Proceedings of the LFG'18 Conference

University of Vienna

Miriam Butt, Tracy Holloway King (Editors)

2018

CSLI Publications

http://csli-publications.stanford.edu/LFG/2018

The papers are published under a CC-BY license:
http://creativecommons.org/licenses/by/4.0/

Contents

1 Editor's Note **4**

I Contributions to the Main Conference **5**

Hasiyatu Abubakari: Information Structure and the Lexical-Functional Grammar Framework **4**

Shatha Alruwaili, Louisa Sadler: Negative Coordination in (Turaif) Arabic **25**

Alex Alsina, Fengrong Yang: Catalan Intransitive Verbs and Argument Realization **46**

Roxana-Maria Barbu, Ida Toivonen: Romanian Object Clitics: Grammaticalization, agreement and lexical splits **67**

Tina Bögel, Saeed Reza Yousefi, Mahinnaz Mirdehghan: Vafsi Oblique Pronouns: Stress-related placement patterns **88**

Kersti Börjars, Christopher Hicks, John Payne: Interdependencies in Chinese Noun Phrases **109**

Maris Camilleri, Louisa Sadler: Schematising (Morpho)Syntactic Change in LFG: Insights from grammaticalisation in Arabic **129**

Paloma Carretero García: Dative Arguments in Psychological Predicates in Spanish **150**

Dewei Che, Adams Bodomo: A Constraint-Based Analysis of the Objects of VO Verbal Compounds in Mandarin Chinese **171**

Pui Lun Chow: The Malefactive Topic Role in Cantonese Indirect Passives **191**

Matthew Gotham, Dag Trygve Truslew Haug: Glue semantics for Universal Dependencies **208**

Tibor Laczkó: Modelling Possession and Agreement in Hungarian DPs: A paradigmatic approach **227**

Helge Lødrup: Prominent Internal Possessors and Backward Possessor Raising: Norwegian *ryggen på ham* 'the back on him' **248**

Moritz Meßmer, Mark-Matthias Zymla: The Glue Semantics Workbench: A modular toolkit for exploring Linear Logic and Glue Semantics **268**

Agnieszka Patejuk: Incorporating Conjunctions in Polish **283**

Agnieszka Patejuk, Adam Przepiórkowski: Predicative Constructions with Infinitival and Clausal Subjects in Polish **304**

Péter Sz[Pleaseinsertintopreamble]cs: A COMP-less Approach to Hungarian Complement Clauses **325**

Péter Szűcs: Operator Fronting in Hungarian **343**

1 Editor's Note

The 2018 Conference on Lexical Functional Grammar was held at the University of Vienna, Austria. The program committe for LFG18 were John Lowe and Ida Toivonen. We would like to thank them for again coordinating a very efficient and effective review process that in coordination with the local organizers resulted in an interesting and varied program. We would also like to thank the executive committee and the abstract and final paper reviewers, without whom the conference and the proceedings would not have been possible in this form.

The local organization consisted of Che Dewei, Hasiyatu Abubakari, Izabela Jordanoska and Muriel Assmann and was headed by Adams Bodomo and Daniel Büring, all of the University of Vienna. This team ensured an incredibly smooth organization of the conference, complete with an interesting workshop, wonderful invited speakers and stunning social activities. Our thanks go out to them for their very hard work in ensuring a successful conference.

This year's conference included a Workshop on Information Structure: Form and Interpretation, submitted papers from that workshop have been included in the proceedings.

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 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.

Hard Copy: All of the papers submitted to the LFG18 proceedings are available in one large pdf file, to be viewed and printed with Adobe Acrobat. The proceedings' file was created via pdflatex tools and a script written by Stefan Müller. We are highly indebted to him for the use of the script. We thank Emma Pease at CSLI Publications for having accompanied the LFG Proceedings over the years and making sure they become accessible and stay accessible. Finally, we thank Dikran Karagueuzian at CSLI Publications for his continuous support of our proceedings and our community.

Part I
Contributions to the Main Conference

Information Structure and the Lexical-Functional Grammar Framework

Hasiyatu Abubakari
University of Professional Studies, Accra

Proceedings of the LFG'18 Conference

University of Vienna

Miriam Butt, Tracy Holloway King (Editors)

2018

CSLI Publications

pages 4–24

http://csli-publications.stanford.edu/LFG/2018

Keywords: information structure, focus constructions, topic constructions, Kusaal

Abubakari, Hasiyatu. 2018. Information Structure and the Lexical-Functional Grammar Framework. In Butt, Miriam, & King, Tracy Holloway (Eds.), *Proceedings of the LFG'18 Conference, University of Vienna*, 4–24. Stanford, CA: CSLI Publications.

Abstract

This paper explores a formal analysis of Information Structure in Kusaal. It is observed that the i-structure is inadequately resourced to account for the various subcategories of discourse notions; more specifically the difference between information focus and contrastive focus in Kusaal and by extension some African languages. These two subtypes of focus are observed to have identical i-structures, resulting in ambiguity, although the c-structures may be different especially with languages where overt morphological particles play important roles in expressing the discourse statuses of constituents. To address the issue, an additional feature, DTYPE, with a value that subcategories subtypes of focus and topic notions is introduced in the i-structure. Another feature referred to as DFORM shows values that may either be morphologically or phonologically realised on an individual language basis.

1.0. Introduction[1]

This paper discusses a formal analysis of Information Structure which basically includes focus and topic constructions. The Lexical-Functional Grammar (LFG) architecture is premised on multiple levels of representation mediated through mapping. One such level of representation for Information Structure is the i(nformation)-structure (King 1997; Mycock 2006, 2013; Dalrymple and Nikolaeva 2011; Butt 2014). Previous attempts have been made at capturing finer grained components of Information Structure such as background, given, focus and topic within the i-structure projection (Butt 2014). This paper is intended to make further suggestions on capturing subtypes of discourse notions: information focus/ new information, contrastive focus, familiarity topic, contrastive topic etc, by building on the combined studies of previous attempts (Butt 2014; Choi 1996; Mycock 2006, 2013, Butt and King 1996 etc). This is deemed necessary because the existing analyses of focus constructions within the i-structure projection are insufficiently resourced to express correctly the statuses of given notions in Kusaal. The i-structure projection as it stands does not distinguish between subtypes of discourse notions such as information focus, contrastive focus, selective focus etc. although the c-structure from which the i-structure is mapped may have overt morphological markings for distinguishing various

[1] This paper is a modified version of aspects of chapter five in Abubakari (2018).
My sincere appreciation goes to all the anonymous reviewers for comments and suggestions which have greatly helped in improving this paper.

subtypes of focus i.e. in the case of Kusaal and several African languages. The mismatch between the c-structure and the i-structure makes the latter under-specified in expressing the exact notion conveyed in the c-structure most especially in instances involving contrastive focus which often results in ambiguity. The ambiguity arises because the same i-structure is projected for both information focus and contrastive focus as will be detailed soon. My aim, in this paper, is to indicate some challenges within previous proposals for Information Structure in LFG and further make suggestions as to how these problems can be resolved.

The discussion in this paper is divided into five (5) sections. After this section, section (2) will explore previous analyses of focus and topic constructions within the LFG literature. I will illustrate problems for these proposals with data from Kusaal. This will be followed by section (3) which will propose the introduction of additional features in the i-structure in an attempt to solve the problems raised in section (2). I further discuss how this proposal can be made universal to accommodate other languages whether discourse notions are expressed phonologically, syntactically, morphologically or by a combination of two or more of these strategies. Section (4) provides sample analyses using the proposed features in the i-structure for topic and focus constructions. Finally section (5) gives a summary of the paper.

2.0. Previous analyses of Information Structure in LFG

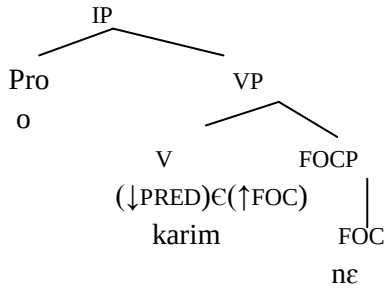
Following Falk (2001:58-59) and Bresnan and Mchombo (1987:757) grammatical function is the underlying concept behind the f-structure in LFG. Syntactic elements can simultaneously perform both grammatical and discourse functions. This has served as the main motivation behind the representation of both grammatical and (grammaticalized) discourse functions in the f-structure.

Examining the interaction that goes on between syntax, prosody and by extension morphology in encoding discourse functions King (1993/1995) and Choi (1996) opine that the introduction of an information (discourse) structure, in addition to the separation of the constituent structure from the functional structure, puts LFG in a better position to account for these interactions. This intervention became necessary in attempts to resolve issues of over-scoping when assigning discourse function to f-structure heads (King

1993/1995) as illustrated in example (1) in answer to the question 'What did he do?'

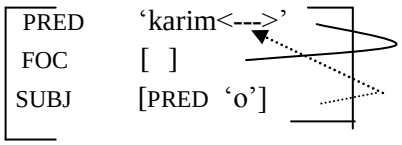
(1) a. Ò kārīm né.
3SG read FOC
'It is reading that he did (not for example *sleeping*)'

b. (↓PRED)Є(↑FOC)



The annotation thus (1b) results in over scoping of the focus domain as illustrated in the f-structure in (1c). By focusing the head 'read<SUBJ>' both the core meaning of the PRED and its argument get included in the focused domain although the interpretation of the contrastive focus in this sentence excludes all other constituents except for the verb (see King 1997).

c.



In finding a solution to the problem of over scoping, King (1997) suggests an approach which involves two basic parts. The first is to posit an i(nformation)-structure projection distinct from that of f-structure. The second is to remove the argument structure of the predicate, thus employing only the core grammaticalized discourse meaning in the i-structure. This means that the i-structure should refer to just the core meaning of the predicate excluding its arguments (see King 1997: 9-12; Butt and King 2000:11).

Below are two constructions: example (2) is an information focus whilst example (3) is a contrastive focus construction (É. Kiss 1989). Focus elements are annotated with ($\uparrow_i$ FOC), (↓PRED FN) Є($\downarrow_i$ REF) (also King 1997) whilst others are annotated as BACKGROUND (BGD). The c-structure

projections are further mapped on to the f-structure and the i-structure projections. Of particular interest is the comparison between the c-structure in (2b) and the i-structure in (2c) on the one hand and the c-structure in (3b) and the i-structure in (3c) on the other hand.

(2) a. What did he do?

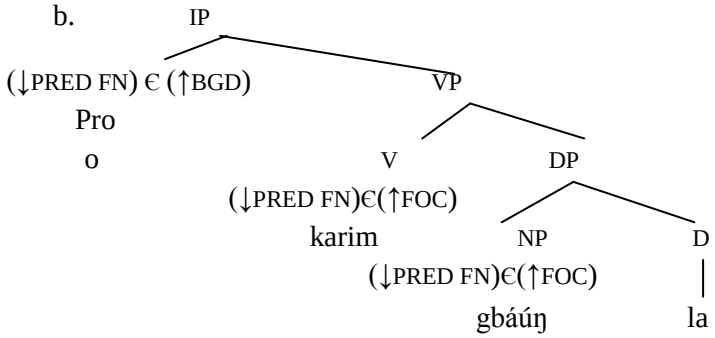
Ò kārīm gbáúŋ lá.

3SG read book DEF

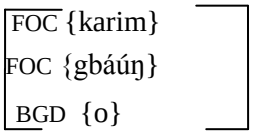
'He read the book.'

Information Focus= read the book

b.



c. I-structure



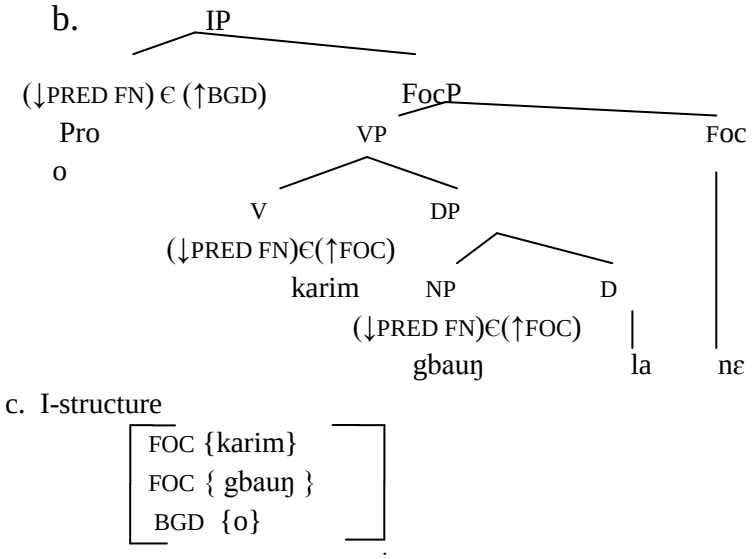
(3) a. What specifically did he do?

Ò kārīm gbáúŋ lá nέ.

3SG read book DEF FOC

'He READ THE BOOK' (as opposed to him selling the news paper for instance)'

Contrastive Focus =read the book



2.1. Problem one: Ambiguity in i-structure

Notice that the i-structures for the sentences in (2) and (3) are underspecified for the subtype of focus category they express. While *kārīm gbáúŋ lá* 'read the book' in (2c) is an information focus *kārīm gbáúŋlá* 'read the book' in (3c) is a contrastive focus and yet there are no specifications to facilitate the correct interpretation of each focus type. There is what I term 'discourse status under-specification' between the information in the c-structure and what is projected in the i-structure. To ensure a complete mapping of subtypes of discourse functions from the c-structure to the i-structure, it is important that the latter projection should reflect the exact discourse type in the well-resourced c-structure for maximum discourse effect and interpretation. Since all projections in LFG are mediated by mapping, and are independent structures, the i-structure does not efficiently express the desired discourse interpretations between contrastive and information focus since these two have the same i-structures. To address the ambiguity between (2c) and (3c), I will introduce a discourse feature in the i-structure with a corresponding value.

Another well acknowledged proposal on discourse information in LFG is the work of Choi (1996) who builds on the proposal of Vallduví (1992, 1993) to propose a four way distinction using two primitive distinctions in Information Structure [±New] and [±Prominent]. Vallduví (1992) divides Information Structure into *focus* and *ground*. He further subcategorizes *ground* into *link*

and *tail* where elements in the former are assumed to be more prominent than elements in the latter. Vallduví (1992) does not divide focus into subgroups. Building on this proposal, Choi (1996) divides focus into *contrastive focus* and *completive focus* (information focus), where contrastive focus is assumed to be 'more prominent' compared to completive focus. Choi (1996)'s four way distinction of Information Structure is captured in (4) below.

(4)

	+Prom	-Prom
-New	Topic	Tail
+New	Contrastive Focus	Completive Focus

From the diagram, topic and contrastive focus share the identical feature [+Prominent] distinguishing them from their less prominent counterparts tail (Background) and completive focus. Completive focus and contrastive focus share the same feature [+New] since they both introduce new referents into the discourse and what distinguishes the two is 'prominence'. Whilst completive focus is [-PROM], contrastive focus is [+PROM]. Proposing to represent the features [PROM] and [NEW] in the i-structure will not be a major issue but the question is as to whether this can serve a cross linguistic purpose.

