, even though they are direct objects syntactically. Cognate objects do not appear without modification (*great* and *terrible* in our examples). Also, the verbs that take cognate objects are not regular transitive verbs: most direct objects are not possible: **sleep the bed, *laugh a friend*. Asudeh et al. (2014) treat cognate objects as modifying adjuncts and not as arguments in the compositional semantics. However, they are nevertheless OBJECTs in the f-structure.

Asudeh and Giorgolo (2012) similarly make use of LFG's parallel levels in order to account for optional arguments of verbs such as *eat* and *read*. In their analysis, the omitted but understood objects are absent at c-s and f-s but present at s-s. This analysis is attractive and seems to work well for the examples that Asudeh and Giorgolo adress. However, it raises an interesting question: Should all entailed core participants be analyzed as s-s arguments? For example, do we want to posit that all verbs that require instruments (*sweep, slice*, etc.) have (possibly unexpressed) instrument arguments at s-s? Do we want to posit that verbs that take four or more conceptual participants (*buy, rent, dispatch, expatriate*, etc.) take four or more arguments at s-s? Asudeh and Giorgolo (2012) seem to limit their analysis to optional elements that are syntactic arguments when they are present in the syntax. Instruments of verbs like *sweep* and *chop* and price phrases of verbs like *sell* and *rent*, by contrast, are "understood" participants but they do not behave syntactically like arguments when they appear in the sentence.

4.6 Section summary

This section provided a brief overview of how arguments and adjuncts are represented at the different LFG levels. A-structure is central: this structure lists the elements that the researcher deem are the genuine linguistic arguments of each predicate. There are prototypical ways in which these arguments map to elements at other levels. Mismatches are not uncommon, as pointed out with reference to a number of examples in this section. These mismatches explain why it is sometimes difficult to identify a given participant as an argument or an adjunct.

5 Previous proposals

Many previous scholars have noted that it is not always easy to determine whether an element is an argument (see section 3.3 for references). I suggested above that difficult cases can be explained by taking the multifaceted nature of grammar into account. Several other interesting proposals for how to handle this issue have been put forward, and a few of them will be briefly reviewed here.

Zaenen and Crouch (2009) argue that all semantically marked obliques should be treated as adjuncts in ParGram, because they are computationally clunky to parse and they lead to too many ambiguities. However, they seem to imply that this is an interim solution, because they remark (p. 647): "It seems then that in the current state of affairs no linguistic theory is developed enough to give criteria that allow us to straightforwardly distinguish arguments from adjuncts in many cases. So, even in the cases where we can hope one day to make the distinction based on syntactic and lexical criteria we are not able to do it now."

Arka (2014) argues that there is no clear-cut argument-adjunct distinction. His claim is based on the observation that Balinese locatives that can undergo applicativization do not exactly correspond to locatives that would normally be classified as arguments. Example (27) (from Arka 2014) shows that some but not all OBLIQUE locatives (arguments) can undergo applicativization:

(27) Tiang ngentung-in anak-e ento / *kema lulu.
1 AV.throw-APPL person-DEF that to.there rubbish
'I threw rubbish to the person/there.'

Example (28) (also from Arka 2014) shows that some but not all ADJUNCT locatives can be applicativized:

(28) a. Tiang pules (di dampar-e / di alas-e)
1 sleep at bench-DEF in forest-DEF
'I slept on the bench / in the forest.'

b. Tiang mules-in dampar-e / ?*alas-e
1 AV.sleep-APPL bench-DEF forest-DEF
'I slept on the bench / ?*in the forest.'

Specific properties of the locative phrase, not the valency of the verb, determine whether it can appear as an applied object or not.

Based on these data, Arka (2014) concludes that the distinction between arguments and adjuncts is gradient. He proposes an *argument index*: A syntactic unit is assigned an argument index between 1 and 0, and the index is calculated based on 14 characteristics (6 general, 8 language-specific). An index value of 1.00 indicates "definitely a core argument" and an index of 0.00 indicates "definitely an adjunct". Arka further proposes that only locatives that receive a high *argument index* can be promoted to applied objects.

Arka (2014) shows that locatives that are nominal, affected by the event, specific, and individuated in space can be applicativized. According to Arka, this is because these factors grant them a high argument index value, which means they are more argumentlike than other locatives.

Arka's interesting proposal is an explicit attempt to account for the intuition that certain elements (e.g., nominals and participants directly affected by the event) are especially argumentlike, or suitable for argumenthood. In the proposal spelled out in section 4, this intuition would be captured less directly: at each level of grammar, elements can be realized in ways that are more or less compatible with argumenthood. An element can be argumentlike at all levels or adjunctlike at all levels. However, there may be mismatches. For example, a c-structure PP can be an argument, even though PPs are more commonly adjuncts than arguments.

Arka suggests that criteria for argumenthood can be specific to languages and even individual constructions. I propose instead that some of the factors he discusses are indeed language-specific, but they do not determine argumenthood but rather applicativization. In other words, the criteria Arka identifies are simply constraints on the applicativization of locatives in Balinese. Whether or not a locative can undergo applicativization does not depend on its argument index. Instead, it is determined by the factors that Arka convincingly argues are relevant: word class, specificity, individuation, and affectedness.

Rákosi (2006, 2012) analyzes adjuncts that seem somewhat argument-like. He specifically addresses circumstantial PPs such as instruments and benefactives. He proposes that these PPs are *thematic adjuncts*: adjuncts that receive a thematic role. He contrasts examples like the following (Rákosi, 2006):

(29) a. This appeals **to me**. THEMATIC ARGUMENT
b. This is important **to me**. THEMATIC ADJUNCT
c. **To me**, this is nice. NON-THEMATIC ADJUNCT

According to Rákosi, a participant PP such as *to me* in (29b) receives thematic specifications labeled with the features [+/-m, +/-c] of Reinhart (2002). The feature [m] indicates *mental state* and [c] indicates *cause change*.

It seems to me that our proposals are broadly compatible. Rákosi's theory is based on Reinhart (2002) whose theta system is intended as "the central system of the systems of concepts" (Reinhart, 2002, 229). The features are thus intended to cover conceptual structure, which is relevant for the linguistic system, but the two systems are nevertheless separate.

In my view, the proposal sketched here complements Rakosi's proposal rather than competes with it. Many phrases that Rákosi (2006, 2012) analyzes as thematic adjuncts would be treated here as core conceptual event participants that are not linguistic arguments. For example, that would be the analysis of *to me* in (29b), while *to me* in (29a) is an argument, and *to me* in (29c) is a regular adjunct that is not a core participant. Adopting Reinhart's system for a fuller understanding of the adjuncts would not in principle contradict my proposal. However, there are cases where my core conceptual event participants that are not linguistic arguments do not align with Rákosi's thematic adjuncts. For example, recall that some verbs (e.g., *cut*) require instruments as event participants whereas others merely allow them (e.g., *break*). Rákosi treats the required instruments as arguments and the allowed instruments as thematic adjuncts. I would treat them all as adjuncts, while recognizing that the required instruments are core event participants conceptually. Also, Rákosi treats comitative *with*-phrases in a sentence like *John cleaned the room with Kate* like thematic adjuncts, but this class of phrases does not align with core event participants. The difference seems to boil down to the fact that my proposal assumes that there is a distinction between adjuncts that denote necessary participants of the event/state denoted by the verb and other adjuncts. Rákosi's proposals does not seem to adopt this distinction.

6 Concluding remarks

The argument-adjunct distinction is foundational to many analyses. It is often easy to identify arguments, but some cases are less straightforward. I propose in this paper that the unclear cases can be explained by recognizing that there can be mismatches between linguistic levels and also between grammar and our general conceptualization of events that predicates refer to. The levels can be schematically described as in Table 3.

In LFG, the a-s representation determines which elements are arguments. However, the a-s analysis relies at least in part on information represented at other levels. Elements with argument functions at f-s are likely to correspond to a-s arguments. In English, elements that are nominal and specifiers or complements of the verb at c-s are likely to correspond to a-s arguments, but c-s representations can vary quite drastically between languages. Participants that are conceptually necessary event participants (entailed participants) are likely to correspond to arguments. Mismatches are not uncommon, and they complicate argumenthood judgments. However, careful analysis reveals that what might seem like gradience or uncertainty is in fact a reflection of the flexibility of mappings between levels.

Table 3: Levels

INTUITION	THEORETICAL	LFG
world knowledge	event participants	not linguistic
storage	initial argument list	lexicon, a-s
manipulation	altered argument list	a-s
syntactic info & relations	abstract syntax	f-s
expression	surface syntax	c-s
interpretation	semantics	s-s

In many cases, there are no mismatches. Consider a sentence like *The woman picked berries in the forest. The woman* and *berries* are straightforward arguments/argumentlike at all levels and *in the forest* is an adjunct at all levels.

Arguments and adjuncts can be compared to other linguistic concepts that are useful and widely adopted even though it might be difficult to pinpoint a definition that is universally accepted and clearly covers all and only the appropriate cases. Examples that come to mind are subjects, word classes like nouns and verbs, vowels/consonants, and tense/lax vowels. Even the basic notion *word* is difficult to define. My view is that these concepts are all based on important intuitions and it makes more sense to put effort into understanding what the intuitions reflect rather than rejecting the concepts altogether.

Acknowledgements

I want to thank Raj Singh, Katie Van Luven, Ash Asudeh, Jamie Findlay and an anonymous reviewer for providing insightful comments on a draft of this paper. I would also like to thank the organizers and participants of LFG21 for creating a supportive and constructive context for presenting this material. The questions and comments I received there were very helpful. The paper has been greatly improved by the input I have received, but all remaining shortcomings are solely my own.

References

Ackema, Peter. 2015. Arguments and adjuncts. In Tibor Kiss and Artemis Alexiadou (eds.), *Syntax: Theory and Analysis*, pages 246–274, Berlin: De Gruyter.

Ágel, Vilmos and Fischer, Klaus. 2009. Dependency Grammar and Valency Theory. In Bernd Heine and Heiko Narrog (eds.), *The Oxford Handbook of Linguistic Analysis*, pages 223–255, Oxford: Oxford University Press.