2.2. Problem two: Prominence not a universal distinguishing feature in discourse notions

Prominence is not an exclusive feature of contrastive focus in Kusaal. It can as well be realised on information focus constituents as illustrated in (5b-c) following the context in (5a).

(5) a. Context: Assuming a context where a child is beaten but the culprit is not known. Whilst A in (5b) thinks *Aduku beat the child*, B in (5c) corrects A by indicating that *it is the man who beat the child*. The use of the long form of the noun *Aduku* instead of *Aduk* is a mark of emphasis accompanied by strong prominence. Kusaal has long and short forms of lexical items; the long forms are used in questions, negations and mostly for marking emphasis whereas the short forms are used elsewhere. Though *Aduku* in (5b) is an example of completive/information focus, it is as prominent as *dau la* 'the

man' in (5c) which is an example of contrastive focus further marked by the subject focus particle *ń*.

b. A: [Àdúk.ú]$_{+N+P}$bʊ^{l} bííg lá.
Aduk.Emph. beat child DEF
'Aduku beat the child.

c. B. Àyéí, [dáú lá]$_{+N+P}$ ń bʊ̄^{l} bííg lá.
no man DEF FOC beat.perf. child DEF
'No, it is the man who beat the child (not the woman, not Aduk)

Prominence as demonstrated can be a feature of both information focus and contrastive focus in Kusaal. The difference between information focus (5b) and contrastive focus (5c) is morphologically encoded in the presence of the particle *ń* in the case of the latter whilst same is not in the case of the former. In essence the feature [±PROM] cannot be used to distinguish between contrastive focus and information focus in Kusaal.

More recent studies on Information Structure in LFG which are closely related to the objectives of this paper but with slightly different goals and approaches include (Bodomo & Marfo 2005; Dalrymple and Nikolaeva 2011; Mycock 2006; 2013, Butt 2014). The central objectives of these studies have been word order, question formation and the relationship therein with Information Structure. The basic aim of this paper is to suggest an approach which will see the introduction of values that represent subtypes of discourse functions: contrastive focus, information focus, contrastive topic, familiarity topic etc in the i-structure projection in the LFG architecture. Below are short reviews of some of the above mentioned studies.

Marfo and Bodomo (2004) following Choi (1999; 2001) and Lee (2001) use the profile in (6) to describe the similarities between Q-words and focused constituents in Akan.

(6) Focus $\begin{bmatrix} \text{NEW} + \\ \text{PROM} + \end{bmatrix}$ Q-word $\begin{bmatrix} \text{NEW} + \\ \text{PROM} + \end{bmatrix}$

They argue that the two have identical c-structure and f-structure but they differ in the i-structure. To distinguish the focus type in wh-fronting and contrastive focus construction in the i-structure, they use the terms 'F-TYPE NEUTRAL' and 'F-TYPE CONTRASTIVE' for wh-fronting and contrastive focus

respectively (see Marfo and Bodomo 2005:199). The main analytical tool used by Marfo and Bodomo (2005) is OT-LFG.

Mycock (2013) considers the discourse functions of question words. In this work, she looks at the various possibilities that arise from the work of Butt (2012) and argues that question words can belong to the Information Structure categories Topic and Completive Information. Question words can have the same values as non-interrogatives for the information features [±NEW] and [±PROM] as suggested by Butt & King (1996). To mark the difference between interrogative and non-interrogative constituents Mycock augments Butt &King's (1996) system by having question words fully populate the information feature space. This proposal introduces an interface feature Q that is potentially relevant at multiple levels of the grammar in line with Dalrymple & Mycock 2011; Mycock &Lowe 2013.

Butt (2014) works on 'Question and Information Structure in Urdu/Hindi', where particular attention is devoted to word order variations involving wh-elements in constituent and polar questions in Urdi/Hindi. Butt, in her analysis, assumes an LFG architecture in which the i-structure is represented as a separate projection (in line with King 1997, Mycock 2006) but instead of the feature-based notions of topic, focus, background and completive information, Butt uses the basic notions of topic, focus and givenness and also allows for finer grained distinctions between these categories following Krifka's work.

The idea of introducing finer grained distinctions between discourse functions in the proposal of Butt (2014) and Choi (1999) are relevant to this paper as the same idea is adopted but in a different form and with more refined details. As indicated earlier, the analysis in this paper looks at subtypes of the notions of focus thus (information focus, contrastive focus etc), and subtypes of the notion of topic thus (familiarity topic, contrastive topic etc.) where discourse particles and phrases play integral roles.

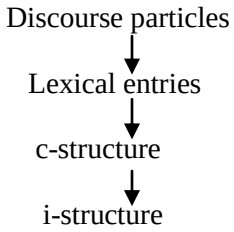
3.0. Towards a solution

The proposal to introduce finer grained details in the i-structure (Choi 1999; Butt 2014) serves as the foundation upon which the analysis for subtypes of discourse notions is built. I will suggest the introduction of additional discourse features in the i-structure to solve the issue of ambiguity. This proposal also suggests a path where language specific discourse strategies for

various Information Structure notions can be captured in the i-structure. The whole intervention as suggested here is a further development on the combined approaches of King (1997); Choi (1996) and Butt (2014). Since discourse particles are meaning distinguishing morphemes, it is paramount to include them in the i-structure to distinguish subtypes of discourse functions in a way close to the use of [+New, +Prom] by Choi (1996) to account for the various discourse notions in selected European languages: German, Russian, and English in the 'skeletal f-structure'.

African languages are predominantly particle-centred when it comes to the expression of discourse notions. These particles, generally referred to as discourse particles, cannot be excluded from a projection purposely designed to express the discourse statuses of constituents. Just as TENSE is primitive to the PREDICATE, thus the verb, so are these particles to discourse constituents such as focus and topic constituents. For this reason, we need to find a different way to treat them instead of considering them on a par with functional particles and eliminating them entirely from both the f-structure and the i-structure. Since the i-structure is the projection designated for discourse function, discourse particles should be added to the i-structure. In general, information in the i-structure becomes ambiguous if it is not adequate to express completely the discourse distinctions that are made in the c-structure especially in instances involving languages where discourse notions are expressed morphologically. Below is a suggestion of how these particles should be integrated from inception to finish in any analysis involving Information Structure.

(7) Suggested path for discourse particles[2]



[2]The a-structure is not included in this path since its function does not overlap directly with grammaticalized discourse function. The f-structure is also left out because of the issue of over-scoping of discourse domain discussed previously following King 1997. However, the c-structure maintains the value for discourse functions which is subsequently projected into the independent i-structure.

All discourse particles should be adequately captured in the lexical entries, represented in the c-structure and further mapped on to the i-structure. This ensures that discourse particles are fully accessible to the i-structure for a holistic discourse interpretation and a complete mapping between c-structure and i-structure.

3.1. Introducing DTYPE and DFORM

I propose a feature in the i-structure referred to as Discourse Type (DTYPE). DTYPE will have attributes that provide further details of the discourse subtype: contrastive focus, completive/information focus and topic. The value for DTYPE will correspond to the discourse status of the constituent in question together with the corresponding particle if any or the feature specification of the said discourse status determined by the language in question. Discourse particles or feature specifications will be referred to as discourse form (DFORM). In other words, a DFORM is a further break down of how a language expresses its DTYPE which may be morphological, phonological or otherwise. For instance a DTYPE can have the value {contrastive focus} and DFORM of the value {né} for Kusaal and DTYPE value {contrastive focus} with a corresponding feature specification, thus, DFORM value {+ NEW +PROM} for German. The predicate functor (PRED FN) is represented in the i-structure as REF(ERENCE). This is mainly aimed at distinguishing subcategories of discourse functions (focus and topic). More specifically, within the i-structure, each discourse function (focus and topic) is still set valued, but each item of the set is an AVM which contains the following:

i. The PRED FN is coded as REF(ERENCE)
ii. The DTYPE, is an abstract meaning like "contrastive" or "completive" and is a subtype of DF
iii. The DFORM, is the particle form such as né and /or intonation or prosodic information such as [±PROM] or null [Ø] for any given DTYPE.

Additionally, the value of DTYPE may have a corresponding relationship with the value of DFORM with the latter being morphologically, phonologically, or syntactically encoded in the particle used or the phonological features associated with the said notion. This will also be entirely language dependent since different languages have different discourse particles that may also be tied to specific discourse strategies (Abubakari 2018).

The rule below serves the purpose of identifying values of DTYPES with corresponding DFORMS on language specific basis.

(8)

	Values
DTPYE:DFORM	{information focus: ɑ}
	{contrastive focus: x}
	{familiarity topic: φ}
	{contrastive topic: q}
	etc

(where ɑ,x, φ, and q are particles if any or features such as [±New] or [±Prom] or others)

I provide values for both DTYPES and DFORMS in (9) Kusaal and (10) German, English and Russian.

(9)

Kusaal	Value
DTPYES:DFORMS	{in-situ focus}
	{contrastive focus: n}
	{contrastive focus: nɛ}
	{contrastive focus: ka}
	{familiariy topic: Ø, -N, +Prom}
	{contrastive topic: yaa an, -N,+Prom}

The rule for Kusaal in (10) implies that the DTPYE: information focus is morphologically null, there are no corresponding particles (DFORMS) for this discourse subtype hence the use of {Ø} value. It is infelicitous to use the DFORM values [+New] and [+Prom] since the same values apply to contrastive focus in Kusaal. On the other hand, contrastive focus has different DFORM values for its subtypes: in-situ subject focus: *n*, in-situ non-subject focus: *nɛ*, and ex-situ non-subject focus *ka* (Abubakari forthcoming; 2018). These values are included for purposes of providing finer grained details of various discourse functions as the case may be. Topics, in Kusaal, are also subcategorized into two: familiarity topic and contrastive topic (Abubakari 2018). These are further distinguished by the absence of the special topic phrase in the former, resulting in a DFORM value of {Ø, -N, +Prom} while the latter has the said phrase, resulting in a DFORM value of {yaa an, -N,+Prom} .

(10) English, German, Russian Values

DTYPES:DFORMS → {information focus: +N, -Prom}
DTYPES:DFORMS → {contrastive focus: +N, +Prom}
DTYPES:DFORMS → {topic:-N,-Prom}

In the absence of overt morphological markings, the DFORM values, [±NEW] and [±PROM] are used to set apart the differences between contrastive focus, information focus and topic in English, German and Russian (see Choi 1996).

In this section, I have proposed that the features DTPYE and DFORM be introduced in the i-structure with values that specify the status of a discourse constituent. I have indicated how languages can apply the rule in generating the needed mechanism to disambiguate discourse constituents in the i-structure. In the next section, I will focus on providing an analyses of both focus and topic constituents with data from Kusaal.

4.0. Sample analyses

In this section, I intend to provide sample analyses demonstrating the implementation of the proposal in section 3. The analyses fall in three categories: argument focus, VP focus and IP focus and topic constructions. For each analysis I will begin from the lexical entry to the c-structure followed by the i-structure.

4.1. Category one: Argument focus

In answer to the question in (11), the sentence in (11a) is information focus construction and that in (12a) is contrastive focus construction.

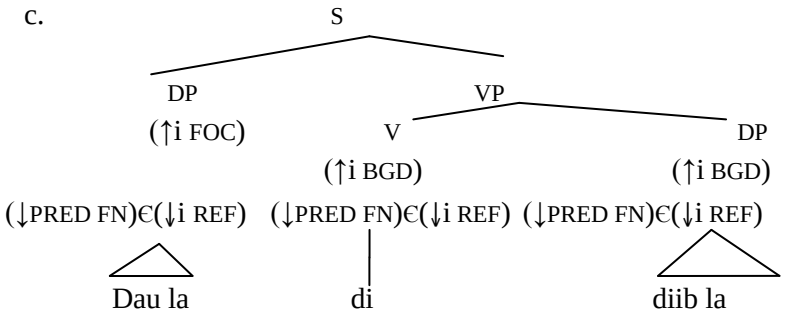
(11) Q.: Who ate the food?

a. Dáú lá dī díib lá.
man DEF eat food DEF
'The MAN ate the food.'

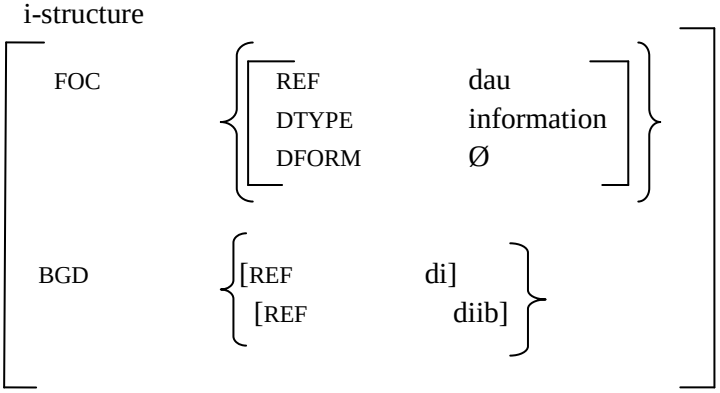
b.Lexical entries: Dáú lá dī díib lá.

Dau	N(ˆPRED) = 'dau'
Di	V (ˆPRED) = 'di<(ˆSUBJ) (ˆOBJ)>'
Diib	N(ˆPRED) = 'diib'

c.



d. i-structure



The discourse status of *dáú* 'man' is explicitly expressed from the lexicon to the i-structure. The value of DTYPE specifies that the focused constituent in question *dáú* 'man' subcategorizes as an information focus constituent. Each level of the architecture independently expresses this status which is mapped from one projection to the other.

Consider the contrastive focus construction in (12a).

(12) a. Àyéí, bíís lá ń dí dííb lá.
no children DEF FOC eat food DEF
'It is the children that ate the food.'

b.Lexical entries

Biis	N(ˆPRED) = 'biis'
n	FP[3](ˆi DFORM) = 'n'

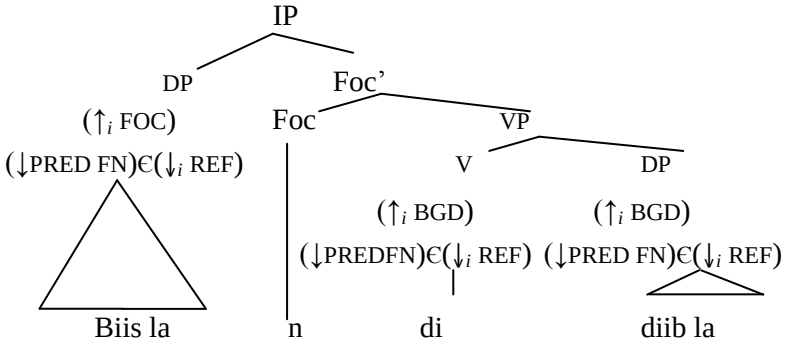
[3] FP refers to 'Focus Particle' and for projection into i-structure, arrows are subscripted with an 'i'.

(ˆi DTYPE) = CONTRASTIVE FOCUS

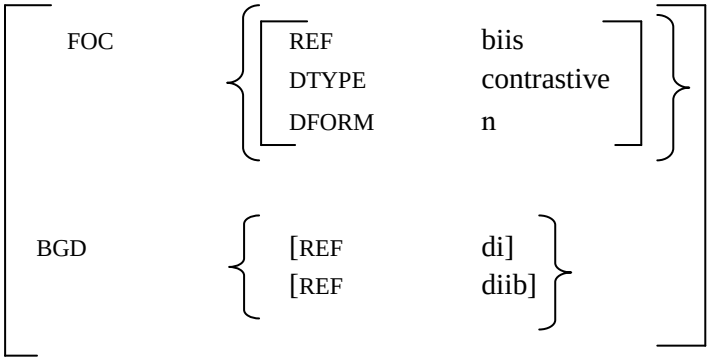
Di V(ˆPRED) = ‘di<(ˆSUBJ) (ˆOBJ)>’

Diib N(ˆPRED) = ‘diib’

c. C-structure equation



d. i-structure



From the lexical entry through to the c-structure and subsequently the i-structure, the subtype of the discourse status of the focused constituent is clearly specified as contrastive focus. Unlike the c-structure in (11c), the c-structure in (12c) has a projection for a focus particle which hosts the focused subject at the specifier of Foc. The focused particle *ń* which is listed in the lexical entries conveys relevant information regarding the focused constituent. The same information is inherently mapped on to the i-structure by the predicate attribute DFORM with the value *n*. Finally, the focused constituent *bíís* ‘children’ in the i-structure can be argued to have all the necessary resources that fully identify its discourse subcategory.

Having considered an example involving in-situ contrastive subject focus in (12), the example in (13) is a demonstration of in-situ contrastive focus with an object.

(13) Did the children eat the fruits or the food?

a. Bíís lá sà dí né dííb lá.

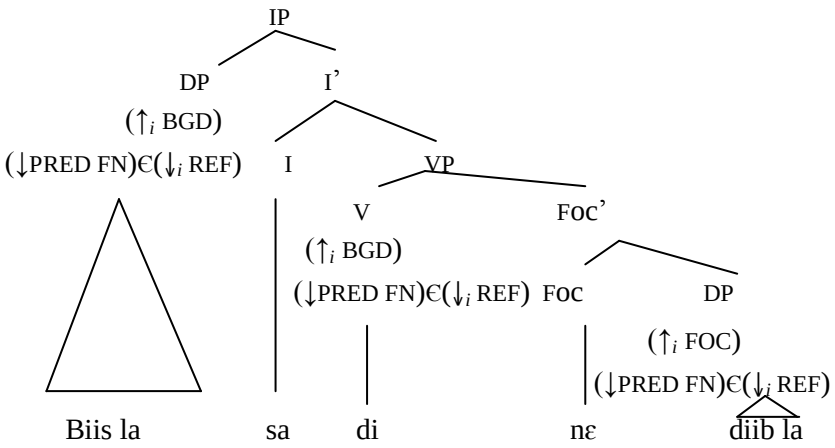
children DEF PAST eat FOC food DEF

'It is the food that the children ate (yesterday).'

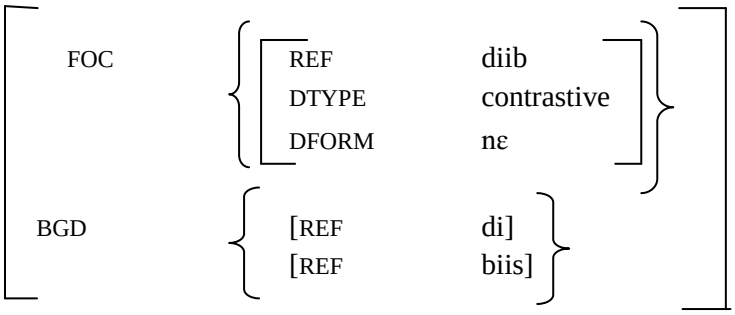
b. Lexical entries

Biis	N(ˆPRED) = 'biis'
Di	V(ˆPRED)='di<(ˆSUBJ) (ˆOBJ)>'
Diib	N(ˆPRED) = 'diib'
né	FP (ˆi DFORM) = 'né'
	(ˆi DTYPE) = CONTRASTIVE FOCUS

c. c-structure



d. i-structure



The status of the focused element *dííb* 'food' is specified as contrastive by virtue of the particle *né*. The representation of this particle from the lexical entries through to the i-structure ensures full specification and coherent discourse interpretation in the various projections.

4.2. Category two: VP and IP focus

To mark VP or IP focus, the focus particle *né* occurs after the focused VP or IP, i.e., at clause internal right periphery (Abubakari forthcoming). The response in (14ii) is a surprised response which is out of the hearer's expectation. It is used in a context where 'no one is supposed to eat a particular food'. The entire response, i.e., the IP is focused with the particle *né*, emphasising that some people defied the said order.

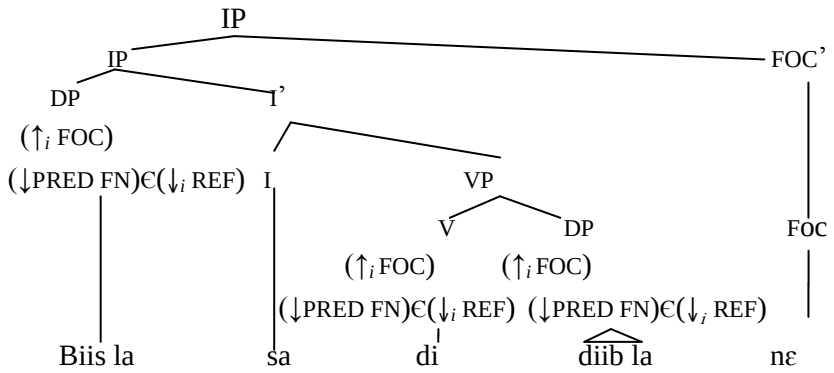
(14) a. i. Q: What happened?

ii. Ans: [Bíís lá sà dī dííb lá né]$_{F}$
children DEF PAST eat food DEF FOC
'THE CHILDREN ATE THE FOOD (yesterday).'

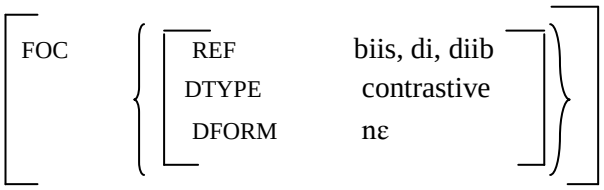
b. Lexical entries

Biis	N(ˆPRED) = 'biis'
Di	V(ˆPRED)='di<(ˆSUBJ) (ˆOBJ)>'
Diib	N(ˆPRED) = 'diib'
nɛ	FP(ˆi DFORM) = 'nɛ'
	(ˆi DTYPE) = CONTRASTIVE FOCUS

c. c-structure



d. i-structure



Finally the i-structure shows that the entire IP is focused. Every constituent in this structure is contrastively focused as they all share the single DTYPE with the value contrastive focus and the same DFORM of the value *nɛ*. The same discourse information is traceable from the lexical entries through to the c-structure and finally to the i-structure.

4.3. Category three: Subcategories of topics and the i-structure

Two types of topic are indentifies in Kusaal: familiarity topic (15) and contrastive topic (16) (Abubakari 2018).

(15) a. Dííb lá, ò sà dī lì.
food DEF 3SG PAST eat it
'The food, s/he ate it (yesterday).'

(16) a. Yáˈá án dííb lá, ò sà dī lì.
if COP.be food DEF 3SG PAST eat it
'As for the food, I ate it (yesterday).'

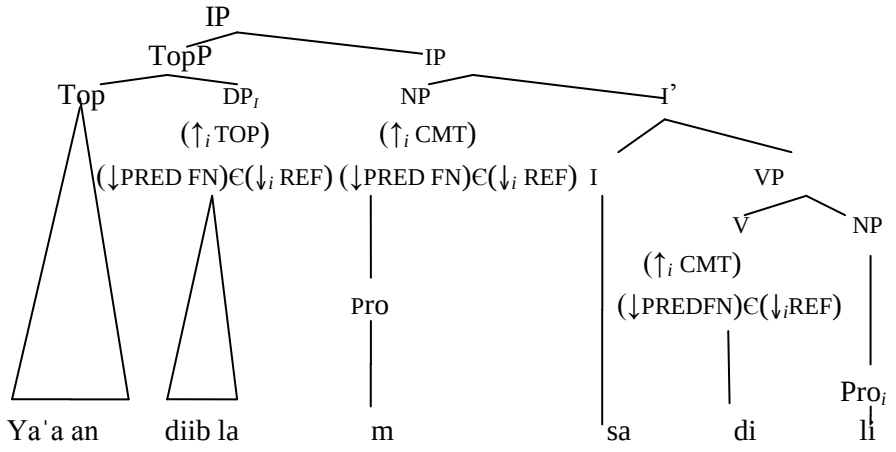
Topic constituents that are qualified by the special topic phrase are classified as contrastive topics and those without the topic phrase are categorized as

familiarity topics on pragmatic grounds (Abubakari 2018). DTYPE is either valued as {contrastive topic} with a corresponding DFORM which is valued as {yáá án} or DTYPE {familiarity topic} with a corresponding DFORM {ø} for Kusaal. Below are the various stages and projections for the contrastive topic construction in (16) within the proposed analysis. (17a) is the lexical entries, whilst (17b) is the c-structure and (17c) is the i-structure projection.

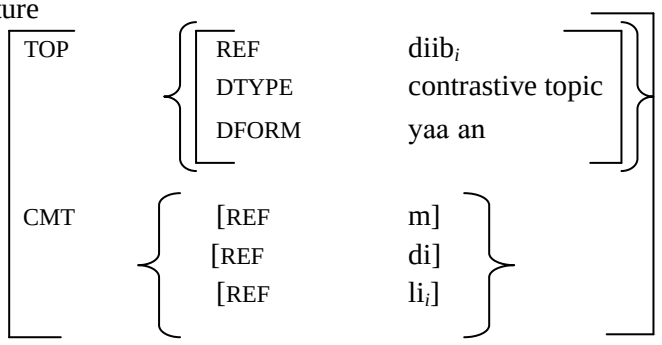
(17) a. Lexical entries

Diib	N(ˆPRED) = 'diib'
Yaˈa an	TP[4](ˆi DFORM) = 'yaˈa an'
	(ˆi DTYPE) = CONTRASTIVE TOPIC
Di	V(ˆPRED) = 'di<(↑SUBJ) (↑OBJ)>'
O	PRO (ˆPRED) = 'o'
Li	ANAPHORIC PRO (ˆPRED) = 'li'

b. c-structure



c. i-structure



[4] TP: Topic phrase

The i-structure is able to set the difference between a familiarity topic construction which is without the topic phrase and contrastive topic construction with the topic phrase by virtue of the values of their respective DTYPEs and corresponding DFORMs.

5.0. Conclusion

This paper set out to discuss a formal account of Information Structure in Kusaal using the Lexical Functional Grammar framework. The main purpose has been to point out issues in previous analyses of focus constructions in the i-structure projection and to suggest possible ways of addressing the problem(s). Generally, it was found that the i-structure is inadequately resourced to account for the various subtypes of discourse notions; more specifically the difference between information focus and contrastive focus. These two subtypes of focus are observed to have identical i-structures although their c-structures may be different especially in languages where overt morphological particles play important roles in expressing the discourse statuses of constituents. The impossibility of differentiating between subtypes of focus in the i-structure results in ambiguity and under specification of discourse interpretations. In addressing the problem, I introduced an additional feature, DTYPE, with a value that specifies subtypes of focus and topic notions in the i-structure. DTYPE can have a value, for example, {contrastive focus} or {information focus}. Another feature referred to as DFORM shows values that may either be morphologically or phonologically realised on individual language basis. For instance the feature values [±New] and [±Prom] are suggested for some European languages whilst the morphological features: *n*, *nɛ* and *ka* are used for Kusaal.

References

Abubakari, Hasiyatu. (2018). *Aspects of Kusaal Grammar: The Syntax-Information Structure Interface*. PhD Dissertation, African Studies Department: University of Vienna.

Abubakari, Hasiyatu. (Forthcoming). Contrastive Focus Particles in Kusaal. *Proceedings of ACAL47*. Berkeley: University of California.

Bresnan, Joan & Mchombo, Sam A. (1987). Topic, Pronoun, and Agreement in Chicheŵa. *Language*, Vol.63, 4:741-782.

Butt, Miriam. (2014). Questions and Information Structure in Urdu/Hindi. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of LFG14 Conference*. CSLI Publications.

Butt, Miriam & Tracy, Holloway King. (1996). Structural Topic and Focus without Movement. In *Proceedings of the First Annual LFG Conference*. Stanford: CSLI Publications.
Choi, Hye-Won. (1999). *Optimizing Structure in Context: Scrambling and Information Structure*. Stanford, CA: CSLI Publication.
Choi, Hye-Won. (1996). *Optimizing Structure in Context: Scrambling and Information Structure* (Doctoral dissertation, Stanford University). Retrieved from http://roa.rutgers.edu/files/147-1096/147-1096-CHOI-0-2.PDF
Dalrymple, Mary & Nikolaeva, Irina. (2011). *Objects and Information Structure*. Cambridge: University Press.
Falk, N. Yehuda. (2001). *Lexical-functional grammar: an introduction to parallel constraint-based syntax*. Stanford, CA: CSLI Publications.
É. Kiss, Katalin. (1998). Identificational Focus versus Information Focus. *Language* 74:245-273.
King, Tracy Holloway. (1997). Focus domains and information-structure. In Miriam Butt and Tracy Holloway King (eds). *Proceedings of the LFG97 Conference*. Stanford: CSLI Publications.
King, Tracy Holloway. (1995). Focus domain and information –structure. In Miriam Butt and Tracy Holloway King (eds.) *Proceedings of the LFG97 Conference*. Stanford: CSLI Publications.
King, Tracy Holloway. (1993). *Configuring Topic and Focus in Russian*. Doctoral dissertation, Stanford University. Published by CSLI Publication, 1995.
Marfo, Charles. & Bodomo, Adams. (2004). Information Processing in Akan Question-Word Fronting and Focus Constructions. In Miriam Butt and Tracy Holloway King (eds). *Proceedings of the LFG04 Conference.* Stanford: CSLI Publications.
Mycock, Louise & John, J. Lowe. (2013). The Prosodic Marking of Discourse Functions. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG13 Conference*. Stanford, CA: CSLI Publications.
Mycock, Louise. (2006). The Typology of Constituent Questions: A Lexical-Functional Grammar Analysis of wh-questions. PhD. Thesis, University of Manchester.
Vallduví, Enric. (1993). Information packaging: A survey. Research Paper HCRC/RP-44.
Vallduví, Enric. (1992). *The Informational Component*. New York: Garland Press.

Negative Coordination in (Turaif) Arabic

Shatha Alruwaili
University of Essex

Louisa Sadler
University of Essex

Proceedings of the LFG'18 Conference

University of Vienna

Miriam Butt, Tracy Holloway King (Editors)

2018

CSLI Publications

pages 25–45

http://csli-publications.stanford.edu/LFG/2018

Keywords: Arabic, Turaif, negation, coordination

Alruwaili, Shatha, & Sadler, Louisa. 2018. Negative Coordination in (Turaif) Arabic. In Butt, Miriam, & King, Tracy Holloway (Eds.), *Proceedings of the LFG'18 Conference, University of Vienna*, 25–45. Stanford, CA: CSLI Publications.