Al-Aqarbeh, Rania and Sprouse, Jon. 2021. Island effects and resumption in spoken Jordanian Arabic: an auditory acceptability judgment study, ms.

Aldosari, Saad. 2015. *The Role of Individual Differences in the Acceptability of Island Violations in Native and Non-native Speakers*. Ph. D.thesis, University of Kansas.

Andrason, Alexander. 2018. The argument-adjunct distinction: Applied nominal and locative phrases in Xhosa. *Linguistic Discovery* 16(2), 47–71.

Apresjan, Yuri D. 1992. *Lexical Semantics: User's guide to contemporary Russian vocabulary*. Ann Arbour: Karoma Publishers.

Arka, I Wayan. 2014. Locative-related roles and the argument-adjunct distinction in Balinese. *Linguistic Discovery* 12(2), 56–84.

Arnold, Doug and Sadler, Louisa. 2012. Affected experiencers and mixed semantics in LFG/Glue. In Miriam Butt and Tracy H. King (eds.), *Proceedings of the LFG12 Conference*, pages 1–20, Stanford, CA: CSLI Publications.

Asudeh, Ash. To appear. Glue Semantics. *Annual Review of Linguistics* 8, 1–21.

Asudeh, Ash and Giorgolo, Gianluca. 2012. Flexible Composition for Optional and Derived Arguments. In Miriam Butt and Tracy H. King (eds.), *Proceedings of the LFG12 Conference*, pages 64–84, Stanford, CA: CSLI Publications.

Asudeh, Ash, Giorgolo, Gianluca and Toivonen, Ida. 2014. Meaning and valency. In Miriam Butt and Tracy H. King (eds.), *Proceedings of the LFG14 Conference*, pages 66–88, Stanford, CA: CSLI Publications.

Baker, C.L. 1978. *Introduction to Generative-Transformational Syntax*. New York: Prentice-Hall.

Barbu, Roxana-Maria. 2015. *Verbs and participants: Non-linguists' intuitions*. Masters Thesis, Carleton University.

Barbu, Roxana-Maria. 2020. *On the psycholinguistics of argumenthood*. Ph. D.thesis, Carleton University.

Barbu, Roxana-Maria and Toivonen, Ida. 2016a. Arguments and adjuncts at the syntax-semantics interface. In Ellen Thompson (ed.), *Proceedings of FLYM3*, pages 1–12, Miami, FL.

Barbu, Roxana-Maria and Toivonen, Ida. 2016b. Event participants and linguistic arguments. In A. Papafragou, D. Grodner, D D. Mirman and J.C. Trueswell (eds.), *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*, pages 1961–1966, Austin, TX: Cognitive Science Society.

Boland, Julie E. 2005. Visual arguments. *Cognition* 95, 237–274.

Boland, Julie E. and Blodgett, Allison. 2006. Argument status and PP-attachment. *Journal of Psycholinguistic Research* 35, 385–403.

Borgonovo, Claudia and Neeleman, Ad. 2000. Transparent adjuncts. *Canadian Journal of Linguistics* 45, 199–224.

Bosse, Solveig, Bruening, Benjamin and Yamada, Masahiro. 2012. Affected experiencers. *Natural Language and Linguistic Theory* pages 1–35.

Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Oxford, UK: Blackwell.

Bresnan, Joan, Asudeh, Ash, Toivonen, Ida and Wechsler, Stephen. 2016. *Lexical-Functional Syntax, Second edition*. Hoboken, NJ: Wiley-Blackwell.

Britt, M. A. 1994. The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases. *Journal of Memory and Language* 33, 251–283.

Carnie, Andrew. 2006. *Syntax: A Generative Introduction*. Oxford: Blackwell.

Cennamo, Michela and Lenci, Alessandro. 2019. Gradience in subcategorization? Locative phrases with Italian verbs of motion. *Studia Linguistica* 73(2), 369–397.

Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: MIT Press.

Christie, Elizabeth. 2015. *The English Resultative*. Ph. D.thesis, Carleton University.

Dalrymple, Mary, Lowe, John J. and Mycock, Louise. 2019. *The Oxford Reference Guide to Lexical Functional Grammar*. Oxford: Oxford University Press.

Di Giovanni, Daniel. 2016. *Neural and predictive effects of verb argument structure*. Masters Thesis, Carleton University.

Donohue, Cathryn and Donohue, Mark. 2004. On the special status of instrumentals. In Miriam Butt and Tracy H. King (eds.), *Proceedings of the LFG04 Conference*, pages 209–225, Stanford, CA: CSLI Publications.

Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67(3), 547–619.

Dowty, David. 2003. The Dual Analysis of Adjuncts and Complements in Categorial Grammar. In Ewald Lang, Claudia Maienborn and Cathrine Fabricius-Hansen (eds.), *Modifying Adjuncts*, Berlin: Mouton de Gruyter.

Falk, Yehuda. 2009. Islands: a mixed analysis. In Miriam Butt and Tracy H. King (eds.), *Proceedings of the LFG09 Conference*, pages 261–281, Stanford, CA: CSLI.

Falk, Yehuda. 2011. Multiple-gap constructions. In Miriam Butt and Tracy H. King (eds.), *Proceedings of the LFG11 Conference*, pages 194–214, Stanford: CSLI.

Findlay, Jamie Y. 2016. Mapping theory without argument structure. *Journal of Language Modelling* 4(2), 293–338.

Findlay, Jamie Y. 2020. Mapping Theory and the anatomy of a verbal lexical entry. In Butt and Toivonen (eds.), *Proceedings of the LFG20 Conference*, pages 127–147, Stanford: CSLI.

Forker, Diana. 2014. A canonical approach to the argument/adjunct distinction. *Linguistic Discovery* 12(2), 27–40.

Frisch, S., Hahne, A. and Friederici, Angela. 2004. Word category and verb-argument structure information in the dynamics of parsing. *Cognition* 91(3), 191–219.

Goldberg, Adele and Ackerman, Farrell. 2001. The pragmatics of obligatory adjuncts. *Language* 77(4), 798–814.

Goodluck, Helen, Tsiwah, Frank and Saah, Kofi. 2017. Island constraints are not the result of sentence processing. In *Proceedings of the Linguistic Society of America 2*, pages 1–5.

Grimshaw, Jane and Vikner, Sten. 1993. Dependent indefinites. In Eric Reuland and Werner Abraham (eds.), *Knowledge and Language, Volume II: Lexical and*

conceptual structure, pages 143–155, Dordrecht: Kluwer.
Haegeman, Liliane. 1994. *Introduction to Government and Binding Theory*. Oxford: Blackwell.
Hankamer, Jorge and Sag, Ivan. 1976. Deep and surface anaphora. *Linguistic Inquiry* 7(3), 391–428.
Hedberg, Nancy and DeArmond, Richard C. 2009. On complements and adjuncts. *Snippets* 19, 11–12.
Hofmeister, Philip, Casasanto, Laura and Sag, Ivan. 2012a. How do individual cognitive differences relate to acceptability judgments? A reply to Sprouse, Wagers, & Phillips. *Language* 88(2), 390–400.
Hofmeister, Philip, Casasanto, Laura and Sag, Ivan. 2012b. Misapplying working memory tests: A reductio ad absurdum? *Language* 88(2), 408–409.
Hofmeister, Philip and Sag, Ivan. 2010. Cognitive constraints and island effects. *Language* 86(2), 366–415.
Hole, Daniel. 2015. Arguments and adjuncts. In Tibor Kiss and Artemis Alexiadou (eds.), *Syntax: Theory and Analysis. An International Handbook*, pages 1285–1321, Berlin and New York: De Gruyter Mouton.
Huang, C.-T. J. 1982. *Logical Relations in Chinese and the Theory of Grammar*. Ph. D.thesis, Massachusetts Institute of Technology, Cambridge, MA.
Huhmarniemi, Saara. 2009. Extraction islands in Finnish. *Biolinguistica Fennica Working papers* 1, 21–78.
Huhmarniemi, Saara. 2012. *Finnish A′-movement: Edges and Islands*. Ph. D.thesis, University of Helsinki, Helsinki, Finland.
Jackendoff, Ray. 1990. *Semantic Structures*. Cambridge, MA: MIT Press.
Johnson, Kyle. 2003. Towards an etiology of adjunct islands. *Nordlyd* 31(1), 187–215.
Junker, Marie-Odile and Toivonen, Ida. 2015. East Cree ghost participants. In Miriam Butt and Tracy H. King (eds.), *Proceedings of the LFG15 Conference*, pages 125–144, Stanford, CA: CSLI Publications.
Kibort, Anna. 2004. *Passive and passive-like constructions in English and Polish*. Ph. D.thesis, University of Cambridge.
Kifle, Nazareth Amlesom. 2011. *Tigrinya applicatives in lexical Functional Grammar*. Ph. D.thesis, University of Bergen.
Koenig, Jean-Pierre, Mauner, Gail and Bienvenue, Breton. 2003. Arguments for adjuncts. *Cognition* 89, 67–103.
Kroeger, Paul R. 2004. *Analyzing Syntax: A Lexical-functional Approach*. Cambridge: Cambridge University Press.
Kush, Dave, Lohndal, Terje and Sprouse, Jon. 2018. Investigating variation in island effects: A case study of Norwegian wh-extraction. *Natural Language and Linguistic Theory* 36(3), 743–779.
Lakoff, George and Ross, John R. 1966. Criterion for Verb Phrase Constituency, technical Report NSF-17, Aiken Computation Laboratory, Harvard University.
Lee, Jiyeon and Thompson, Cynthia K. 2011. Real-time production of arguments

and adjuncts in normal and agrammatic speakers. *Language and Cognitive Processes* 26(8), 985–1021.

Levin, Beth and Rappaport Hovav, Malka. 2005. *Argument Realization*. Cambridge: Cambridge University Press.

Lewis, Heather. 2004. *The with-phrase theme in English: Argument or adjunct?*. Masters Thesis, University of Canterbury.