Abstract

We discuss the combination of negation and coordination in an Arabic construction which is somewhat akin to the *neither...nor* construction in English and many other languages. In Arabic however, the form marking the non-initial conjunct is transparently related to the *and* coordinator rather than the *or* form. We provide an analysis of bisyndetic negative coordination expressing both sentential and constituent negation, and also as negative concord in certain contexts. We draw exclusively on data from the Turaif variety of Arabic in our discussion. The central facts concerning the use and distribution of the bisyndetic negative coordination construction are broadly similar across the Arabic vernaculars.

1 Introduction

We discuss the combination of negation and coordination in an Arabic construction somewhat akin to the *neither...nor* construction in English illustrated in (1) and (2) (for the coordination of predicates and arguments respectively), corresponding in logical meaning to the monosyndetic examples with a single *and/or* in (3).

(1) John neither washed nor dried the dishes.

(2) Leo ate neither the rice nor the carrots.

(3) John did not wash the clothes and did not hang them out to dry (either).
John did not wash or dress.
Leo did not eat rice or carrots.

Haspelmath (2004, 2007) describes coordinated structures of the bisynthetic (and polysynthetic) types such as (1) and (2) as instances of 'emphatic coordination' (or focusing coordination), arguing that where every term has a negative coordinator, the terms are indicated as being in some sort of contrast. On the other hand, the examples in (3) have a less 'emphatic' flavour.

Our discussion draws on data from the Turaif dialect (of Saudi Arabia), but the facts are broadly similar in other contemporary varieties of Arabic. The bisynthetic construction, also referred to as emphatic bisynthetic coordination (Haspelmath, 2004, 2007) is illustrated in (4) and (5). We focus in particular on the use of *wala* and *lā*.

(4) a. mansōr **mā** akal l-ruz **wala** šarab
Mansour.M NEG eat.PFV.3SGM DEF-rice NEG.CONJ drink.PFV.3SGM
l-gahwa
DEF-coffee
Mansour neither ate the rice nor drank the coffee.

[†]This work was partially funded by Leverhulme Major Research Fellowship MRF-2016-048. Support from this source is gratefully acknowledged. Thanks also to the Essex LFG research group, the audience at LFG 2018 and two reviewers for very useful feedback and discussion.

b. mansōr **lā** akal l-ruz **wala** šarab
Mansour.M NEG eat.PFV.3SGM DEF-rice NEG.CONJ drink.PFV.3SGM
l-gahwa
DEF-coffee

Mansour neither ate the rice nor drank the coffee.

(5) **lā** ʔaħmad **wala** mhammad ǧ-aw
NEG Ahmad.M NEG.CONJ Mohammad come.PFV-3PLM

Neither Ahmad nor Mohammad came.

The element *wala* is polysemous - we will gloss the *wala* which appears in this construction as NEG.CONJ and refer to it as 'coordination *wala*'. We will gloss *lā* as NEG (reflecting its etymological source). Coordination *wala* is transparently related to a combination of the conjunction and a negative particle. The main questions which we address here are: (i) does coordination *wala* contribute negation or is it simply restricted to a negative environment?; (ii) what are the constraints on the constructions illustrated above and how can they be captured in LFG?

2 Agreement, Coordination, Disjunction

In Turaif Arabic we find full agreement in both SVO and VSO word orders (SVO is the common or default word order). Both 3SG and 3PL show gender agreement (i.e. there is a 3PLF form in this variety of Arabic). When the agreement controller is coordinate we find fully resolved agreement in SVO order and both fully resolved and closest conjunct agreement (CCA) in VSO word order. With disjunctive agreement controllers, we find a closest conjunct agreement pattern in both word orders. As we will see in section 4.3, the *lā...wala* negative coordination structure exhibits its coordinative (rather than disjunctive) nature by following the agreement pattern of *w* 'and' (6) rather than *yā* 'or' (7).

(6) a. huda w mansōr ǧ-aw
Huda.F CONJ Mansour.M come.PFV-PLM

Huda and Mansour came.

b. huda w nora ǧ-an
Huda.F CONJ Noura.F come.PFV-3PLF

Huda and Noura came.

(7) yā abō-i yā ʔumm-i raħ ti-ǧ-i
either father.M-1SG.GEN or mother.F-1SG.GEN FUT 3SGF-come.IMPV
bokra
tomorrow

Either my father or my mother will come tomorrow.

3 Negation

3.1 Sentential Negation

Like many other vernacular Arabics, sentential negation in verbal sentences in Turaif Arabic uses the particle *mā* immediately before the verbal element, as in (8). This verbal strategy with *mā* also extends to use with pseudo-verbs as in (9).[1]

(8) a. ʕali mā kitab l-waǧib
Ali.M NEG write.PFV.3SGM DEF-homework
Ali didn't do the homework.

b. ʕali mā ya-ktib l-waǧib
Ali.M NEG 3SGM-write.IMPV DEF-homework
Ali doesn't do the homework.

(9) huda mā ʕinda-ha/maʕa-ha sayyārah
Huda.F NEG with-3SGF.GEN/with-3SGF.GEN car
Huda doesn't have a car.

Sentential negation with non-verbal predicates (other than the set of pseudo-verbs which exhibit the verbal strategy with *mā*) uses the particle *mū*, and its inflectional counterparts which show agreement with the subject, in (10).[2]

(10) a. ʕali imdars
Ali.M teacher.SGM
Ali is a teacher.

b. ʕali mū/mahu imdars
Ali.M NEG/NEG.3SGM teacher.SGM
Ali is not a teacher.

The distribution of *lā* is much more constrained in vernacular Arabic than it is in Modern Standard Arabic. In connection with Palestinian Arabic, Hoyt observes: "In Classical Arabic and early forms of the dialects (c.f. Blau, 1967), the *la*-particle was itself ambiguous between three uses: (i) expressing present tense verbal negation; (ii) expressing existential or categorial negation (Arabic *nafi lǧins* "negation of the kind") ...; and (iii) negative imperatives. Of these, (i) and (ii) have

[1] The term pseudo-verb is used for forms diachronically related to prepositions and nouns which do not inflect as regular verbs (but by means of a GEN affix coding the SUBJ) but exhibit verbal functions, including that of occurring as the main sentential predicate. The use of the verbal strategy for negation distinguishes the pseudo-verb from its prepositional counterpart (such as the locative prepositions *ʕind* 'at' and *maʕ* 'with').

[2] As shown by (10) there is no copula in the affirmative predicational clause with present tense interpretation. It is sometimes claimed that *mū* (and inflectional variants) is a form of copula verb. Whether or not this is the case, the main point here is that (10b) is an instance of sentential negation.

largely been reduced to formulaic borrowings from Standard Arabic, leaving negative imperatives...as the primary productive use of *lɛʔ*". (Hoyt, 2010, 108). (11) illustrates existential or categorial negation in MSA.

(11) lā šakka-a fī ḏālika
NEG doubt-ACC.INDEF on that.FSG
There's no doubt about that. (Ryding, 2005, 179)

Beyond the coordinative constructions we discuss here, the particle *lā* occurs only in the prohibitive (negative imperative), shown in (12), in Turaif Arabic.[3]

(12) a. ʔktib l-waǧib!
write.IMPV DEF-homework
Do the homework!

b. lā ta-ktib l-waǧib!
NEG 2SGM-write.IMPV DEF-homework
Don't do the homework!

3.2 Other Strong Negative Elements

Alongside sentential negative particles, there are certain other expressions in Turaif Arabic which have inherently negative meaning, as shown by the fact that they may occur as fragment negative answers to questions. Of these, relevant to the current topic, we find (i) the negative (pronominal) quantifier *māħad* 'no one' and (ii) the negative quantifier *wala* 'not even one' in its scalar focus particle (SFP) use, which combines with an indefinite NP. Note that, as observed in section 1, *wala* is polysemous, and indeed there are other languages where the same word form occurs both in SFP and negative coordinator uses, such as *ani* in Polish (cf. also Russian, Hungarian, Modern Greek and Romanian (Haspelmath, 2004)).

The examples in (13) show that *māħad* 'no one' is an inherently negative word appearing in preverbal position, and as a consequence combining it with the sentential negation marker leads to a 'double negative' interpretation, as in (13b). It does not occur in postverbal position where instead we find the corresponding item *ʔaħad* anyone, which does not itself express any negative meaning. As shown in (13c) to convey 'no one' it will occur in the context of a preceding sentence negative *mā, mū*, etc. SFP *wala*, illustrated in (14), also has an inherently negative meaning preverbally, which is negated if the sentential negative occurs in the same sentence. However, it also occurs in postverbal position, as in (15) where it behaves like *ʔaħad* in that it requires a preceding sentence negator to convey its usual negative meaning. Thus strong preverbal SFP *wala* contributes negation while weak

[3]The (positive) imperative is formed of the imperfective stem (without the agreement prefix), and an epenthetic augment. We gloss it simply as an imperfective stem. The prohibitive is formed of the imperfective stem with second person inflection, preceded by *lā*.

postverbal SFP *wala* occurs in a "negative context", and arguably exhibits negative concord (NC). However as Lucas (2009, 187) claims, "the Arabic varieties that exhibit true negative concord are fewer than what is claimed in the literature".

(13) a. māħad ǧ-a l-yōm
no.one come.PFV-3SGM DEF-day
No one came today.

b. māħad mā ǧ-a l-yōm
no.one NEG come.PFV-3SGM DEF-day
No one didn't come today. (= Everyone came today.)

c. mā ǧ-a ʔaħad l-yōm
NEG come.PFV-3SGM one DEF-day
No one came today.

(14) a. wala ṭālib ǧ-a l-yōm
NEG.SFP student.SGM come.PFV-3SGM DEF-day
Not even a (single) student came today.

b. wala ṭālib mā ǧ-a l-yōm
NEG.SFP student.SGM NEG come.PFV-3SGM DEF-day
Not even a single student didn't come today.
(= Every student came today.)

(15) a. mā ǧ-a wala ṭālib l-yōm
NEG come.PFV-3SGM not.even student.SGM DEF-day
Not even a (single) student came today.

b. *ǧ-a wala ṭālib l-yōm
come.PFV-3SGM not.even student.SGM DEF-day
Intended: Not even a (single) student came today.

Following Przepiórkowski and Patejuk (2015) (see also Sells (2000), Laczkó (2014) and Laczkó (2015)) on the syntactic aspects of such negative items, we will represent the distinction between constituent negation and eventuality negation at f-structure, using two features ENEG and CNEG (standing for eventuality negation and constituent negation).[4] Thus an example such as (14b) with NQ SFP *wala* and a realisation of sentential negation will be represented as in (16).[5]

[4]For the data which we discuss here, it would in principle be possible to replace the features ENEG and CNEG by a more general feature NEG, available in the f-structures corresponding to sentences and their dependents alike. Discussion of the further aspects of negation in Arabic which do in fact motivate the maintenance of the ENEG/CNEG distinction we make use of here would take us too far afield.

[5]The feature SFOC simply provides a syntactic indication of the scalar focussing property of negative quantifier *wala*. It is not important in the present context.

(16)
$$\begin{bmatrix} \text{PRED} & \text{`COME}< \text{SUBJ} > \text{'} \\ \text{ENEG} & + \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`STUDENT'} \\ \text{CNEG} & + \\ \text{NUM} & \text{SG} \\ \text{SFOC} & + \end{bmatrix} \\ \text{ADJ} & \left\{ \begin{bmatrix} \text{PRED} & \text{`TODAY'} \end{bmatrix} \right\} \end{bmatrix}$$

4 Negative Coordination

There are several strategies for expressing the coordination of negated predications in Turaif Arabic. In particular, although neither *lā* nor *wala* are used as markers of sentential negation they occur in widespread strategies for negative coordination.

4.1 With Verbal Predicates

The examples in (17) involve coordination at the lexical level where we see three variants are possible: *mā ... w mā* in (17a) involves the standard marker of verbal negation on the first conjunct and the coordinating particle *w* followed by the standard marker of verbal negation on the second conjunct (and any subsequent conjuncts); *mā ... wala* combining the standard marker of verbal negation on the first conjunct with negative conjunction *wala* (17b); and finally *lā ... wala* which marks negation on the first conjunct using the negative element *lā* combined with the negative conjunction *wala* before the second conjunct (and any subsequent conjuncts), in (17c).

(17) a. huda mā naẓẓaff-at w mā rattib-at l-bēt
Huda.F NEG clean.PFV-3SGF CONJ NEG tidy.PFV-3SGF DEF-house.SGM
Huda did not clean and did not tidy the house.

b. huda mā/lā naẓẓaff-at wala rattib-at
Huda.F NEG/NEG clean.PFV-3SGF NEG.CONJ tidy.PFV-3SGF
l-bēt
DEF-house.SGM
Huda neither cleaned nor tidied the house.

c. huda lā naẓẓaff-at wala rattib-at
Huda.F NEG clean.PFV-3SGF NEG.CONJ tidy.PFV-3SGF
l-bēt
DEF-house.SGM
Huda neither cleaned nor tidied the house.

These three strategies are all equally available to cases of coordination with a shared subject, at the VP and I′ levels, as shown in (18) and (19).

(18) a. mansōr mā akal l-ruz w mā šarab
Mansour.M NEG eat.PFV.3SGM DEF-rice CONJ NEG drink.PFV.3SGM
l-gahwa
DEF-coffee

Mansour did not eat the rice and did not drink the coffee.

b. mansōr mā/lā akal l-ruz wala
Mansour.M NEG/NEG eat.PFV.3SGM DEF-rice NEG.CONJ
šarab l-gahwa
drink.PFV.3SGM DEF-coffee

Mansour neither ate the rice nor drank the coffee.

(19) a. huda mā kān-at ta-lʕab riyāḍa w mā
Huda.F NEG be.PFV-3SGF 3SGF-play.IMPV sport.3SGF CONJ NEG
kān-at t-rūḥ n-nādi
be.PFV-3SGM 3SGF-go.IMPV DEF-gym

Huda didn't either play any sport or go to the gym.

b. huda mā/lā kān-at ta-lʕab riyāḍa wala
Huda.F NEG be.PFV-3SGF 3SGF-play.IMPV sport.SGF NEG.CONJ
(kān-at) t-rūḥ n-nādi
be.PFV-3SGF 3SGF-go.IMPV DEF-gym

Huda didn't either play any sport or go to the gym.

Things are different with coordination at the sentential level. In this case, the pattern seen in (17a), (18a) and (19a) in which *mā* occurs immediately adjacent to the verb in each conjunct, is grammatical, as in (20a). However, the patterns which combine sentence-internal *mā* or *lā* on the first conjunct with *wala* on the second conjunct are ungrammatical, and we find instead that *lā* occurs before the first conjunct. We will return briefly to discussion of IP coordination in section 5.3.