Lowe, John. 2015. Complex predicates: an LFG+glue analysis. *Journal of Language Modelling* 3(2), 413–462.

Moura, Heronides and Miliorini, Rafaela. 2018. Toward a comprehension of an intuition: Criteria for distinguishing verbal arguments and adjuncts. *Alfa: Revista de Linguística* 62(3), 575–593.

Müller, Christiane. 2019. *Permeable islands: A contrastive study of Swedish and English adjunct clause extractions*. Ph. D.thesis, Lund University, Lund.

Needham, Stephanie and Toivonen, Ida. 2011. Derived arguments. In Miriam Butt and Tracy H. King (eds.), *Proceedings of the LFG11 Conference*, pages 401–421, Stanford, CA: CSLI Publications.

Pham, Catherine, Covey, Lauren, Gabriele, Alison, Aldosari, Saad and Fiorentino, Robert. 2020. Investigating the relationship between individual differences and island sensitivity. *Glossa: A Journal of General Linguistics* 5(1), 94, 1–17.

Przepiórkowski, Adam. 2016. How *not* to distinguish arguments from adjuncts in LFG. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy H. King and Stefan Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar*, pages 560–580, Stanford: CSLI.

Rákosi, György. 2006. On the need for a more refined approach to the argument-adjunct distinction: The case of dative experiencers. In Miriam Butt and Tracy H. King (eds.), *Proceedings of LFG06*, Stanford, CA: CSLI Publications.

Rákosi, György. 2012. Non-core participant PPs are adjuncts. In Miriam Butt and Tracy H. King (eds.), *Proceedings of the LFG12 Conference*, pages 524– 543, Stanford, CA: CSLI Publications.

Reinhart, Tanya. 2002. The theta system – an overview. *Theoretical Linguistics* 28, 229–290.

Rissman, Lilia. 2013. *Event participant representations and the instrumental role: A cross-linguistic study*. Ph. D.thesis, Johns Hopkins University, Baltimore.

Rissman, Lilia, Rawlins, Kyle and Landau, Barbara. 2015. Using instruments to understand argument structure for gradient representation. *Cognition* 142, 266–290.

Ross, John. 1967. *Constraints on Variables in Syntax*. Ph. D.thesis, Massachusetts Institute of Technology, Cambridge, MA.

Russo, Lara. 2021. *Arguments, adjuncts and instruments in English and Turkish*. Masters Thesis, Carleton University.

Shapiro, Lewis P., Zurif, Edgar and Grimshaw, Jane. 1989. Verb representation and sentence processing: Contextual impenetrability. *Journal of Psycholinguis-*

tic Research 18, 223–243.

Sprouse, Jon, Caponigro, Ivano, Greco, Ciro and Cecchetto, Carlo. 2016. Experimental syntax and the variation of island effects in English and Italian. *Natural Language and Linguistic Theory* 34, 307–344.

Sprouse, Jon, Wagers, Matt and Phillips, Colin. 2012a. A test of the relation between working memory capacity and syntactic island effects. *Language* 88(1), 82–123.

Sprouse, Jon, Wagers, Matt and Phillips, Colin. 2012b. Working-memory capacity and island effects: a reminder of the issues and the facts. *Language* 88(2), 401–407.

Stepanov, Arthur. 2007. The end of CED? Minimalism and extraction domains. *Syntax* 10(1), 80–126.

Tallerman, Maggie. 2005. *Understanding Syntax, 2nd ed*. London: Hodder Arnold.

Thompson, Cynthia K., Bonakdarpour, B and Fix, SC. 2010. Neural mechanisms of argument structure processing in agrammatic aphasia and healthy age-matched listeners. *Journal of Cognitive Neuroscience* 22(9), 1993–2011.

Thompson, Cynthia K., Bonakdarpour, B., Fix, SC, Blumenfeld, HK, Parrish, TB, Gitelman, DR and Mesulam, MM. 2007. Neural correlates of verb argument structure processing. *Journal of Cognitive Neuroscience* 19(11), 1753–1767.

Toivonen, Ida. 2013. English benefactive NPs. In Butt and King (eds.), *Proceedings of the LFG13 Conference*, pages 503–523, Stanford: CSLI.

Truswell, Robert. 2007. Extraction from adjuncts and the structure of events. *Lingua* 117, 1355–1377.

Truswell, Robert. 2011. *Events, Phrases and Questions*. Oxford Studies in Theoretical Linguistics, Oxford: Oxford University Press.

Tutunjian, Damon and Boland, Julie E. 2008. Do we need a distinction between arguments and adjuncts? Evidence from psycholinguistic studies of comprehension. *Language and Linguistic Compass* 2, 631–646.

Van Luven, Katie. 2018. *Pseudoclefts*. Masters Thesis, Carleton University.

Van Luven, Katie and Toivonen, Ida. 2018. Argumenthood tests: notes, manuscript in preparation, Carleton University.

Vater, Heinz. 1978. On the possibility of distinguishing between complements and adjuncts. In Werner Abraham (ed.), *Valence, Semantic Case, and Grammatical Relations: Workshop studies prepared for the 12th International Congress of Linguists*, pages 21–45, Amsterdam and Philadelphia: John Benjamins.

Whaley, Lindsay. 1993. *The Status of Obliques in Linguistic Theory*. Ph. D.thesis, State University of New York, Buffalo.

Williams, Alexander. 2015. *Arguments in syntax and semantics*. Cambridge: CUP.

Wood, Jim and Zanuttini, Raffaela. 2018. Datives, data and dialect syntax in American English. *Glossa* 2(1), 87, 1–22.

Zaenen, Annie and Crouch, Dick. 2009. OBLs hobble computations. In Miriam Butt and Tracy H. King (eds.), *Proceedings of the LFG09 Conference*, pages 644–654, Stanford, CA: CSLI Publications.

Perfective non-past in Modern Greek

Mark-Matthias Zymla
University of Konstanz

Alexandra Fiotaki
University of Ioannina

Proceedings of the LFG'21 Conference

On-Line

Miriam Butt, Jamie Y. Findlay, Ida Toivonen (Editors)

2021

CSLI Publications

pages 332–352

http://csli-publications.stanford.edu/LFG/2021

Keywords: tense, aspect, Greek, perfective non-past, pnp, dependent verb form, future shift, semantic annotation, parTMA, glue semantics, LiGER, description-by-analysis

Zymla, Mark-Matthias, & Fiotaki, Alexandra. 2021. Perfective non-past in Modern Greek. In Butt, Miriam, Findlay, Jamie Y., & Toivonen, Ida (Eds.), *Proceedings of the LFG'21 Conference, On-Line*, 332–352. Stanford, CA: CSLI Publications.

Abstract

This paper presents an analysis of the Greek tense/aspect system with a particular focus on the perfective non-past. It relies on the ParGram feature space for morphosyntactic analysis and the ParTMA semantic annotation scheme for semantic analysis. The whole analysis is computationally implemented, making use of the Xerox Linguistic Environment (XLE), the Glue Semantics Workbench (GSWB), and a new system for rewriting syntactic representations inspired by XLE's transfer system called LiGER.

1 Introduction

The present paper contributes to the large body of work providing computational resources for Lexical Functional Grammar building mainly on the international ParGram effort (Butt et al. 2002). More specifically, we present the continuation of the development of the Modern Greek (MG) ParGram grammar (Fiotaki and Tzortzi 2016) by adding a semantic interpretation component. This semantic resource is grounded in a description-by-analysis approach to Glue semantics and is implemented in the LiGER system (Linguistic Graph Expansion and Rewriting) that operates on top of the morphosyntactic analysis provided by the Xerox Linguistic Environment (XLE; Crouch et al. 2017).

In this paper, we present this system by focusing on the treatment of tense and aspect in the MG XLE grammar. The Greek tense/aspect system provides a complex picture that requires both syntactic and semantic analysis to be captured appropriately. Traditionally, the Greek verbal system is organized on the basis of tense (past/non-past) and aspect distinctions (perfective/imperfective; Holton et al. 1997 and Mozer 2009). The perfective non-past (henceforth PNP) plays a special role in this paradigm in that it is the only verb form that cannot occur freely in matrix clauses but requires certain licensors to be available. For this reason, it has been labeled 'dependent' in the literature (Holton et al. 1997). Consider the example in (1), where the PNP is ungrammatical. Unlike the imperfective non-past (INP) verb form, it needs to be licensed as shown here by virtue of the *na* (complementizer).

(1) O Christos *grap-sei/graf-ei ena gramma
DEF Chris write-*PNP.3SG/INP.3SG a letter
Chris is writing a letter.

(2) O Christos thimithik-e na grap-sei/graf-ei ena gramma
DEF Chris remember-3SG C write-PNP.3SG/INP.3SG a letter
Chris remembered to write/writing a letter.

The main goal of this work is to make a proposal for analyzing the PNP embedded in complement clauses introduced by *na* (Fiotaki and Lekakou 2018). The gist of

[†]We want to thank the audience of the 2021 LFG conference and the reviewers for their feedback. Furthermore, we are grateful to Katerina Kalouli for her helpful comments. This work has been funded by the Deutsche Forschungsgemeinschaft (DFG) within the project CUEPAQ, Grant Number 455910360, as part of the Priority Program "Robust Argumentation Machines (RATIO)" (SPP-1999).

this proposal is to distinguish between proper tenses and dependent tenses via the resources that they contribute to a Glue semantics derivation. While proper tenses introduce a tense operator and an evaluation time, dependent tenses (such as the PNP) pick up a temporal variable provided by some governing operator (similar to relative tenses). Crucially, the PNP never introduces an evaluation time itself, distinguishing it from other tense forms. This handling of resources is covered by semantic rewrite rules that produce Glue semantics meaning constructors, following description-by-analysis approaches (e.g., Andrews 2008). In the process, we refine the layered annotation proposed by the ParTMA annotation scheme (Zymla 2017, Zymla and Sulger 2017). More generally, this analysis follows ideas from Kusumoto (2005) and Giannakidou (2009) in terms of its semantic modeling.

The paper is organized as follows. §2 discusses the tense and aspect system in MG. In §3 the modeling of the semantics of tense and aspect is presented. §4 provides a detailed implementation of the morphosyntactic and semantic analysis of the examined verb type/PNP. We conclude and summarize our findings in §5.

2 Tense and aspect in Greek

In this section, we present the key observations of the MG tense/aspect system that provide the foundation for our syntactic and semantic analysis. We briefly describe the general paradigm but then quickly zoom in on the perfective non-past as the main point of interest.

2.1 The basic paradigm

In line with descriptive grammars of MG (Mackridge 1987, Holton et al. 1997, see also Mozer 2009), we take the Greek verbal system to be organized based on tense and aspect distinctions, namely past/non-past and perfective/imperfective.[1] The distinction between the perfective and the imperfective aspect appears morphologically in all grammatical verbal forms, whereas tense only appears as part of indicative clauses.

	perfective(`PE`)	imperfective(`IP`)
past(`PAST`)	e-grap-s-a	e-graf-a
non-past(`NON_PAST`)	grap-s-o	graf-o

Table 1: Modern Greek basic tense/aspect paradigm

In Table 1 the corresponding analysis of the Greek verb morphology is presented.[2] Let us first look at the lower left cell of the table: the imperfective non-past

[1]The PNP is historically related to the aorist subjunctive of Classical/Ancient Greek. Based on that, "traditionalists" maintain that MG has a morphological subjunctive. There are numerous reasons to reject this theory (see Tsangalidēs (1999) and (Sampanis 2012) for full discussion)

[2]For a more detailed discussion of Greek verbal morphosyntax see Mackridge (1987), Joseph

verb form is built of the stem *graf-* and the non-past tense marker (*-o*). In contrast, the imperfective past carries an unstressed (*e-*) before the stem *graf-* and the suffix (*-a*) to provide a past tense form. Similarly, perfective past carries an unstressed (*e-*) before the stem *graf-*and the tense marker (*-a*), but the latter is preceded by the marker for perfective aspect (*-s-*). Finally, the perfective non-past verb type is formed of the stem *graf-* followed by the perfective marker (*-s-*) and the non-past tense marker (*-o*).

2.2 The perfective non-past

The basic paradigm is straightforward; however, in this section, we show that the distribution of the perfective non-past indicates a more complex picture, where this particular form plays a special role that is not fully anticipated by the previously introduced 2x2 distinction.

2.2.1 The data

Our empirical investigation is based on data retrieved from the Hellenic National Corpus (henceforth HNC; http://hnc.ilsp.gr/), a balanced online monolingual corpus of MG texts developed by the Institute for Language and Speech Processing (ILSP). It currently contains approximately 50 million words and is constantly being updated (Hatzigeorgiu et al. 2000). There, we examined the occurrence of the sequence 'main verb + *na* + subordinate clause', extracting 7508 sentences. The general interpretation of the PNP in these sentences consists of future temporal reference and perfective viewpoint, i.e., a fully realized, episodic interpretation of the underlying eventuality (see also Fiotaki and Lekakou 2018).

The data suggests that the PNP only occurs in future shifted contexts. This will be crucial later, but first, we are going to examine its status as a 'dependent' verb form which is an important factor in modeling the syntax/semantics interface.

2.3 Tense and aspect of the PNP

The PNP has been called 'dependent' (Holton et al. 1997, Tsangalidēs 1999, Lekakou and Nilsen 2009, Giannakidou 2009), as, unlike the other finite verbal forms in MG, it cannot be used to form a full clause as shown in (1). It requires one of the following licensors: either the particle *na* which serves as a complementizer (see (2)), the future marker *tha* in (3), a optative/hortative marker (e.g. *as* shown in (4)), or other temporal connectives such as *prin* 'before' (Fiotaki and Lekakou 2018 and references cited therein). These are all operators that are able to shift forward the evaluation time of the verb they embed (directly or indirectly).

(3) Tha pei tin istoria ap'ekso
FUT say.PNP.3SG the story by heart
She will tell the story by heart.

and Smirniotopoulos (1993), Holton et al. (1997), and Ralli (2005) among others.

(4) As pei tin istoria ap'ekso
OPT say.PNP.3SG the story by heart
Let him tell the story by heart.

Example (5) illustrates the pattern 'main verb + embedding an INP', whereas the example (6) exemplifies the pattern we focus on in this paper: 'main verb + embedding a PNP'. In the first example, the event time is prior to the utterance time. The example (6) yields an event time that starts now and moves forward open-endedly, indicating a future shifted interpretation.

(5) Thimat-ai na leei tin istoria ap'ekso
remember-3SG C say.INP.3SG the story by heart
S/he remembers that she was telling the story by heart.

(6) Thimat-ai na pei tin istoria ap'ekso
remember-3SG C say.PNP.3SG the story by heart
S/he remembers to tell the story by heart.

As pointed out above and following Fiotaki and Lekakou (2018), the PNP is allowed in those embedded clauses in which the semantics of the main verb impose a future orientation of the embedded eventuality. Example (7) illustrates this point: the lexical semantics of the verb *iposxomai* 'promise' are compatible with a future-shifted complement as it is only possible to make a promise concerning the future. In comparison, verbs like *vlepo* 'see', *akuo* 'hear', *arxizo* 'start' disallow the PNP in the embedded clause as shown in (8)–(9). This is because they impose a simultaneous interpretation of their complement, or they force a habitual interpretation of the embedded eventuality. Thus, they are temporally or aspectually incompatible.

(7) Iposxethik-e na epistrep-sei ta xrhmata sintoma
Promise-3SG C return-PNP.3SG the money soon
S/he promised to return the money soon.

(8) Ton vlep-ei na *diasxi-sei/diasxiz-ei to potami
Him see-3SG C cross-*PNP/INP.3SG the river
S/he sees him crossing the river.

(9) Arxis-e na *pai-xei/paiz-ei podosfero eksi xronon
Start-3SG C play-*PNP/INP.3SG football six years
S/he started to play football when he was six years old .

In summary, we propose that the PNP introduces relative future temporal reference in embedded contexts. Although we cannot argue that the PNP itself constitutes a morphological subjunctive, it gets licensed by markers that are related to mood such as *na*.[3] In the next section, we focus on the temporal properties of the PNP and its interaction with the licensor *na*.

[3]Giannakidou (2009) mentions that *na* carries properties of both a subjunctive marker and a complementizer. Thus, the subjunctive flavor of the PNP is arguably a result of the interaction between it and *na*. An interesting result of this interaction is that PNP is preferred in non-veridical contexts.

3 Formal semantics for the PNP

Now that we have outlined the main assumptions of our analysis, we provide a formalization of mainly the semantic properties of the PNP (see Fiotaki and Tzortzi 2016 and section 4.1 for the basic morphosyntactic analysis and the revisions made for this paper). In this paper, we focus on the temporal properties of the PNP: first, its dependent nature with respect to temporal reference and, second, its obligatory future shift in relation to perfective aspect.

Following most research in the semantics of tense and aspect, we build our analysis on ideas from Reichenbach (1947) and its successors, as explained in the next section. Starting in section 3.2, the proposal is tailored towards integration in LFG. More concretely, we present a two-component semantic analysis that makes use of the ParTMA annotation scheme (Zymla and Sulger 2017, Zymla 2017). Section 3.2.1 describes semantic feature structures based on the ParTMA annotation scheme, which we call the ParTMA template. Mapping f-structures to these semantic features is the first part of our semantic analysis. Section 3.2.2 explains how this template can be interpreted, using a description-by-analysis approach to Glue semantics. The concrete implementation is presented in section 4.

3.1 The semantics of tense and aspect

Before proposing a compositional semantics for the PNP, let us first discuss the semantics on a more conceptual level. In general, temporal reference, i.e. the semantics we assume to be underlying morphosyntactic tense markers, is associated with locating a reference or topic time with respect to an evaluation time, usually the time of utterance or speech time (Reichenbach 1947).[4] The corresponding semantics are usually specified using time intervals (type *i*). However, pure interval semantics fall short when considering embedded contexts (for an elaborate discussion, see Kusumoto 1999). For this reason, we use situation semantics, i.e., entities of type s, to encode tense/aspect information, where situations are abstract entities with at least one property, e.g., world/time coordinates which are relevant for our analysis. Thus, we may sometimes call the reference time the reference situation.

Grammatical aspect, the semantic exponent of which we call viewpoint, encodes the (temporal) relation between the reference situation and the corresponding eventuality, where an eventuality is a situation describing a state or event and its participants (i.e., the information encoded in the predicate-argument structure). Viewpoint distinguishes between an external and an internal view of a given eventuality. More specifically, perfective viewpoint describes an eventuality as a whole, whereas imperfective viewpoint provides an internal viewpoint focusing on a specific part of the underlying eventuality (see, for example, Comrie (1976), Smith (2013)). In the next two sections, we focus on the semantic properties of the PNP.

[4]Reference time is the term used by Reichenbach, whereas the term topic time has been coined by Klein (1994), whose work is one of the works that started the *neo-Reichenbachian* tradition. We will stick to the former term for the sake of consistency.

3.1.1 The semantics of non-past temporal reference

Generally, temporal reference is encoded in terms of temporal constraints on situations using relations, such as the $\prec$-relation for precedence. Thus, the past tense is understood as a mechanism to temporally locate a situation before some evaluation time. On the most basic level, the non-past is the inverse: a mechanism that locates the reference situation in the present or the future. Thus, we need to get an understanding of these three basic tenses.