(20) a. mansōr mā gaʕad min n-nōm, w ʕali mā
Mansour.M NEG wake.PFV.3SGM from DEF-sleep, CONJ Ali.M NEG
ǧ-a min d-dawām
come.PFV-3SGM from DEF-work

Mansour did not wake up and nor did Ali come from work.

b. *mansōr mā/lā gaʕad min n-nōm, wala
Mansour.M NEG/NEG wake.PFV.3SGM from DEF-sleep, NEG.CONJ
ʕali ǧ-a min d-dawām
Ali.M come.PFV-3SGM from DEF-work

Mansour did not wake up and nor did Ali come from work.

c. lā mansōr gaʕad min n-nōm, wala ʕali
NEG Mansour.M wake.PFV.3SGM from DEF-sleep, NEG.CONJ Ali.M
ǧ-a min d-dawām
come.PFV-3SGM from DEF-work
Mansour did not wake up and nor did Ali come from work.

4.2 With Non-Verbal Predicates

Negative coordination of non-verbal predicates is grammatical with all three strategies, as shown below. Where *mā* occurred in corresponding verbal sentences in (17) - (19) we find *mū* or its inflected forms.

(21) a. huda mi fī l-bēt wa mi fī d-dawām
Huda.F NEG.3SGF in DEF-house CONJ NEG.3SGF in DEF-work
Huda is not at work and not at home.

b. huda mi/lā fī l-bēt wala fī d-dawām
Huda.F NEG.3SGF/NEG in DEF-house NEG.CONJ in DEF-work
Huda is neither at home nor at work.

(22) a. huda mi ṭuīl-a wa mi giṣīr-a
Huda.F NEG.3SGF tall-SGF CONJ NEG short-SGF
Huda is neither tall nor short.

b. huda mi/lā ṭuīl-a wala giṣīr-a
Huda.F NEG.3SGF/NEG tall-SGF NEG.CONJ short-SGF
Huda is not tall and not short.

4.3 With Nominal Dependents

In sections (4.1) and (4.2) we have seen a number of patterns for expressing sentential or eventuality negation. The possibilities are much more restricted when it comes to the constituent negation of coordinate nominal arguments such as subject and object. Since these are nominal **arguments**, rather than main sentential predicates, neither *mā* nor *mū* are possible marking the coordinate argument; hence the only pattern which arises is that combining *lā* on the first conjunct with *wala* on the second (and any subsequent) conjunct. Parallel to what we saw above in section 3.2 for certain negative words such as SFP *wala*, the negative coordination of arguments with *lā...wala* preverbally is inherently negative (see (23a)) and can combine with sentential negation to give a double negative meaning, as in (23b). Again like SFP *wala*, postverbal negative coordination with *lā...wala* exhibits negative concord (NC) and requires the presence of sentential *mā* (see (24)). The agreement behaviour that we see is the coordination-appropriate pattern for this variety of Arabic — full (resolved) agreement in SVO and both fully resolved

and CCA agreement in VSO (examples (23) and (24) show resolved agreement and (25) illustrates CCA). When we have the CCA agreement pattern with a (negative) coordinate subject it is possible to drop the *lā* marking the first conjunct, as in (25).

(23) a. lā ʔaħmad wala mhammad ǧ-aw
NEG Ahmad.M NEG.CONJ Mohammad.M come.PFV-3PLM
Neither Ahmad nor Mohammad came.

b. lā ʔaħmad wala mhammad mā ǧ-aw
NEG Ahmad.M NEG.CONJ Mohammad.M NEG come.PFV-3PLM
Neither Ahmad nor Mohammad didn't come.
(= Both Ahmad and Mohammad came.)

(24) *(mā) ǧ-aw lā ʔaħmad wala ʕali
NEG come.PFV-3PLM NEG Ahmad.M NEG.CONJ Ali.M
Neither Ahmad nor Ali came.

(25) mā ǧ-at (lā) huda wala ʕali
NEG come.PFV-3SGF NEG Huda.F NEG.CONJ Ali.M
Neither Huda nor Ali came.

The same positional dependent alternation between NEG in (26b) and NC (in 26a) readings arises with non-subject arguments to verbs, as illustrated in (26). Negative coordination of arguments to non-verbal predicates such as the pseudo-verb *ʕind* 'have' is parallel in all respects, as in (27a) and (27b).

(26) a. ʕali mā šarab lā gahwa wala šāy l-yōm
Ali.M NEG drink.PFV.3SGM NEG coffee NEG.CONJ tea DEF-day
Ali has drunk neither coffee nor tea today.

b. lā gahwa wala šāy šarab ʕali l-yōm
NEG coffee-SGF NEG.CONJ tea.SGM drink.PFV.3SGM Ali.M DEF-day
Ali has drunk neither coffee nor tea today.

(27) a. mā ʕind-i (lā) raχṣ-a wala sayyār-ah
NEG have-1SG.GEN NEG license-SGF NEG.CONJ car-SGF
I have neither a license nor a car.

b. lā raχṣ-a wala sayyār-ah ʕind-i
NEG license-SGF NEG.CONJ car-SGF have-1SG.GEN
I have neither a license nor a car.

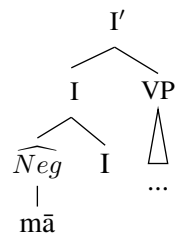
5 Sentential Negation and Negative Coordination: Analysis

We start by considering the *mā* .. *w mā* .. pattern illustrated in (17a), (18a) and similar examples. In these examples the SUBJ is outside the coordinate structure (in terms of c-structure) and distributed in (in terms of f-structure). Sentential negation is independently marked in each conjunct by the negative particle *mā*, and the conjunction *w* defines CONJTYPE as AND and CONJFORM as W for the coordinate structure as a whole.

(28)
$$\begin{bmatrix} \text{CONJTYPE} & \text{AND} \\ \text{CONJFORM} & \text{W} \\ \left\{ \begin{bmatrix} \text{PRED} & \text{`EAT}< \text{SUBJ, OBJ} > \text{'} \\ \text{ENEG} & + \\ \text{SUBJ} & \boxed{1}\ \begin{bmatrix} \text{PRED} & \text{`MANSOUR'} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`SWEETS'} \end{bmatrix} \end{bmatrix} \right. \\ \left. \begin{bmatrix} \text{PRED} & \text{`DRINK}< \text{SUBJ, OBJ} > \text{'} \\ \text{ENEG} & + \\ \text{SUBJ} & \boxed{1} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`COFFEE'} \end{bmatrix} \end{bmatrix} \right\} \end{bmatrix}$$

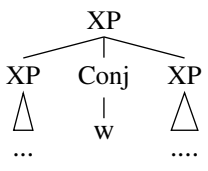
The negative particle *mā* is obligatorily adjacent to the verb (and is a morphologically bound form in some vernaculars). We treat it as a non-projecting word adjoined to I and defining ENEG = +. For the conjunction *w*, two possible analyses are plausible. We adopt the flat structure in (31a) as the more standard assumption. The alternative would be the structure shown in (31b) in which the conjunction forms a constituent with the following conjunct.

(29)

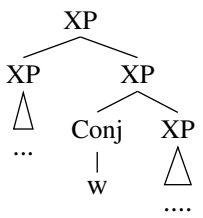


(30) *mā* $\widehat{\text{Neg}}$ (↑ ENEG) = +

(31) a.



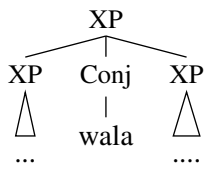
b.



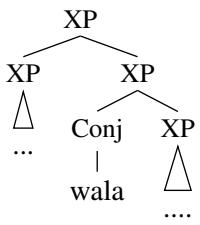
5.1 The wala conjunct

Consider now coordination *wala* as in (18b) and other similar examples above. Coordination *wala* (distinct from SFP *wala*) occurs only before non-initial conjuncts, expresses coordination and contributes negation to the following conjunct. Again there are two possible structures, differing in whether *wala* forms a constituent with the second conjunct or occurs at the level of the coordinate structure as a whole.

(32) a.

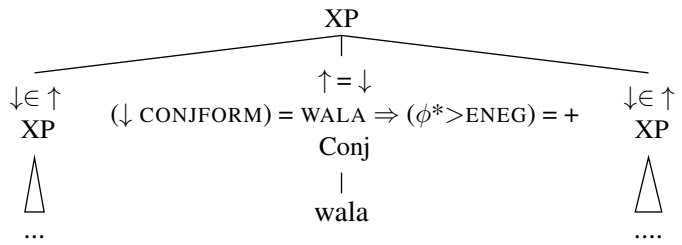


b.



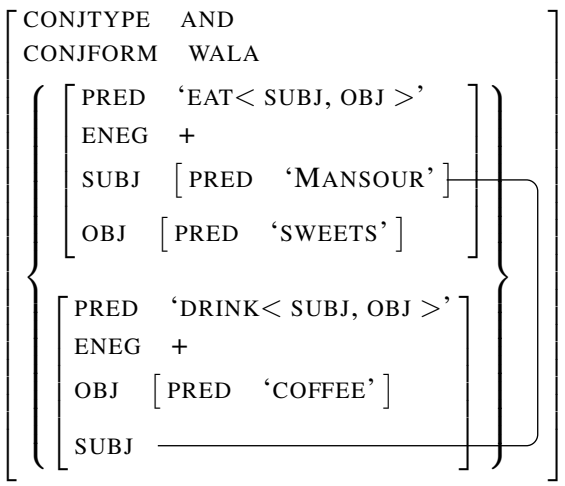
In a flat structure we would require annotations along the lines shown in (33), where the conditional A $\Rightarrow$ B $\equiv_{df}$ $\neg$ A $\vee$ (A$_c$ $\wedge$ B) (Bresnan et al. (2015, 64) originally proposed in Andrews and Manning (1999)), and where *> denotes the right sister of a node and ϕ* > the f-structure of that node (Dalrymple, 2001, 120). The f-description ($\downarrow$ CONJFORM) = WALA $\rightarrow$ (ϕ*> ENEG) = + assigns ENEG = + to the right sister (the following conjunct) provided that the CONJFORM of the coordinate structure as a whole is CONJFORM =$_c$ WALA. This in turn is provided by the lexical description of conjunction *wala*, in (34). The conjuncts themselves have the standard $\downarrow \in \uparrow$ annotation. The features CONJFORM and CONJTYPE are non-distributive; when a non-distributive feature is definied on a set the attribute and its value is a property of the set as whole: for example, the f-description ($\uparrow$ CONJFORM) = WALA in (34) defines the CONJFORM value of the coordinate structure as a whole, as shown in (35). See Dalrymple (2001, 156-158) for the distinction between distributive and non-distributive features.

(33)



(34) *wala* Conj ($\uparrow$ CONJFORM) = WALA
($\uparrow$ CONJTYPE) = AND

(35)



In the case of non-binary negative coordination, as stated in (36) this requires all conjunctions to be *wala* (which corresponds to the facts).

(36) *Iterating Coordination Schema*

$$\begin{array}{lccc} \text{XP} \longrightarrow & \text{XP} & (\ \text{Conj} & \text{XP}\ )^{+} \\ & \downarrow \in \uparrow & \uparrow = \downarrow & \downarrow \in \uparrow \end{array}$$

$$(\downarrow \text{CONJFORM}) = \text{WALA} \rightarrow \phi^{*} > (\text{ENEG}) = +$$

The approach outlined above does seem to permit an analysis of the appropriate facts, though perhaps at a cost of a certain amount of technical machinery.[6] Among the drawbacks of this approach (with a flat c-structure) however, are that *wala* cannot lexically define its conjunct's ENEG feature using the notation $\phi\ *\ >$ because it has no sister. While the intuition is that *wala* directly contributes ENEG information, this information is introduced constructionally.[7]

We now consider an alternative analysis using (32b). On this approach, other coordinate structures involve the flat coordination structure (so *wala* must be excluded from this), but *wala* coordination (alone) involves the special coordination schema in (37), which must be limited to this type of coordination.The c-structure rule for the conjunct XP is shown in (38): the inside-out f-description ($\in \uparrow$) ensures that the f-structure of the XP is a member of a set. Treating ENEG as an instantiated (and non-distributive) feature will ensure that (38) applies only once in each conjunct. The element which we have called coordination *wala* (to distinguish it from SFP *wala*) specifies both negative and coordinative information in f-structure,

[6] As given above, this approach actually permits the first conjunct to be either affirmative or negative, but in the general case, both (all) conjuncts are negative if *wala* is used. To rule out coordination of this type with mixed polarity across the conjuncts, a further condition could be added to the f-description of the Conj node.

[7] An alternative possibility, still maintaining the flat c-structure, is that this is an instance of lexical sharing, involving a Conj node and a Neg node initial within the following conjunct, however this also requires us to make provision for a special negative *lā* which is not found outside of negative coordination.

and so the question arises as to whether it is categorially a Conj or a Neg. In (38) we have treated it categorially as a Neg element, as (38) is potentially also appropriate for negative incidental adjuncts, which we cannot discuss here. For the coordination data, it would also be possible to treat *wala* NEG.CONJ categorially as a conjunction.

(37) *Negative Coordination Schema*

$$\begin{array}{lcccc} \text{XP} & \longrightarrow & \text{XP} & & \text{XP}^{+} \\ & & \downarrow \in \uparrow & & \downarrow \in \uparrow \\ & & (\downarrow \text{ENEG}) =_c +_ & & (\downarrow \text{CONJFORM}) =_c \text{WALA} \\ & & (\downarrow \text{CONJFORM}) \neq \text{WALA} & & \end{array}$$

(38) $\text{XP} \longrightarrow \begin{array}{cc} \text{Neg} & \text{XP} \\ \uparrow = \downarrow & \uparrow = \downarrow \\ (\in \uparrow) & \end{array}$

(39) *wala* Neg
$(\uparrow \text{CONJFORM}) = \text{WALA}$
$(\uparrow \text{ENEG}) = +_$
$((\in \uparrow) \text{CONJTYPE}) = \text{AND}$

The analysis of (18b) (the variant with *mā*) in this approach is as follows. In the first conjunct ENEG = +_ is contributed by *mā*, a non-projecting word introduced as sister to the verbal element (see (29)): the CONJFORM annotation on the first daughter of (37) prevents *wala* occurring in this conjunct. The lexical entry for *mā*, revised to treat ENEG as an instantiated feature, is shown in (40). (37) requires the second conjunct to have the feature CONJFORM = WALA which is satisfied by adjunction of *wala* using (38). The f-structure is shown in (41).

(40) $m\bar{a}\ \widehat{\text{Neg}}$ $(\uparrow \text{ENEG}) = +_$

(41)
$$\left[\begin{array}{ll} \text{CONJTYPE} & \text{AND} \\ \left\{\begin{array}{l} \left[\begin{array}{ll} \text{PRED} & \text{`EAT}\!<\text{SUBJ, OBJ}\!>\text{'} \\ \text{ENEG} & +_ \\ \text{SUBJ} & \left[\text{PRED} \quad \text{`MANSOUR'}\right] \\ \text{OBJ} & \left[\text{PRED} \quad \text{`RICE'}\right] \end{array}\right] \\ \left[\begin{array}{ll} \text{PRED} & \text{`DRINK}\!<\text{SUBJ, OBJ}\!>\text{'} \\ \text{ENEG} & +_ \\ \text{CONJFORM} & \text{WALA} \\ \text{OBJ} & \left[\text{PRED} \quad \text{`COFFEE'}\right] \\ \text{SUBJ} & \end{array}\right] \end{array}\right\} \end{array}\right]$$

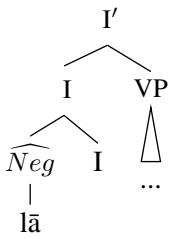
5.2 Status of *lā*

We can now turn to the status of *lā*, the special marker of negation which occurs only on the initial conjunct of a coordinate phrase in examples such as (42).