The semantics of past and future tense are often treated as temporal quantifiers. We employ the same approach, following Kusumoto (1999, 2005). These quantifiers simply ensure that the reference time is ordered appropriately with respect to the evaluation time, as exemplified for past temporal reference in (10-a). Conceptualizing the present tense is less straightforward. According to Abusch (1998), present tense is not necessarily constrained by some temporal relation, but rather denotes a temporal variable itself, which she calls n. This n usually but not necessarily corresponds to the evaluation or utterance situation in the temporal quantifiers for past and future temporal reference as well. Furthermore, tenses may remain uninterpretable in certain contexts. For example, Kusumoto (2005) assumes that, in complement clauses, a temporal quantifier is not obligatory. This means that such embedded tenses have two states: either they introduce a temporal quantifier or they are bound by a temporal quantifier higher up in the derivation. This explains the ambiguity that arises as part of the sequence-of-tense phenomenon (Abusch 1988, 1997, Grønn and von Stechow 2010)). In both cases, the tense of the embedded clause is interpreted relatively. Thus, tenses in complement clauses do generally not introduce an evaluation time.

What does this mean for MG non-past tense? First, we could assume that an ambiguity arises between a present interpretation and an interpretation of future temporal reference as initially suggested. Considering the discussion above, this would mean that there would be an ambiguity between two semantically different elements (a variable of type s and a quantifier of type $\langle st, st\rangle$, i.e., a modifier of sets of situations). However, this suggests challenges at the level of composition due to a type mismatch between present tense and future tense.

Giannakidou (2009) proposes a different interpretation of non-past: rather than denoting a disjunction of two different meanings, she suggests that, at least in the case of the PNP, non-past temporal reference denotes a time interval whose initial point is saturated by the evaluation time, and which extends infinitely into the future. Furthermore, she assumes that the non-past itself cannot introduce an evaluation time differentiating it from Kusumoto's tenses, thus making it dependent.

We adopt this view, but rather than treating non-past as a temporal variable in the same vein as Abusch's present tense and Giannakidou's non-past, we unify the treatment of all tenses to quantifiers to simplify the compositional process overall. This means tenses can occur in three forms. In (10-a), a typical temporal quantifier is shown. It is introduced together with an evaluation time, t^* of type s that is supposed to saturate its t λ-slot in matrix clauses (we ensure this later by matching the semantics with the corresponding Glue semantics resources). Example (10-b) illustrates the temporal quantifier associated with non-pastness based on Giannakidou's (2009) proposal: there exists some situation t' that starts at the evaluation time provided by the t λ-slot, and that extends infinitely into the future. This type of tense is defective in the sense that it can not provide an evaluation time, which distinguishes it from (10-a). As mentioned above, a tense marker might remain uninterpretable. This is, for example, the case when it is governed by another temporal operator of the same kind (Kusumoto 2005). In this case, such a variable is abstracted over to combine with an element higher up in the derivation. The abstraction step falls out naturally if we use a Glue semantics framework. This means that it does not need to be stipulated as a separate step in the semantic composition but surfaces in terms of different handling of the corresponding resources (cf. Kusumoto 2005). The next step is to encode these three different behaviors of tenses in a semantic feature structure suitable for a Glue-style composition. This is done in section 3.2.1. Before that, let us briefly discuss the semantics of viewpoint.

(10) a. **past:** $\lambda P_{<s,t>} \lambda t_s . \exists t'_s [t \prec t' \wedge P(t')]$
b. **non-past:** $\lambda P_{<s,t>} \lambda t_s . \exists t'_s . [t' = i(t, \infty) \wedge P(t')]$

3.1.2 The semantics of perfective viewpoint

Other than the non-past component of the perfective non-past, the aspectual component behaves more or less as expected in that it describes an episodic occurrence of an eventuality. As Giannakidou (2009) notes, when the perfective is applied to an eventuality that is a state, a meaning shift is induced that makes the underlying state eventive. Thus, we can generally assume that perfective viewpoint describes a situation in which the underlying eventuality is fully realized, i.e., the eventuality is fully contained within the reference time or situation provided by temporal reference (Comrie 1976, Bohnemeyer and Swift 2004). We model this as an existential quantifier over situations, which takes as arguments the corresponding eventuality and relates it to the reference time via a part-of relation. This is exemplified in (11-a). Conversely, imperfective viewpoint is encoded as a universal quantifier over situations following classical modal semantics for imperfective aspect and, in particular, progressive aspect, e.g., Dowty (1977). The restrictor of the quantifier encodes the flavor of imperfectivity, mainly distinguishing between a progressive and a habitual interpretation (Arregui et al. 2011, 2014).

(11) a. **perfective:** $\lambda P_{<s,t>} \lambda s_s . \exists s'_s [s' < s \wedge P(s')]$
b. **imperfective:** $\lambda P_{<s,t>} \lambda s_s . \forall s'_s . [MB_{prog/hab}(s', s) \rightarrow P(s')]$

Combining perfective aspect with present tense forces a future-shifted interpretation in many languages, e.g., Slavic languages and Urdu (De Wit 2016). This is compatible with the tenses we have specified above since the reference situation specified by non-past temporal reference extends into the future, ruling out a simultaneous interpretation of the corresponding eventuality.

3.2 A two-component semantic analysis for tense and aspect

Now that we have discussed the general properties of the tense/aspect semantics we presuppose for PNP, it is time to flesh out the formalization within a compositional framework. Concretely, we want to translate insights from the works cited above into a Glue semantics analysis. However, first, we explore the compositional process we envision more generally.

We have outlined above that tense applies on top of aspect and aspect applies to a given eventuality. This is formalized in a type-driven composition in Kratzer (1998). The order of composition proposed there is shown in Figure 1. The VP is modeled as a set of events that is bound by aspect (viewpoint following the present conventions). Viewpoint returns a set of intervals which, in turn, combines with semantic tense (temporal reference). The result is a set of time intervals.

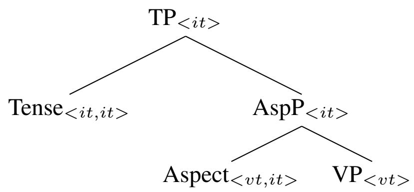


Figure 1: Compositional framework for tense and aspect

Since the present framework uses situations to model both tense and aspect (rather than time intervals and eventualities), the order of application needs to be restricted in another way than semantic typing. We solve this issue by using typing via Glue semantics which encodes both structural and type information at the same time based on the Curry-Howard isomorphism (Curry et al. 1958). The underlying structure for the composition is provided by the ParTMA annotation scheme.

3.2.1 The ParTMA annotation scheme

The parTMA annotation scheme provides an LFG-inspired annotation template that can be implemented in LFG's projection architecture as part of the semantic structure.[5] The ParTMA template contains three different components: i) a temporal reference component, ii) a viewpoint aspect component, and iii) a component for encoding lexical aspect. We focus on the first two components.

The first part of the two-component semantic analysis is a set of rules that generates a fully annotated ParTMA template from an f-structure input, i.e., the rules specify the syntax/semantics interface on a feature level (Zymla and Sulger 2017, Zymla 2017). As in the description-by-analysis tradition in Glue semantics, these rules match f-structure inputs and introduce additional information accordingly (Crouch and King (2006), Crouch (2005) for computational approaches and, e.g., Andrews et al. (2007), Andrews (2008) for theoretical discussion).

[5]The template can be structurally aligned with different morphosyntactic inputs, while the internal structure is preserved to cater to the specific needs of tense/aspect semantics.

$$\boxed{ID}\begin{bmatrix} \text{LEX-ASP} & \begin{bmatrix} \ldots \end{bmatrix} \\ \text{TEMP-REF} & \begin{bmatrix} \textbf{T-REF} ::= \text{'PAST'} \mid \text{'PRESENT'} \mid \text{'FUTURE'} \mid \\ \text{'NON-PAST'} \mid \text{'NON-PRESENT'} \mid \text{'UNSPEC'} \\ \textbf{T-RESTR} ::= \text{'IMMEDIATE'} \mid \text{'RECENT'} \mid \text{'NON-RECENT'} \\ \textbf{EVAL} ::= \text{'FID'} \end{bmatrix} \\ \text{VIEWPOINT} & \begin{bmatrix} \textbf{ASPECT} ::= \text{'PERFECTIVE'} \mid \text{'IMPERFECTIVE'} \\ \textbf{RESTR} ::= \text{'PROGRESSIVE'} \mid \text{'HABITUAL'} \mid \text{'ONGOING'} \mid \\ \text{'TERMINATED'} \mid \text{'CULMINATED'} \end{bmatrix} \end{bmatrix}$$

Figure 2: ParTMA eventuality template

Figure 2 presents the relevant part of the ParTMA template. As shown there, temporal reference specifies not only the temporal relation but also the evaluation time. As discussed in section 3.1.1, this is crucial for distinguishing relative tenses from absolute tenses and also plays a role in semantic phenomena not discussed in this paper, such as sequence-of-tense and double-access readings.

Mapping f-structure features onto these semantic features is relatively straightforward. In the next section, we explain how to use description-by-analysis methods to interpret the ParTMA annotation scheme while preserving the more hierarchical structure proposed in Figure 1 at the beginning of the section.

3.2.2 Semantic interpretation of the ParTMA annotation scheme

The ParTMA annotation template is interpreted via a description-by-analysis Glue semantics component that assigns meaning constructors to the feature structures introduced in the previous section. The goal is to make use of the compositional hierarchy proposed in section 3.2 to derive the semantics of tense and aspect.

Traditionally, in LFG, tense and aspect are treated as modifiers of a clause that appear as functional features in the f-structure (Butt et al. 1999). This is also the treatment generally proposed in description-by-analysis analyses (e.g., Andrews 2008). Haug (2008) proposes a different approach, distinguishing between modification of eventuality time and reference time, following Klein (1994). The architecture presented in this paper is similar but more fine-grained.

As shown in Figure 3 and Figure 4, the eventuality is specified at the level of `LEX-ASP` and is then related to the reference situation by viewpoint and finally to the situation of evaluation by temporal reference. Instead of encoding the hierarchy in terms of different semantic types as shown in section 3.2, we use linear logic to guide the composition. Together with the semantics specified in section 3.1.1 and 3.1.2 we can thus produce meaning constructors allowing for a Glue semantics calculation of the tree in Figure 3. Our goal of following the composition order in Figure 1 is thus achieved: We have mapped the relatively flat structure of the

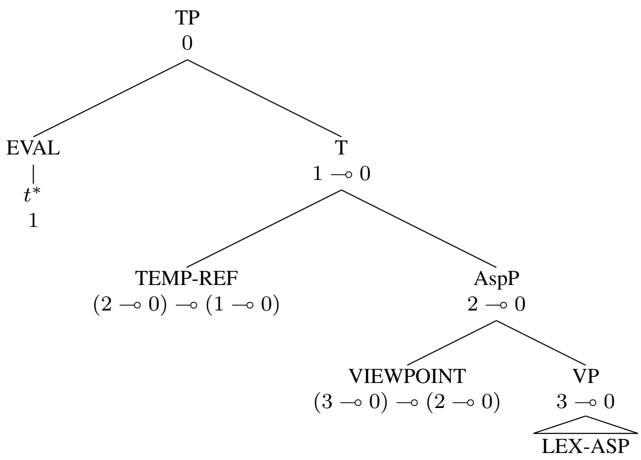


Figure 3: Order of composition of the ParTMA features

$$
{}_0\begin{bmatrix}
\text{TEMP-REF} & {}_2\begin{bmatrix} \ldots & \\ \text{EVAL} & {}_1\big[\ldots\big] \end{bmatrix} \\
\text{VIEWPOINT} & {}_3\big[\ldots\big] \\
\text{LEX-ASP} & {}_4\big[\ldots\big]
\end{bmatrix}
$$

Figure 4: Indexation of the ParTMA template

ParTMA template onto a tree structure for the sake of the underlying composition.

This has another reason in addition to following the proposal made by Kratzer (1998). This reason is technical: since the logic that guides the composition is inherently commutative (Asudeh 2012), glue approaches to compositional semantics suffer from abundant spurious ambiguity, in particular, in the computational domain. By encoding structural constraints within the Glue side of meaning constructors, commutativity is generally preserved while constraining the resulting combinatory explosion in areas where it is not required.

However, one of the disadvantages of this system is that features can not remain unspecified, since it would break the chain of composition that percolates all semantic information to the top node of the tree. Thus, the semantic interpretation component of the ParTMA template interprets unspecified features as identity functions that simply pass semantic information up the tree. This is not a very costly processing step. The benefits of avoiding combinatory explosion outweigh the cost of this additional step.

4 Implementation

In this section, we illustrate parts of the implementation of the analysis described above within the XLE and the LiGER system for modifying XLE's syntactic output. First, we will briefly recap the morphosyntactic modeling of the PNP within the XLE grammar and the resulting f-structures in the next section, then we explain the implementation of the two-component semantic analysis in section 4.2.

4.1 Morphosyntax in the Greek XLE grammar

The current version of the fragment of Modern Greek presents progress on the MG XLE grammar (Fiotaki and Tzortzi 2016), in particular with respect to the treatment of tense and aspect (TA) by adopting the ParGram TAM (Tense/Aspect/-Mood) scheme (Butt et al. 2002). As discussed in section 2.1, in MG there are four verb forms that are annotated for tense and aspect. The annotation scheme used is exemplified in the INP lexical entry *paizei* 'plays' presented in Figure 5.

```
paizei V * @(OPT-TRANS PAIZW)
            (^ TNS-ASP TENSE)=NON_PAST
            (^ TNS-ASP ASPECT)=IP
            (^ TNS-ASP MOOD)= indicative
           @(TRANSL play)
           @(PERS 3)
           @(NUM SG).
```

Figure 5: Lexical entry: *paizei*

A lexical entry encoding the PNP for the verb *paizw* 'play' is given in Figure 6. It is annotated for the tense and aspect features, but also for its inability to occur on its own in matrix contexts, with the feature `DEPENDENT` and the value `YES`. This feature is an artifact. It was implemented based on the descriptive analysis discussed in section 2.1 and is used as a stipulation to avoid over-generation of the syntactic component with respect to the PNP. However, it is not required for the semantic analysis. More concretely, the semantic annotation of the PNP proposed in this paper rules out the same parses in a syntactic grammar that does not contain the feature `DEPENDENT` (see section 4.3).

```
paiksei V *  @(OPT-TRANS PAIZW)
              (^ TNS-ASP TENSE)=NON_PAST
              (^ TNS-ASP ASPECT)=PE
              (^ TNS-ASP MOOD)= indicative
             @(DEPENDENT YES)
             @(TRANSL play)
             @(PERS 3)
             @(NUM SG).
```

Figure 6: Lexical entry: *paiksei*

A slightly simplified sample f-structure analysis is given in Figure 7 for the sentence ‘I don’t think that he will win the race.’ The main verb in the presented example is the verb *pistevo* ‘believe’. The particle *na* is treated as a complementizer (see Roussou (2000) among others) and it is encoded in the c-structure and surfaces in the f-structure by virtue of the `COMP-FORM` feature with the value *na*. The tense/aspect information for the PNP is given under the attribute `TNS-ASP`.

$$
1\left[\begin{array}{ll}
\multicolumn{2}{l}{\text{PRED `PISTEVO<[1-SUBJ:PRO],[5:KERDISEI]>'}} \\
\text{SUBJ} & \left[\text{PRED `PRO'}\right] \\
\text{COMP} & 5\left[\begin{array}{ll}
\multicolumn{2}{l}{\text{PRED `KERDIZO<[5-SUBJ:PRO],[10:AGONAS]>}} \\
\text{SUBJ} & \left[\text{PRED `PRO'}\right] \\
\text{OBJ} & \left[\text{PRED `AGONAS'}\right] \\
\text{TNS-ASP} & \left[\begin{array}{l}\textbf{ASPECT PE, TENSE NON\_PAST,} \\ \text{MOOD INDICATIVE}\end{array}\right] \\
\multicolumn{2}{l}{\textbf{COMP-FORM } \text{NA}\textbf{, DEPENDENT YES}, \ldots}
\end{array}\right] \\
\text{TNS-ASP} & \left[\begin{array}{l}\textbf{ASPECT IP, TENSE NON\_PAST,} \\ \text{MOOD INDICATIVE}\end{array}\right] \\
\multicolumn{2}{l}{\text{CLAUSE-TYPE DECL, NE\_FORM DE, \ldots}}
\end{array}\right]
$$

Figure 7: f-structure: *I don’t think he will win the race.*

4.2 Implementation of the semantics

The two-component semantic analysis approach for handling tense/aspect in MG presented in 3.2.2 is implemented in a system called LiGER (Linguistic Graph Expansion and Rewriting).[6] The system is inspired by the XLE’s transfer system, which has proven to be quite versatile. For example, it has been used to implement a semantic parser (Crouch 2005, Crouch and King 2006) and a reasoning engine (Bobrow et al. 2007), beyond its initially envisioned use as a system for machine translation (Frank 1999).[7] It is grounded in the wish to make linguistic annotation resources more cross-compatible. More concretely, the goal is to use a uniform graph format for linguistic annotations as inspired by Ide and Bunt (2010) and, more generally, the efforts concerned with interoperable annotation schemes.

The general architecture of the system presented in this paper is illustrated in Figure 8. The XLE is used to produce a morpho-syntactically annotated treebank (see section 4.1). The resulting parses are then accessed by the LiGER system and rewritten one by one, adding a semantic graph structure and the meaning constructors for the semantic derivation as leaves of this structure. These are given to the

[6]https://github.com/Mmaz1988/abstract-syntax-annotator-web

[7]Unfortunately, the transfer system is not supported by newer versions of the XLE and, thus, by more recent efforts concerning the XLE.

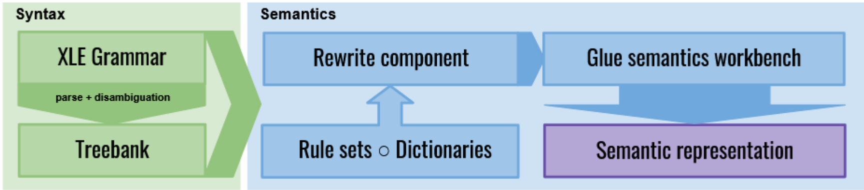


Figure 8: Annotation pipeline for DBA semantics

Glue Semantics Workbench (GSWB) to calculate the Glue derivation and produce the final semantic representation (Meßmer and Zymla 2018).[8] The GSWB provides a framework for computational work in Glue semantics, offering a modular structure to work with different meaning languages and linear logic provers.[9]

4.3 Description-by-analysis in LiGER

As described in section 3.2, we consider two different sets of rules for deriving the semantic analysis. One set establishes the semantic structure provided by the ParTMA template and the second set of rules "interprets" the semantic structure to produce the corresponding meaning constructors. We, therefore, call the first set semantic construction rules and the second set semantic interpretation rules.

In LiGER, these are encoded in terms of an ordered set of rules with rewriting capabilities. As discussed in section 3.2.2, this is because all features are required to be specified even if their value is unspecified The special role of features with the value `unspecified` is to introduce identity functions that raise semantic resources within the derivation tree in Figure 3. This is similar to manager resources used in, e.g., Haug (2008), Asudeh (2004).

Furthermore, as shown in Zymla and Sulger (2017), certain kinds of meaning shifts can be encoded in terms of layered interpretations of the respective features. This layering is encoded within the rewrite rules. Where applicable, first, the default interpretation rule applies, and if the semantics require it, the default value is rewritten into the appropriate value.

4.4 Semantic construction

Rules consist of a left-hand side and a right-hand side separated by an arrow ('==>'). The syntax that is used to specify the two sides is inspired by the INESS query language (Rosén et al. 2012). Hash signs in combination with alphanumeric values serve as variables over f-structure, or more generally, graph nodes (the f-structure is in principle a directed (acyclic) graph). Nodes may be connected via

[8] https://github.com/Mmaz1988/GlueSemWorkbench_v2

[9] See, for example, XLE+Glue which allows meaning constructors to be defined directly within XLE's output representations, while working with different meaning languages (Dalrymple et al. 2020).

relations or labeled edges from a graph perspective. The right-hand side of a rule expands a matching graph with the nodes and edges it specifies. Thus, the rule in (12) initializes the feature `T-REF` (for temporal reference) as `undefined` for all `TNS-ASP` nodes, ensuring that all such nodes are interpreted. The ParTMA template is, thus, constructed under the `SEM` attribute, which is mapped onto an unused variable (i.e., a variable that does not occur on the left-hand side). The `SEM` relation can be understood as a mapping to a semantic structure.