(42) mansōr lā akal l-ruz wala šarab
Mansour.M NEG eat.PFV.3SGM DEF-rice NEG.CONJ drink.PFV.3SGM
l-gahwa
DEF-coffee
Mansour neither ate the rice nor drank the coffee.

In Turaif Arabic (and other vernaculars), *lā* marks negation in the initial conjunct of negative coordination and provides some additional emphatic, focussing, or related information as compared to the counterpart sentences with *mā* or *mū* on the first conjunct (this is not dissimilar to the choice between *not A or B* and *neither A nor B* in English). Beyond this use in coordination, *lā* only occurs (vestigially) in fixed collocations, and as part of the negative imperative (prohibitive). In Classical Arabic and MSA, on the other hand, *lā* appears as a marker of sentential negation in a position immediately adjacent to the imperfective form of the verb, hence in a structure similar to (29), (see (43)).

(43)



While *lā* in Turaif Arabic may appear immediately adjacent to the verb (as it does in (42)) it is **not** restricted to this position and so does not share the positional restrictions of its CA/MSA cognate. It may occur initially (before the subject) in the negative coordination of sentences, as shown in (44). While *mā* is a non-projecting $\widehat{\text{Neg}}$ word immediately adjoined to the verb in I, the syntax of *lā* is like that of coordination *wala*: it combines with a following phrase (including an IP) in accordance with (38). The proposed lexical description for *lā* is shown in (45).

(44) lā mansōr gaʕad min n-nōm, wala ʕali
NEG Mansour.M wake.PFV.3SGM from DEF-sleep, NEG.CONJ Ali.M
ǧ-a min d-dawām
come.PFV-3SGM from DEF-work
Neither did Mansour wake up nor Ali come (home) from work.

(45) *lā* Neg (↑ CONJFORM) = LĀ
(↑ ENEG) = +_
((∈ ↑) CONJTYPE) = AND

(46) *Negative Coordination Schema* where XP ≡ { IP | I′ | VP | AP | PP}

XP ⟶	XP	XP^{+}
	$\downarrow \in \uparrow$	$\downarrow \in \uparrow$
	$(\downarrow$ ENEG$) =_c +_-$	$(\downarrow$ CONJFORM$) =_c$ WALA
	$(\downarrow$ CONJFORM $) \neq$ WALA	

5.3 Further Issues

The analysis of cases in which the main sentential predicate in each conjunct is non-verbal will follow straightforwardly from the above, given an appropriate sentential analysis for cases of non-verbal predication in Arabic. In these sentence types, ENEG can be marked by *mū* and its variants or *lā* on the initial conjunct and by *mū* and its inflectional variants or *wala* on the non-initial conjuncts. If *lā* is used, then *wala* is required on subsequent conjuncts. However there is a remaining issue concerning negative coordination of full IPs (see the data in (20). If (46) applies to IP (as stated above), then it will additionally (and incorrectly) permit *lā* and *mā* in clause internal position in the first conjunct (the ungrammatical pattern in (20b). One possibility (which we do not explore further here) is that there are *additional* linearisation constraints which require the NEG element to be initial in each conjunct. Another possibility is that negative coordination of IPs is excluded from (46) and falls instead under the rule for saturated arguments discussed in section 6 below.

6 Negative Coordination of Dependents

We can now turn to the negative coordination of dependents, illustrated by an example such as (47). There are considerable reasons for concluding that the syntactic f-structure analysis of such examples should reflect rather directly the external syntactic manifestation which involves the coordination within the dependent, with each conjunct showing constituent negation.[8] These are that the pattern of agreement between the subject and the predicate is consistent with the pattern we find with conjunction rather than disjunction in Turaif Arabic, and the combination of (preverbal) negative coordination of dependents with the expression of predicate negation gives rise to a double negative reading, shown in (48). Accordingly we take the f-structure for (47) to be as in (49), with (50) for the 'double negative' reading in (48).

(47) lā ʔaħmad wala mhammad ǧ-aw
NEG Ahmad.M NEG.CONJ Mohammad come.PFV-3PLM
Neither Ahmad nor Mohammad came.

[8]That is, in contrast to closely mirroring the interpretation. We leave matters of interpretation to one side here, but note that the interpretation (at least the most salient) is ¬ came(Ahmad) ∧ ¬ came(Mohammad) or equivalently, ¬ (P ∨ Q).

(48) lā ʔaħmad wala mhammad mā ǧ-aw
NEG Ahmad.M NEG.CONJ Mohammad NEG come.PFV-3PLM
Neither Ahmad nor Mohammad didn't come.
(= Both Ahmad and Mohammad came.)

(49)
$$\begin{bmatrix} \text{PRED} & \text{`COME}\langle\ \text{SUBJ}\ \rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix} \text{CONJTYPE} & \text{AND} \\ \left\{ \begin{bmatrix} \text{PRED} & \text{`AHMAD'} \\ \text{CNEG} & +_{_} \\ \text{CONJFORM} & \text{LĀ} \end{bmatrix}, \begin{bmatrix} \text{PRED} & \text{`MOHAMMAD'} \\ \text{CNEG} & +_{_} \\ \text{CONJFORM} & \text{WALA} \end{bmatrix} \right\} \end{bmatrix} \end{bmatrix}$$

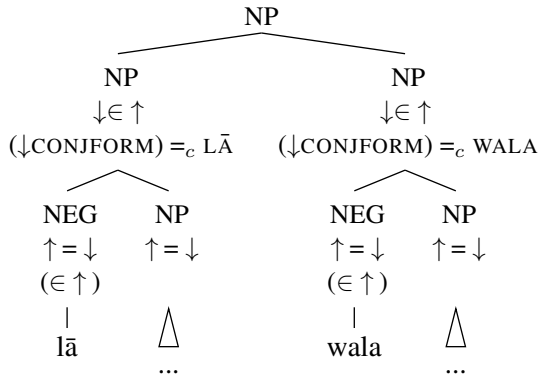
(50)
$$\begin{bmatrix} \text{PRED} & \text{`COME}\langle\ \text{SUBJ}\ \rangle\text{'} \\ \text{ENEG} & + \\ \text{SUBJ} & \begin{bmatrix} \text{CONJTYPE} & \text{AND} \\ \left\{ \begin{bmatrix} \text{PRED} & \text{`AHMAD'} \\ \text{CNEG} & +_{_} \\ \text{CONJFORM} & \text{LĀ} \end{bmatrix}, \begin{bmatrix} \text{PRED} & \text{`MOHAMMAD'} \\ \text{CNEG} & +_{_} \\ \text{CONJFORM} & \text{WALA} \end{bmatrix} \right\} \end{bmatrix} \end{bmatrix}$$

If we are right about this, then we need lexical descriptions for *lā* and *wala* in their CNEG incarnation, alongside the lexical descriptions which are motivated by the use of these conjunctions in sentential negation (adjoined to verbs and pseudo-verbs and their projections), if we maintain the assumption that ENEG and CNEG are distinct attributes (rather than instances of the same attribute in different f-structures). So in addition to (45) (for *lā*) and (39) (for *wala*) we postulate (51) and (52), alongside a version of the *Negative Coordination Schema* for dependents, in (53). Note that there are real distributional differences between *wala* and *lā* in their (clausal) predicate negating and argument negating functions: in the latter function we require *lā* on the first conjunct, whereas in the former negation can be realised in a variety of different ways, as we have seen.

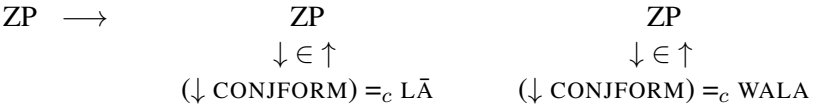
(51) *lā* Neg $(\uparrow \text{CONJFORM}) = \text{LĀ}$
$(\uparrow \text{CNEG}) = +_{_}$
$((\in \uparrow)\ \text{CONJTYPE}) = \text{AND}$

(52) *wala* Neg (↑ CONJFORM) = WALA
(↑ CNEG) = +_
((∈ ↑) CONJTYPE) = AND

(53)



(54) *Negative Coordination Schema: Dependents*
where ZP ≡ { NP |DP | PP}



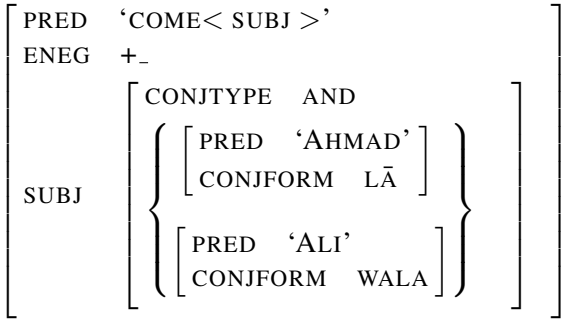
6.1 Negative Coordination of Dependents: Negative Concord

We now turn briefly to the question of *lā...wala* nominal (and prepositional) dependents as negative concord elements. While *lā...wala* marking of dependents in preverbal position marks negation (as we have seen above), in the postverbal position *lā...wala* constitutes an instance of (non-strict) negative concord. Because position is a crucial factor, there is an apparently irreducible syntactic component to the phenomenon of negative concord (which arises postverbally).

(55) mā ğ-aw lā ʔaħmad wala ʕali
NEG come.PFV-3PLM NEG Ahmad.M NEG.CONJ Ali.M
Neither Ahmad nor Ali came.

In (55), *mā* contributes ENEG = +_, and *lā* and *wala* are NC items, so they do not contribute CNEG = +_ to their respective f-structures:

(56)



Przepiórkowski and Patejuk (2015) briefly outlines an approach to the (strict) negative concord items in Polish *nikt* 'nobody.NOM' and its inflectional counterparts, which occur in the context of the marker of sentential negation *nie*, in (57) (NW stands for n-word). They associate an inside-out constraint with these NC items which requires ENEG to be defined as + in the appropriate containing f-structure, as shown for *nikt* 'nobody.NOM' in (58).

(57) Nikt nie lubi nikogo.
nobody.NW.NOM NEG likes nobody.NW.GEN
Nobody likes anybody. Polish: Przepiórkowski and Patejuk (2015, 330)

(58) *nikt (nobody)* N $(\uparrow \text{CASE}) = \text{NOM}$
$((\text{XCOMP}^* \ \text{GF}^+ \in \uparrow) \ \text{ENEG}) =_c +$

For the non-strict NC element *lā ... wala*, we need to treat the negative coordination of an argument as introducing CNEG or as a case of NC depending on its position with respect to the verb and the expression of sentential negation (non-strict NC language). The NC interpretation arises if there is ENEG in the clause **and** the marker of ENEG precedes the conjunctive negative markers (*lā* and *wala*). The interpretation as a marker of constituent negation (CNEG) arises if there is no marker of ENEG **and** no marker of TNS which f-precede the conjunctive negative markers. To capture the precedence relations we need **both** the values of the ENEG and the TNS feature to take a position in the f-precedence relation *independent* of the larger (sentential) f-structure. The lexical description for the dependent-marking *wala*, taking account of the fact that it occurs as a marker of CNEG in some circumstances and as an NC marker in other circumstances, would then be along the lines shown in (59) replacing (52). In both negative and NC uses, *wala* defines CONJFORM and CONJTYPE features (first two equations in (59)). Alongside this, either it defines the CNEG feature to be positive (under certain f-precedence conditions, namely when the f-structure it which it appears as an attribute is not f-preceded by either the marker of ENEG or that of TENSE) or the NC feature to be positive (under distinct conditions, namely when it f-precedes these same elements).[9] We use a feature NC here essentially for expository convenience (it would be possible to introduce the appropriate conditions without this feature), but in any case such a feature might eventually turn out to play a role in guiding the mapping to the semantics.

(59) wala *Neg* $(\uparrow \text{CONJFORM}) = \text{WALA} \qquad ((\in \uparrow) \ \text{CONJTYPE}) = \text{AND}$
$\{(\uparrow \text{CNEG}) = +_- \wedge ((\text{GF}^+ \in \uparrow) \text{ENEG}) \ {}_f\not\prec \uparrow \wedge ((\text{GF}^+ \in \uparrow) \ \text{TNS}) \ {}_f\not\prec \uparrow \ |$
$(\uparrow \text{NC}) = + \wedge ((\text{GF}^+ \in \uparrow) \text{ENEG}) =_c +_- \wedge ((\text{GF}^+ \in \uparrow) \text{ENEG}) \ {}_f\prec \uparrow \}$

[9]Note that we assume here that both the values of ENEG and that of the TENSE feature to take a position in the f-precedence relation *independent* of the larger, sentential f-structure.

A Further Option: Note that the meaning of a sentence such as (60) is not equivalent to that of an *and* coordination in the dependent under the scope of sentential negation. That is, it does not correspond to ¬ (P ∧ Q), where P is *drank(ali, coffee)* and Q is *drank(ali, tea)*, but rather it corresponds to meaning ¬ (P ∨ Q). In the light of this we might consider an alternative approach to the dependent data, separating the conjunctive CNEG reading from a disjunctive NC reading in the entries for *wala* and *lā*, leading to entries such as (61) and (62).

(60) ʕali mā šarab lā gahwa wala šāy l-yōm
Ali.M NEG drink.PFV.3SGM NEG coffee NEG.CONJ tea DEF-day
Ali has drunk neither coffee nor tea today.

(61) *wala* Neg
$(\uparrow \text{CONJFORM}) = \text{WALA}$
$((\in \uparrow) \text{CONJTYPE}) = \text{AND}$
$(\uparrow \text{CNEG}) = +_{-}$
$((\text{GF}^{+} \in \uparrow) \text{ENEG})_{f} \not\prec \uparrow$
$((\text{GF}^{+} \in \uparrow) \text{TENSE})_{f} \not\prec \uparrow$

(62) *wala* Neg
$(\uparrow \text{CONJFORM}) = \text{WALA}$
$((\in \uparrow) \text{CONJTYPE}) = \text{OR}$
$(\uparrow \text{NC}) = +$
$((\text{GF}^{+} \in \uparrow) \text{ENEG}) =_{c} +_{-} \wedge ((\text{GF}^{+} \in \uparrow) \text{ENEG})_{f} \prec \uparrow$

However, we note that it is possible to have full (resolved) agreement in VSO order with a NC-marked *lā...wala* SUBJ. This is consistent with conjunction, but disjunctive agreement controllers give rise to a single conjunct pattern. Although this is not conclusive evidence, we do not propose to follow this alternative.

7 Conclusion

We have considered the analysis of the bisyndetic negative coordination strategies in vernacular Arabic, on the basis of data from Turaif Arabic and in particular the combination of *lā* (which does not otherwise occur as a marker of sentential negation) with *wala*, which also has a SFP use. We have argued that *wala* and *lā* in these negative coordinate constructions both negate individual conjuncts and also contribute CONJTYPE information to the coordinate structure as a whole.

We have shown that when *lā... wala* is used in the coordination of dependents (rather than predicates), it gives rise to either a negative reading or a negative concord reading. The conditions under which these interpretations arise are parallel to those for other items in Arabic which show an alternation between a negative and a NC reading, including the element *wala* used as a SFP. The diachronic development of a SFP by compounding a conjunction with a negative marker is attested in a number of languages, and the Arabic facts appear similar in some respects to these cases (see e.g. Gajić (2016) (Serbian), Herburger (2003) (Spanish), Gianollo

(2017) (Latin), Haspelmath (1997) and especially Hoyt (2010) for a discussion of Palestinian Arabic, in which he proposes that the SFP weak-*wala* may have developed out of constructions in which a final disjunct closes off a set of alternatives).

References

Andrews, Avery and Chris Manning. 1999. *Complex Predicates and Information Spreading in LFG*. Stanford, CA: CSLI Publications.

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

Dalrymple, Mary. 2001. *Lexical Functional Grammar*. San Diego, CA: Academic Press.

Gajić, Jovana. 2016. Ni-disjunction as a coordination marker and focus particle. In *Proceedings of ESSLLI 2016 Student Session*, pages 155–166.

Gianollo, Chiara. 2017. Focus-sensitive negation in Latin. *Catalan Journal of Linguistics* 16:51–77.

Haspelmath, Martin. 1997. *Indefinite Pronouns*. Oxford: Oxford University Press.

Haspelmath, Martin, ed. 2004. *Coordinating Constructions*. Amsterdam; Philadelphia: John Benjamins.

Haspelmath, Martin. 2007. Coordination. In T. Shopen, ed., *Language typology and syntactic description*, pages 1–51. Cambridge, UK: Cambridge University Press, 2nd edn.

Herburger, Elena. 2003. A note on Spanish 'ni siquiera', 'even' and the analysis of NPIs. *Probus* 15:237–256.

Hoyt, Frederick MacNeill. 2010. *Negative Concord in Levantine Arabic*. Ph.D. thesis, University of Texas at Austin, Austin,Texas.

Laczkó, Tibor. 2014. Outlines of an LFG-XLE account of negation in Hungarian sentences. In M. Butt and T. H. King, eds., *Proceedings of the LFG14 conference*.

Laczkó, Tibor. 2015. On negative particles and negative polarity in Hungarian. In M. Butt and T. H. King, eds., *Proceedings of the LFG15 Conference*, pages 166–186. Stanford, CA: CSLI Publications.

Lucas, Christopher. 2009. *The development of negation in Arabic and Afro-Asiastic*. Ph.D. thesis, University of Cambridge.

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

Ryding, Karin. 2005. *A Reference Grammar of Modern Standard Arabic*. Cambridge, UK: Cambridge University Press.

Sells, Peter. 2000. Negation in Swedish, where it's not at. In M. Butt and T. H. King, eds., *Proceedings of LFG00*. Stanford, CA: CSLI Publications: http://www-csli.stanford.edu/publications.

Catalan Intransitive Verbs and Argument Realization

Alex Alsina
Universitat Pompeu Fabra

Fengrong Yang
Universitat Pompeu Fabra

Proceedings of the LFG'18 Conference

University of Vienna

Miriam Butt, Tracy Holloway King (Editors)

2018

CSLI Publications

pages 46–66

http://csli-publications.stanford.edu/LFG/2018

Keywords: intransitive verbs, argument realization, mapping theory, verb agreement

Alsina, Alex, & Yang, Fengrong. 2018. Catalan Intransitive Verbs and Argument Realization. In Butt, Miriam, & King, Tracy Holloway (Eds.), *Proceedings of the LFG'18 Conference, University of Vienna*, 46–66. Stanford, CA: CSLI Publications.

Abstract

The goal of this paper is to analyze the behavior of the single direct argument of intransitive verbs in Catalan, including its encoding as a grammatical function, verbal agreement, case assignment, and expression by means of clitics. Our main claim is that the single direct argument of a clause can be a nominative object. We show that the direct argument of intransitive verbs (whether unaccusative or unergative) alternates between subject and object. The proposed analysis diverges from standard versions of LFG, as it allows an external argument to map onto an object and allows a clause to lack a subject, in violation of the Subject Condition. We propose a new mapping theory in which case assignment plays a major role and account for the agreement facts by assuming a set of agreement features of the clause (AGR) that are identified with a grammatical function (GF), not necessarily the subject, by general constraints.

1 Introduction

The topic of this paper is the behavior of the single direct argument[1] of intransitive verbs (the *intransitive argument*, for short). The relevant facts are presented in section 2, showing that the behavior of that argument is split between subject and object. The argument realization theory needed to account for these facts is proposed in section 3, where case assignment plays a crucial role in constraining the mapping of arguments to grammatical functions. The agreement facts are discussed and explained in section 4, adopting the theory of agreement proposed by Alsina and Vigo (2014, 2017). The main conclusions are summarized in section 5.

2 Properties of the sole argument of intransitive verbs

The intransitive argument behaves in some ways like a subject and in some ways like an object. We start by showing its object properties, in 2.1; then, turn to its subject properties, in 2.2, focusing on the agreement facts in 2.3.

2.1 Object properties

En cliticization provides evidence that the intransitive argument can be an object in Catalan. (Other Romance languages, such as Italian and French, show a similar behavior of the cognate clitic *en* or *ne*.) The internal argument of Catalan transitive verbs can be partially or totally expressed by means of the clitic *en*:[2] *en* in (1a) and (1b) replaces *carpetes* 'folders' and *carpetes de plàstic*

† We deeply acknowledge the comments and observations made by the anonymous reviewers and the audience of the 23th LFG conference. Any remaining errors are our own.

[1] By *direct argument* we refer to an argument whose default expression is as a direct grammatical function with an unmarked case feature (nominative or accusative).

[2] Here we are only concerned with one of the two functions of the clitic *en*, which we may call 'partitive' *en*, as it replaces the head noun of an indefinite object and cannot cooccur with it. In the other function—'genitive' *en*—the clitic corresponds to a *de*-complement of the verb or of the verb's object, as shown in (i), where *en* corresponds to the verb's *de*-complement:

noves 'new plastic folders', respectively:

(1) a. Si vols carpetes, en tinc tres de noves.
if want.2p.sg *folder*.f.pl *en*.cl *have*.1p.sg *three of new*.f.pl
'If you need folders, I have three new ones.'

b. Si vols carpetes de plàstic noves, compra'n.
if want.2p.sg *folder*.f.pl *of plastic new*.f.pl *buy*.imp.2p.sg-*en*.cl
'If you need new plastic folders, buy some.' (Alsina 1986:97-98)

The internal argument of Catalan unaccusative verbs patterns with the internal argument of transitive verbs in terms of the *en* cliticization:

(2) Cada dia surten molts trens,
every day leave.pl *many*.m.pl *train*.m.pl
però avui només n'ha sortit un.
but today only en.cl-*have*.sg *leave*.pp *one*
'Everyday many trains leave, but today only one has left.'

Surprisingly, although Catalan transitive verbs do not allow their external arguments to be cliticized by *en*, as in (3), the external argument of unergative verbs nevertheless can be replaced by the *en* clitic, as in (4):[3]

(3) a. * N'aprovaran tres els exàmens.
en.cl-*pass*.fut.pl *three the*.m.pl *exam*.m.pl

b. * N'aprovaran els exàmens tres.
en.cl-*pass*.fut.pl *the*.m.pl *exam*.m.pl *three*
'Three of them will pass the exams.'

(4) a. En ploraran sis quan sàpiguen la veritat.
en.cl *cry*.fut.pl *six when know*.sbjv.pl *the*.f.sg *truth*.f.sg
'Six of them will cry when they find out the truth.'
(Cortés and Gavarró 1997:41)

b. - Com repartirem les conferències?'
how distribute.fut.1p.pl *the*.f.pl *conference*.f.pl
'How should we arrange the conferences?
- Avui en poden parlar dos i demà tres més.
today en.cl *can*.3p.pl *talk*.inf *two and tomorrow three more*
'Today two of them can give a talk and tomorrow three.'
(Gràcia 1989:82)

The possibility of *en* cliticization with unergative verbs in Catalan shows that

---

(i) Podria parlar avui d'aquest problema, però en parlarà demà.
could.3p.sg *speak today of-this problem but en*.cl *speak tomorrow*
'He could speak about this problem today, but he will speak about it tomorrow.'

[3] It has sometimes been claimed that only unaccusative verbs allow *en* cliticization. Here we are describing the facts of speakers who accept *en* cliticization with unergatives as well as with unaccusatives, like Cortés and Gavarró (1997) for Catalan, or Saccon (1995) for Italian. Independent evidence for the claim that *plorar* 'cry' in (4a) and *parlar* 'talk' in (4b) are unergatives comes from tests such as the participial adjunct test in Cortés and Gavarró (1997). Note that *poden* 'can' in (4b) is a restructuring verb, which inherits the argument structure of the dependent verb.

it is not the 'deep object', i.e., the internal argument, that triggers *en* cliticization. Instead, the fact that both unaccusative and unergative verbs allow their single direct argument to be expressed by means of *en* requires assuming that the argument in question is an object (or the 'surface object', in theories like Burzio 1986, or Cortés and Gavarró 1997, among others).[4]

The second argument for the object status of the intransitive argument is past participle agreement. In Catalan, the past participle optionally agrees in gender and number with a third person object clitic, when cooccurring with the perfective auxiliary *haver* 'have'. But this agreement does not happen with a full NP object:

(5) a. La directora ha defensat/*defensada la proposta.
the director have.sg *defend*.pp.m.sg/*f.sg *the*.f.sg *proposal*.f.sg
'The director has defended the proposal.'

b. La directora **l**'ha defensa**da**.
the.f.sg *director*.f.sg *la*.cl.f.sg-*have.sg defend*.pp.f.sg
'The director has defended it.' (Alsina 1996:95)

Past participle agreement is not only possible with objects of transitive verbs, like the one in (5b), but also with the direct argument of intransitive verbs:[5]

(6) a. Perquè aleshores hi haurà una gran tribulació,
because then hi.cl *have*.fut.sg *one*.f.sg *great distress*.f.sg
com no **n**'hi ha hagu**da** cap des de
like not en.cl-*hi*.cl *have*.3p.sg *have*.pp.f.sg *never from*
la creació del món…
the creation of-the world
'For then there will be great distress, as there has not been one since the creation of the world...' (Bible [Mt 24:21][6])

b. **N**'han arribat**s** molt**s**.
en.cl-*have*.pl *arrive*.pp.m.pl *many*.m.pl
'Many have arrived.' (Fabra 1912:160)

The fact that an intransitive argument expressed as the clitic *en* can trigger past participle agreement further confirms that the argument is an object.

The possibility of expressing the intransitive argument as a bare indefinite NP gives additional evidence for the objecthood of this argument. Bare indefinite NPs, which have a non-specific interpretation, can encode the object of a transitive verb, as shown in (1a). However, they cannot be the subject of the verb, as illustrated in (7) with a transitive verb:

---

[4] Notice that, with respect to the phenomena examined in section 2, there is no difference in behavior among one-argument verbs between unaccusatives and unergatives in Catalan, and we refer to this distinction precisely to make this point.

[5] We have not documented past participle agreement with *en* with unergatives, although it is expected to be possible. This may be due to the fact that this construction is infrequent and formal, and not used by many speakers. We leave it to further research to decide whether unergatives are excluded from this construction.

[6] http://www.biblija.net/biblija.cgi?m=Mt+24%2C1-31&l=ca, visiting time: 18:19, 08/07/2018

(7) a. * Arreglen mecànics el teu cotxe.
fix.pl *mecanics*.m.pl *the*.m.sg *your*.m.sg *car*.m.sg
b. * Arreglen el teu cotxe mecànics.
fix.pl *the*.m.sg *your*.m.sg *car*.m.sg *mecanics*.m.pl
'Mechanics fix your car.' (Alsina 1996:104)

By contrast, the intransitive argument can freely be expressed as a bare NP:

(8) a. Cau aigua de la teulada.
fall.sg *water*.f.sg *from the*.f.sg *roof*.f.sg
'Water is falling from the roof.' (Alsina 1995:13)
b. Treballen nens en aquesta fàbrica.
work.pl *child*.m.pl *in this*.f.sg *factory*.f.sg
'Children work in that factory.' (Cortés 1995:64)

The contrast between examples (7) and (8) indicates that both *aigua* 'water' in (8a) and *nens* 'children' in (8b) are objects and not subjects. The evidence from bare NPs, together with *en* cliticization and optional past participle agreement, indicates that the intransitive argument is an object.

2.2 Subject properties

Catalan is known to be a subject pro-drop language: in Catalan, a subject can be null and be interpreted as having a definite referent, whereas an object cannot be null with a definite reading:

(9) a. Els estudiants solen sortir puntualment,
the.m.pl *student*.m.pl *be-used-to*.pl *leave*.inf *punctually*
però avui Ø surten tard.
but today leave.pl *late*
'Students usually leave on time, but today they are leaving late.'
b. Els estudiants no volen estudiar habitualment,
the.m.pl *student*.m.pl not *want*.pl *study*.inf *usually*
però avui Ø estudien molt.
but today study.pl *a-lot*
'Students usually do not want to study, but today they are studying a lot.'

(10) Joan ha llegit el diari avui,
John have.3p.sg *read*.pp.m.sg *the*.m.sg *newspaper*.m.sg *today*
però no llegirà demà.
but not read.fut.3p.sg *tomorrow*
'John has read the newspaper today, but will not read (*it) tomorrow.'

The contrast between (9) and (10) shows that grammatical functions other than the subject in Catalan cannot be null with a definite reading. Therefore, the fact that the intransitive argument in Catalan can be omitted and have a definite referent, as in (9), requires analyzing it as the subject of the clause.

Another subject property is the possibility of being the controlee in a control

construction, because only the subject of the embedded clause can be controlled by the subject or object of the matrix clause, as shown in (11) for a transitive verb in an embedded clause:

(11) N'he obligat molts a examinar el metge.
en.cl-*have*.1p.sg *obligate*.pp *many*.pl *to examine*.inf *the doctor*
'I have obligated many to examine the doctor.'
* 'I have obligated many to be examined by the doctor.'

In contrast, as the object of the embedded clause, the intransitive argument cannot be controlled by an argument of the embedding clause. Examples (12) and (13) illustrate this contrast.

(12) a. * N'he obligat molts a quedar-se'n.
en.cl-*have*.1p.sg *obligate*.pp *many*.m.pl *to stay*.inf-*se*.cl-*en*.cl
'I have obligated many to stay.'

b. N'he obligat molts a quedar-se.
en.cl-*have*.1p.sg *obligate*.pp *many*.m.pl *to stay*.inf-*se*.cl
'I have obligated many to stay.'

(13) a. * N'he obligat molts a estudiar-ne.
en.cl-*have*.1p.sg *obligate*.pp *many*.m.pl *to study*.inf-*en*.cl
'I have obligated many to study'

b. N'he obligat molts a estudiar.
en.cl-*have*.1p.sg *obligate*.pp *many*.m.pl to *study*.inf
'I have obligated many to study.'