(12)
```
#a TNS-ASP #b
==> #a SEM #c & #c TEMP-REF #d &
#d T-REF 'undefined'.
```

Rule (13) is crucial for interpreting the PNP. As shown there, this rule does not introduce an evaluation time. Compare this to rule (14) for the INP in (14) which does so.[10] Note that both of these rules rewrite the initially provided default value. The differences between interpreting the PNP and the INP are subtle since both need to check for potential governing operators. The difference is that the INP may align its evaluation time with a governing operator, whereas the PNP is required do so, since it does not itself have the potential to be saturated by an external evaluation time, i.e., the speech time.

(13)
```
#a TNS-ASP #b TENSE 'NON_PAST' & #b ASPECT 'PE' &
#a SEM #c TEMP-REF #d & #d T-REF 'undefined'
==>#d T-REF 'non-past'.
```

(14)
```
#a TNS-ASP #b TENSE 'NON_PAST' & #b ASPECT 'IP' &
#a SEM #c TEMP-REF #d & #d T-REF 'undefined'
==>#d T-REF 'non-past' & #d EVAL #e & #e CHECK '-'.
```

The rule in (15) serves to account for the lacking evaluation time in the PNP rule. It is one example of a rule that searches for an appropriate licensor of the PNP, which can provide an evaluation time. In other words, the temporal reference of the PNP (`non-past`) is only fully specified in the context of a proper licensor (here a complementizer as indicated by the `COMP` relation. The licensor is searched for via inside-out functional application as indicated by the ^-symbol.)[11] The necessity of an evaluation time to interpret temporal reference is shown later in rule (18-a). From a resource perspective, the evaluation time is co-described with the temporal variable of the matrix verb, leading to the desired relative interpretation.[12] Con-

[10]The `CHECK` feature is used to distinguish between bound and unbound occurrences of tense operators. The default value is `'-'` indicating that the evaluation time is provided externally (i.e., it corresponds to the utterance time) rather than by some other element in the computation such as, for example, a complementizer. This value is rewritten if such a potential binder is found.

[11]LiGER allows to check for the typical LFG relations *(inside-out) functional application* and *(inside-out) functional uncertainty* using the same symbols as the XLE: fa `!`, iofa `^`, and `*` for functional uncertainty respectively.

[12]The rule shown here is particular to a specific licensor for the PNP. However, the approach can be easily extended to others in a straightforward manner, either by simply introducing additional

versely, the PNP does not receive a semantic interpretation in matrix clauses since no evaluation time is specified for it there.

(15)
```
#a TEMP-REF # b T-REF 'non-past' &
#a VIEWPOINT #c ASPECT 'prv' &
#a ^(SEM>XCOMP) #d & #d !(SEM>TEMP-REF) #e EVAL #f
==> #a EVAL #e.
```

The rule for interpreting aspect shown in (16) picks up this s-structure node to add the semantic features for `VIEWPOINT` which encodes the semantic information of the markers for grammatical aspect (`ASPECT` in f-structure).

(16)
```
#a TNS-ASP #b ASPECT 'pe' & #a SEM #c
==> #c VIEWPOINT #d &
#d ASPECT 'prv' & #d A-RESTR 'partOf'.
```

In sum, the semantic construction rules produce a feature structure that reflects the semantic properties of tense and aspect. In the case of the PNP, the perfective aspect behaves as expected in that it follows the common analyses that postulate that the eventuality time is included in the reference time (here encoded as a part-of relation). The temporal dimension of the PNP provides a deficient instance of temporal reference that does not itself introduce an evaluation time. Following the discussion of the semantic interpretation rules, this ensures that the PNP can not occur in matrix clauses since this would lead to a resource deficit. This will become more clear in the context of the semantic interpretation rules discussed next.

4.5 Semantic construction

Let us first take a look at the fairly uncontroversial rules for interpreting viewpoint in (17). There, we present the rules that apply in the case of perfective aspect as encoded by the rules described above in (16). In other words, the following rules take as input the output of the rules presented before.

The quantifier over situations contributed by perfective aspect (see section 3.1.2) is decomposed into its restrictor and scope similar to the treatment of NP quantifiers (Dalrymple et al. 1999). Correspondingly, the rule in (17-a) introduces a `VAR` node and a `RESTR` node for this aspectual quantifier. The next rule uses these additional nodes to establish its restrictor, namely, the part-of relation. The final rule in (17-c) picks up the restrictor in typical Glue semantics fashion to provide a quantifier over situations that picks up the eventuality description and raises it to the level of temporal interpretation.

(17) a.
```
#a SEM #b VIEWPOINT #c
==> #c VAR #d & #c RESTR #e & #c ASP-RESTR #f.
```

rules that follow the same schema or by introducing the evaluation time separately and then linking it to the temporal reference annotation of the embedded eventuality, which would arguably be more in line with Giannakidou's (2009) proposal.

b.
```
#a SEM #b VIEWPOINT #c A-RESTR 'bounded' &
#c VAR #d & #c RESTR #e & #c ASP-RESTR' #f
==> #f GLUE [/s_s.[/t_s.partOf(t,s)]] :
(#d -o (#e -o #c)).
```
c.
```
#a SEM #b VIEWPOINT #c ASPECT 'prv' &
#c VAR #d & #c RESTR #e & #b TEMP-REF #f
==> #c GLUE
[/M_<s,<s,t».[/p_<s,t>.[/s_s.Ez_s[M(s)(z) & p(z)]]]] :
((#d -o (#e -o #c)) -o ((#c -o #b) -o (#f -o #b))).
```

At the level of temporal reference, the procedure is conceptually the same. The restrictor of the temporal quantifier is defined in (18-a). The rule in (18-b) checks whether temporal reference needs to be interpreted by checking for a feature other than `undefined`. In that case, a quantifier is introduced that carries an open situation slot that is saturated by whatever specifies the evaluation time of that quantifier, i.e., the speech time or some governing element, as is the case with the PNP.

(18) a.
```
#a SEM #b TEMP-REF #c T-REF 'non-past' &
#c EVAL #d
==> #c T-REF' #e &
#e GLUE [/t_s.[/t2_s.equals(t,i(t2,∞))]] :
(#c -o (#d -o #c)).
```
b.
```
#a SEM #b TEMP-REF #c T-REF %a &
%a != 'undefined' & #c EVAL #d
==> #c GLUE
[/T_<s,<s,t».[/P_<s,t>.[/s_s.Er_s[T(r)(s) & P(r)]]]] :
((#c -o (#d -o #c)) -o ((#c -o #b) -o (#d -o #b))).
```

Ultimately, the rules illustrated above produce the four meaning constructors presented in (19): two for viewpoint and two for temporal reference. Furthermore, the system provides a semantics for the VP consisting of information about the verb, its arguments, and inner aspect. These are subsumed in the VP placeholder. As shown in Figure 9, the combination of these meaning constructors is straightforward and mimics the compositional process described in section 3.2.2.

In the present pipeline, the derivation is conducted by the GSWB based on all the `GLUE` nodes introduced by the semantic interpretation rules discussed above. As described throughout this paper, the node modified for temporal reference still requires an evaluation time, here represented by the resource corresponding to index 1 (see Figure 9). The grammaticality of the PNP hinges on the fact whether this value is instantiated by rules like the one in (15). Thus, as already stated, the ungrammaticality of the PNP in matrix clauses is a simple case of a lacking resource.

(19) **VIEWPOINT:** $\lambda s_s.\lambda t_s.t <_p s : 4 \multimap (5 \multimap 3)$
$\lambda M_{<s,st>}.\lambda p_{st}.\lambda s_s.\exists z_s[M(s)(z) \wedge p(z)] :$
$(4 \multimap (5 \multimap 3)) \multimap (3 \multimap 0) \multimap (2 \multimap 0)$

$$
\begin{array}{ll}
\textbf{TEMP-REF:} & \lambda t_s.\lambda t''_s.t = i(t'', \infty) : 6 \multimap (7 \multimap 2) \\
& \lambda T_{<s,st>}.\lambda p_{st}.\lambda s_s.\exists r_s[M(r)(s) \wedge p(r)] : \\
& (6 \multimap (7 \multimap 2)) \multimap (2 \multimap 0) \multimap (1 \multimap 0) \\
\textbf{VP:} & \lambda s.VP(s) : 3 \multimap 0
\end{array}
$$

$$
\dfrac{\dfrac{3 \multimap 0 \qquad \dfrac{(4 \multimap (5 \multimap 3)) \multimap (3 \multimap 0) \multimap (2 \multimap 0) \qquad 4 \multimap (5 \multimap 3)}{(3 \multimap 0) \multimap (2 \multimap 0)}}{(2 \multimap 0)} \qquad \dfrac{(6 \multimap (7 \multimap 2)) \multimap (2 \multimap 0) \multimap (1 \multimap 0) \qquad 6 \multimap (7 \multimap 2)}{(2 \multimap 0) \multimap (1 \multimap 0)}}{(1 \multimap 0)}
$$

Figure 9: Partial derivation: Temporal reference and viewpoint of the PNP

This section has illustrated some capabilities of the LiGER system by illustrating some of the rules written for MG. As shown here, the core capability is based on basic graph pattern matching and checking for equations between features. However, it also supports further capabilities, such as checking for long-distance dependencies by using a mechanism inspired by (inside-out) functional uncertainty, which will be explored in future work.

5 Summary

We have presented work on the Modern Greek ParGram grammar developed in the XLE and a refinement of the ParTMA annotation scheme. Concretely, we added a semantic analysis to the grammar that captures the particular properties of the perfective non-past, a dependent verb form. We attributed this dependency to missing information in the semantics of temporal reference, following an analysis by Giannakidou (2009) and the semantic analysis of tense in Kusumoto (1999, 2005).

These ideas have been compiled into a Glue semantics treatment of tense and aspect providing insights into dealing with the resources contributed by different tense markers. More concretely, we have provided a two-component analysis that makes use of the ParTMA annotation scheme as a separate structure that serves as input to an interpretation procedure grounded in description-by-analysis approaches to Glue semantics (Andrews 2008).

On a more general level, the present paper has shown a new way of implementing description-by-analysis Glue semantics: the LiGER system, which is a new open-source resource for adding and rewriting annotations such as XLE's syntactic output. Combining this system with the XLE and the GSWB, we provide an alternative to XLE+Glue (Dalrymple et al. 2020), which specifically aims at covering the areas of Glue semantics research that are left open by this resource.

References

Abusch, Dorit. 1988. *Sequence of Tense, Intensionality and Scope*. Stanford University.

Abusch, Dorit. 1997. Sequence of Tense and Temporal De Re. *Linguistics and Philosophy* 20(1), 1–50.

Abusch, Dorit. 1998. Generalizing Tense Semantics for Future Contexts. In *Events and Grammar*, pages 13–33, Springer.

Andrews, Avery D. 2008. The Role of PRED in LFG + Glue. In *Proceedings of the LFG08 Conference*, pages 47–67.

Andrews, Avery D et al. 2007. Input and Glue in OT-LFG. *Architectures, Rules and Preferences: Variations on Themes by Joan Bresnan* pages 319–340.

Arregui, Ana, Rivero, María Luisa and Salanova, Andrés. 2011. Imperfectivity: Capturing Variation across Languages. *MIT Working Papers in Linguistics* 75, 53–68.

Arregui, Ana, Rivero, María Luisa and Salanova, Andrés. 2014. Cross-Linguistic Variation in Imperfectivity. *Natural Language & Linguistic Theory* 32(2), 307–362.

Asudeh, Ash. 2004. *Resumption as Resource Management*. Ph. D.thesis, Stanford University.

Asudeh, Ash. 2012. *The Logic of Pronominal Resumption*. Oxford University Press.

Bobrow, Daniel G., Cheslow, Bob, Condoravdi, Cleo, Karttunen, Lauri, King, Tracy Holloway, Nairn, Rowan, de Paiva, Valeria, Price, Charlotte and Zaenen, Annie. 2007. PARC's Bridge and Question Answering System. In *Proceedings of the GEAF 2007 Workshop*, pages 1–22.

Bohnemeyer, Jürgen and Swift, Mary. 2004. Event Realization and Default Aspect. *Linguistics and Philosophy* 27(3), 263–296.

Butt, Miriam, Dyvik, Helge, King, Tracy Holloway, Masuichi, Hiroshi and Rohrer, Christian. 2002. The Parallel Grammar Project. In *Proceedings of the 2002 Workshop on Grammar Engineering and Evaluation*, volume 15, pages 1–7, Association for Computational Linguistics.

Butt, Miriam, King, Tracy Holloway, Niño, María-Eugenia and Segond, Frédérique. 1999. *A Grammar Writer's Cookbook*. CSLI Publications.

Comrie, Bernard. 1976. *Aspect: An Introduction to the Study of Verbal Aspect and Related Problems*, volume 2. Cambridge university press.

Crouch, Dick, Dalrymple, Mary, Kaplan, Ronald M., King, Tracy Holloway, Maxwell III, John T. and Newman, Paula. 2017. *XLE Documentation*. Palo Alto Research Center.

Crouch, Richard. 2005. Packed Rewriting for Mapping Semantics to KR. In *Proceedings of the Sixth International Workshop on Computational Semantics (IWCS-6)*, pages 103–114, Tilburg.

Crouch, Richard and King, Tracy Holloway. 2006. Semantics via F-Structure

Rewriting. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG06 Conference*, pages 145–165, Stanford, CA: CSLI Publications.

Curry, Haskell Brooks, Feys, Robert, Craig, William, Hindley, J Roger and Seldin, Jonathan P. 1958. *Combinatory Logic*, volume 1. North-Holland Amsterdam.

Dalrymple, Mary, Lamping, John, Pereira, Fernando and Saraswat, Vijay. 1999. Quantification, Anaphora, and Intensionality. In Mary Dalrymple (ed.), *Semantics and Syntax in Lexical Functional Grammar – The Resource Logic Approach*, pages 39–89.

Dalrymple, Mary, Patejuk, Agnieszka and Zymla, Mark-Matthias. 2020. XLE+Glue – A New Tool for Integrating Semantic Analysis in XLE. In Miriam Butt, Tracy Holloway King and Ida Toivonen (eds.), *Proceedings of the LFG'20 Conference, Australian National University*, Stanford, CA: CSLI Publications.

De Wit, Astrid. 2016. *The Present Perfective Paradox Across Languages*, volume 4. Oxford University Press.

Dowty, David R. 1977. Toward a Semantic Analysis of Verb Aspect and the English 'Imperfective' Progressive. *Linguistics and philosophy* 1(1), 45–77.

Fiotaki, Alexandra and Lekakou, Marika. 2018. The Perfective Non-past in Modern Greek: a Corpus Study. *Scandinavian Journal of Byzantine and Modern Greek Studies* 4, 99–118.

Fiotaki, Alexandra and Tzortzi, Katerina. 2016. Exhaustive Object Control Constructions in Greek: an LFG/XLE Treatment. *Phrase Structure Grammar and Lexical Functional Grammar* pages 276–296.

Frank, Anette. 1999. From Parallel Grammar Development towards Machine Translation. A Project Overview. *Proceedings of Machine Translation Summit VII" MT in the Great Translation Era* pages 134–142.

Giannakidou, Anastasia. 2009. The dependency of the subjunctive revisited: Temporal semantics and polarity. *Lingua* 119, 1883–1908.

Grønn, Atle and von Stechow, Arnim. 2010. Complement Tense in Contrast: The SOT Parameter in Russian and English. *Oslo Studies in Language* 2(1).

Hatzigeorgiu, Nick, Gavrilidou, Maria, Piperidis, Stelios, Carayannis, George, Papakostopoulou, Anastasia, Spiliotopoulou, Athanassia, Vacalopoulou, Anna, Labropoulou, Penny, Mantzari, Elena, Papageorgiou, Harris et al. 2000. Design and Implementation of the Online ILSP Greek Corpus. In *LREC*.

Haug, Dag. 2008. Tense and Aspect for Glue Semantics: The Case of Participial XADJs. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG08 Conference*, pages 291–311, CSLI Publications.

Holton, David, Mackridge, Peter and Philippaki-Warburton, Irene. 1997. *Greek: A Comprehensive Grammar of the Modern Language*. London & New York: Routledge.

Ide, Nancy and Bunt, Harry. 2010. Anatomy of Annotation Schemes: Mapping to GrAF. In *Proceedings of the Fourth Linguistic Annotation Workshop*, pages 247–255.

Joseph, Brian D. and Smirniotopoulos, Jane C. 1993. The Morphosyntax of the

Modern Greek Verb as Morphology and Not Syntax. *Linguistic Inquiry* 24(2), 388–398.

Klein, Wolfgang. 1994. *Time in Language*. Psychology Press.

Kratzer, Angelika. 1998. More structural analogies between pronouns and tenses. In *Semantics and linguistic theory*, volume 8, pages 92–110.

Kusumoto, Kiyomi. 1999. *Tense in Embedded Contexts*. Ph. D.thesis, University of Massachusetts, Amherst.

Kusumoto, Kiyomi. 2005. On the Quantification over Times in Natural Language. *Natural Language Semantics* 13(4), 317–357.

Lekakou, Marika and Nilsen, Øystein. 2009. What Aspect can Tell us about the Future of MUST. *Proceedings of 8th ICGL. Ioannina: University of Ioannina* pages 261–276.

Mackridge, P. 1987. *The Modern Greek Language: A Descriptive Analysis of Standard Modern Greek*. Clarendon Paperbacks, Oxford University Press.

Meßmer, Moritz and Zymla, Mark-Matthias. 2018. The Glue Semantics Workbench: A Modular Toolkit for Exploring Linear Logic and Glue Semantics. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG'18 Conference, University of Vienna*, pages 249–263, CSLI Publications.

Mozer, Amalia. 2009. Άποψη και χρόνος στην ιστορία της Ελληνικής *[Aspect and Tense in the History of Greek]*. 2nd ed. Athens: National and Kapodistrian University of Athens.

Ralli, A. 2005. Μορφολογία*[Morphology]*. Ekd. Pataki.

Reichenbach, Hans. 1947. The Tenses of Verbs. *Elements of Symbolic Logic* pages 287–298.

Rosén, Victoria, Smedt, Koenraad De, Meurer, Paul and Dyvik, Helge. 2012. An Open Infrastructure for Advanced Treebanking. In Jan Hajič, Koenraad De Smedt, Marko Tadić and António Branco (eds.), *META-RESEARCH Workshop on Advanced Treebanking at LREC2012*, pages 22–29.

Roussou, Anna. 2000. On the Left Periphery. *Journal of Greek linguistics* 1, 63–92.

Sampanis, Konstantinos. 2012. "The Modern Greek Subjunctive Mood and its Semantic Features". In Current Trends in Greek Linguistics. Fragaki G., Georgakopoulos Th. and Ch. Themistocleous (eds), 66-91. Newcastle: Cambridge Scholars Publishing. .

Smith, Carlota S. 2013. *The parameter of aspect*, volume 43. Springer Science & Business Media.

Tsangalidēs, Anastasios. 1999. *Will and Tha: a Comparative Study of the Category Future*. University Studio Press.

Zymla, Mark-Matthias. 2017. Comprehensive Annotation of Cross-Linguistic Variation in the Category of Tense. In *12th International Conference on Computational Semantics*.

Zymla, Mark-Matthias and Sulger, Sebastian. 2017. Cross-Linguistically Viable Treatment of Tense and Aspect in Parallel Grammar Development. In *Proceedings of the LFG17 Conference*, CSLI Publications.