The fact that the control relation in (12b) and (13b) is grammatical indicates that the intransitive argument of the embedded clause is the subject. This is further confirmed by the ungrammaticality of (12a) and (13a), in which the clitic *en* appears in the embedded clause. If we assume that *en* cliticization is an object property, the ungrammaticality of (12a) and (13a) follows naturally: as an object, the argument of the embedded clause cannot be controlled.

2.3 Verbal agreement

It is commonly assumed that the agreement trigger of the verb is the subject (Chomsky 1981, 1995, among others). In a simple example with a transitive verb like (14), the auxiliary *haver* is in the third person plural form, agreeing with the subject *els estudiants* 'the students':

(14) Els estudiants han/*ha llegit aquest llibre.
the.pl *student*.pl *have*.3p.pl/*sg *read*.pp.m.sg *this*.m.sg *book*.m.sg
'The students have read this book.'

Intransitive verbs regularly agree with their single direct argument. But we would have a problem if we should assume that the agreement trigger is necessarily the subject: *molts* in (15) would have to be both a subject (as the agreement trigger) and an object (as it is expressed by means of the *en* clitic):

(15) Avui en surten/*surt molts.
today en.cl *leave*.pl/*sg *many*.pl
'Today many are leaving.'

The verbal agreement facts of languages like Icelandic or Hindi indicate that, in such languages, the verb can agree with a grammatical function other than the subject, provided that it is in *nominative* case:

(16) a. Henni líkuðu hestarnir.
she.dat.3p.sg *like*.past.3p.pl *horse*.nom.3p.pl
'She liked the horses' (Icelandic, Sigurðsson 2004:139)

b. Ravii-ne / niinaa-ne kelaa k^{h}aayaa
Ravi-erg.m.sg/ *Nina*-erg.f.sg *banana*-nom.m.sg *eat*.perf.m.sg
'Ravi/Nina ate a banana.' (Hindi, Mohanan 1994:104)

The same assumption will allow us to solve the paradox of (15): the verb agrees with a nominative argument, whether it is a subject or an object, and in (15) the verb in fact agrees with the object, which is nominative.

Independent evidence for the claim that the argument with which the verb agrees is nominative comes from the contrast between nominative and accusative with respect to the use of the preposition *a* 'to'. An indefinite pronoun allows *a*-marking optionally only if it is animate and accusative:

(17) a. (*A) molts llegeixen el llibre.
to *many*.m.pl *read*.3p.pl *the*.m.sg *book*.m.sg
'Many read the book.'

b. En veiem (a) molts.
en.cl *see*.1p.pl *to many*.m.pl
'We see many.'

Nominatives never allow *a*-marking, whether SUBJ (as in (17a)) or OBJ (as in (18)):

(18) En surten (*a) molts.
en.cl *leave*.pl *to many*.m.pl
'Many are leaving.'

From the facts listed above, we conclude that the intransitive argument in Catalan alternates between subject and object, and is always nominative. The intransitive verb agrees with this argument, regardless of the function it takes.[7]

3 Argument realization

In this section we propose the theory of argument realization needed to account for the facts reported in the previous section concerning the expression of the intransitive argument in Catalan. In 3.1, we briefly point out the difficulties that existing theories of argument realization within LFG would face in

[7] The properties discussed in subsection 2.2 cannot be attributed to nominative case, rather than to subjecthood. For example, the controlee has to be the subject and not just a nominative argument, as shown in (12) and (13).

accounting for these facts. In 3.2, an alternative argument realization theory is proposed, in which case assignment is a central element. In 3.3, we show how some of the main facts are derived from this theory, and, in 3.4, we show some constraints on the subject-object alternation.

3.1 Current LFG mapping theories

Current LFG theories of argument realization face two problems with respect to the facts considered in this paper: the treatment of multiple objects and the difficulty in accounting for the subject-object alternation of external arguments, which we will address in turn.

Since its inception, LFG has assumed as a general property of all languages that clauses have at most one unrestricted object and possibly one or more restricted objects. These two kinds of GFs have been designated by different names, including OBJ and OBJθ, to refer to unrestricted and restricted object, respectively, which we shall use for brevity. While the distinction between these two types of object finds strong motivation in asymmetrical languages such as Chicheŵa (see Alsina and Mchombo 1990, 1993, and Bresnan and Moshi 1990, among others), it is unmotivated in many other languages, particularly in languages that make use of grammatical case such as Catalan and the other Romance languages.[8] Therefore, assuming the OBJ/OBJθ distinction for all languages constitutes an unnecessary complication of the analysis of multiple objects in the latter type of language.

As noted already in Alsina (1996), the relevant distinction among objects in Catalan (as well as other Romance languages) is in terms of grammatical case: dative vs. non-dative objects. Stipulating that one of the two objects is an OBJ and the other one an OBJθ plays no role in accounting for the facts in this language and does not allow us to maintain that this distinction has a cross-linguistically valid empirical reflex. The behavior of objects in Catalan is entirely predictable from the presence or absence of dative case. Stipulating that the dative object is the OBJθ is redundant, as it would be to stipulate that the dative object is the OBJ and the non-dative object is the OBJθ. Both dative and non-dative objects can be expressed by means of pronominal clitics (and in some cases dative objects are preferentially expressed in this way), which can be taken to be the equivalent of object marking in the Bantu languages, a property not available to OBJθ. Both dative and non-dative objects can be reflexivized (and reciprocalized), which is the equivalent of reciprocalization in Bantu, another property in which OBJθ does not take part. The failure of dative objects (in contrast with non-dative objects) to alternate with the SUBJ function (i.e., to passivize) is best analyzed by means of a language-particular constraint disallowing dative subjects (see Nominative Subject Constraint (23) below). As is well known, other case-marking languages lack this constraint

[8] Bresnan and Moshi (1990:167) already note that "many languages (including Romance) lack restricted objects altogether". We thank an anonymous reviewer for bringing up this point.

and allow dative subjects, or other subjects with other marked cases (e.g., Icelandic, Hindi-Urdu, etc.).

In addition, importing the OBJ/OBJθ distinction into Catalan would render this distinction devoid of any cross-linguistically valid empirical effect. On the basis of asymmetrical languages such as Chicheŵa, in which the OBJ/OBJθ distinction does play an important role, we can observe that certain properties are only available to OBJ, such as expression by means of an object marker, possibility of passivization, or accessibility to reciprocalization. In Catalan, the two types of objects are available for expression by means of a verbal clitic and for reflexivization. If dative objects were assumed to be OBJθ and non-dative objects were assumed to be OBJ, it would no longer be possible to maintain that certain properties (such as expression by means of object markers or clitics and accessibility to reflexization or reciprocalization) are cross-linguistically properties of OBJ (that is, unavailable to OBJθ).

Therefore, we do not assume that objects in Catalan are represented as either OBJ or OBJθ. Instead, we assume that, cross-linguistically, there can be multiple instances of the GF OBJ and that, in some languages, objects are distinguished by means of grammatical case. Catalan is one of these languages, in which objects can be either dative or non-dative. In languages such as Chicheŵa, where there are no grammatical case distinctions, objects are distinguished between restricted and unrestricted at the level of argument structure. As proposed in Alsina (2001), internal arguments may be marked as R at the level of argument structure, so that there may be at most one internal argument not marked with this feature. This feature makes the argument so marked unavailable to the morphosyntactic properties noted above (object marking, reciprocalization, possibility of passivization).

The proposal that objects are not distinguished in terms of grammatical function, since they all bear the GF OBJ, but may be distinguished either in terms of grammatical case (as in Catalan) or in terms of the presence or absence of the feature R at the level of argument structure (as in Chicheŵa) entails rejecting the four-way classification of grammatical functions found in current versions of the Lexical Mapping Theory (LMT), as in, for example, Levin (1986), Bresnan and Kanerva (1989), Bresnan and Moshi (1990), Kibort (2001, 2009, among others), and Findlay (2016). These versions of LMT assume that there are four basic GFs: SUBJ, OBJ, OBJθ, and OBLθ. These theories also assume a decomposition of these GFs by means of the features [±r] and [±o] and that arguments are classified by means of these features. Since these features combine to yield the four GFs just mentioned, they also need to be discarded in the theory to be advanced in subsection 3.2.

The second problem with current LFG mapping theories can be seen as a consequence of the featural decomposition of GFs just discussed. The classification of an argument by means of one of these features implies the possibility of an alternation between two GFs. If an argument is classified at a-structure as [–r], as is assumed for internal arguments, it can map onto either

SUBJ or OBJ; if it is classified as [+o], it can map onto either OBJ or OBJθ, and so on. This restricts the possible GF alternations. External arguments, such as agents, are assumed to have the [–o] classification, which limits the possible realizations to SUBJ and OBLθ. What is not assumed in current versions of LMT is for external arguments to show a SUBJ/OBJ alternation, but what we find in Catalan is that the intransitive argument, whether internal or external, shows the SUBJ/OBJ alternation.[9] In contrast, the external argument of transitive verbs is constrained to map onto the SUBJ function. This shows that a [–o] argument classification is inadequate for external arguments and that the mapping of external arguments depends in part on the other arguments in the argument structure.

3.2 Argument-to-function mapping theory

The present mapping theory assumes a level of argument structure, or a-structure, and three sets of principles of argument realization, which relate a-structure to f-structure: case assignment principles, argument-to-GF linking rules, and constraints on case features.

A-structure

A-structure consists of the list of arguments of a predicate, without any thematic information, ordered according to the thematic hierarchy, such as the commonly assumed hierarchy based on Givón (1984), Kiparsky (1987), and Bresnan and Kanerva (1989), among others:

(19) Thematic Hierarchy:
ag > ben > recip/exp > inst > th/pt > loc

Arguments are classified into *core* arguments (C) and *non-core* arguments (NC). As we shall see, core arguments are the ones that map onto direct grammatical functions (i.e., SUBJ and OBJ). Core arguments are further divided into external argument (E) and internal argument (I) and represented as such in the a-structure. The external argument E, if there is one, is the most prominent argument in the argument structure. Non-core arguments are those that map onto the indirect function OBL.

Case assignment principles

In this theory, case assignment is crucial for argument realization. For Catalan, we assume that there are three case values—dative, accusative, and nominative—for the core arguments, and that all core arguments must be assigned a case value, according to the following case assignment principles, ordered by priority:

[9] The subject-object alternation of the intransitive argument (both internal and external argument) is also found in Norwegian, Swedish, and Danish, according to Lødrup (1999).

(20) Case Assignment Principles:

i. Assign *dative* case to the more prominent of two internal arguments, or to a goal;[10]

ii. Assign *accusative* case to the less prominent of two *core* arguments that lack case;

iii. Elsewhere, assign *nominative* case to a *core* argument.

Argument-to-GF linking rules

We propose two rules to license the correspondence between arguments and GFs—the Core Argument Rule and the Elsewhere Mapping Rule—and Passivization, as an instance of a morphosyntactic operation that affects the argument-to-GF linking.

The Core Argument Rule requires a core argument (C) to map onto a direct grammatical function (DGF), the class of GFs that consists of SUBJ and OBJ:

(21) Core Argument Rule: C
|
DGF

This rule allows the external argument, as well as an internal argument, to be either SUBJ or OBJ, which is not possible in previous mapping theories like Bresnan and Kanerva (1989), Kibort (2001), or Findlay (2016), for, as noted earlier, the proposal that the external argument is associated with [–o] prevents linking this argument to an OBJ.

The operation of passivization blocks the linkage of the highest argument to a DGF:[11]

(22) Passivization: $\hat{\theta}$
|
DGF

Finally, the Elsewhere Mapping Rule optionally links an argument to OBL:

(23) Elsewhere Mapping Rule: A
|
(OBL)

This rule is ordered after the other linking rules and therefore it applies to arguments to which the Core Argument Rule (21) cannot apply: non-core arguments as well as arguments that have their linkage to DGF cut off by morphosyntactic operations like passive or antipassive. The optionality of this rule captures the idea that in general OBLs are not obligatory. Moreover, this optionality may be overridden by having a lexical entry specifying that an

[10] See Alsina (1996:175) for a detailed discussion.

[11] This allows for cross-linguistic variation. In Spanish and Catalan, passivization prevents the linkage of the highest argument to a direct grammatical function, thus accounting for the *se* passivization/impersonalization with both unergative and unaccusative verbs, but in languages like German or Dutch, we need to rewrite the operation of passivization as 'blocking the linkage of the external argument to a direct grammatical function', since there is no passivization/impersonalization with unaccusative verbs in these languages.

argument is obligatorily mapped onto an oblique.

Constraints on case features

There are some constraints on the association of particular case features with particular GFs. Catalan, along with other Romance languages, but unlike languages such as Icelandic and Hindi-Urdu, requires subjects to be in the nominative case (or, conversely, rules out subjects in a case other than nominative). For example, in Catalan there are no dative subjects (see Alsina 1996) or accusative subjects. To account for this fact, we posit the Nominative Subject Constraint:

(24) Nominative Subject Constraint (specific to Catalan):
*SUBJ [CASE ¬NOM]

The effect of this constraint is to rule out structures with a non-nominative subject. Notice that the implication is unidirectional: subjects must be nominative, but it is not required for a nominative expression to be a subject.

A second case constraint that we need to consider is what we may call the 1 Non-Dative Object Constraint (or 1NDO): a structure allows at most one object that is not dative:

(25) 1 Non-Dative Object Constraint (1NDO):
*[CASE ¬DAT] [CASE ¬DAT]
| |
OBJ OBJ

This constraint rules out a structure with two accusative objects, or with two nominative objects, or with a nominative object and an accusative object. Together with constraint (24), it has the effect of requiring a nominative argument to be the subject if it co-occurs with an accusative object. Notice that the principles and constraints stated so far do not require the presence of a subject in the clause and so it is the 1NDO constraint that forces a nominative to be the subject if there is an accusative in the structure.

3.3 Illustration of the theory

We now provide some examples of how the proposed argument realization theory works in Catalan.

A ditransitive verb like *donar* 'give' is lexically specified with one external and two internal arguments, as represented in (26). The goal argument is the more prominent internal argument, thus, by case assignment principle (20i), it will get dative case. The theme argument, as the less prominent of the two arguments—agent and theme—lacking case, is assigned accusative case, according to principle (20ii). Finally, the external argument receives nominative case by principle (20iii). As for the argument-to-GF mappings, the three arguments, being core arguments, are required to map onto a direct GF by the Core Argument Rule (21). However, the goal and theme arguments can only be realized as OBJ according to the Nominative Subject Constraint (24) and the nominative agent argument must be realized as SUBJ in order to avoid

violating the 1NDO constraint (25). The representation in (26) and subsequent ones show the thematic roles of the arguments involved merely for convenience, as they are not part of the a-structure or of the f-structure; the a-structure is shown in angled brackets; the case features assigned to each argument are shown on the line below it and, on the next line, are the corresponding GFs; the relevant principles are given in parentheses.

(26) *Donar* 'give'

(ag)	(go)	(th)	
< E	I	I	>
nom	dat	acc	(Case assignment principles (20))
SUBJ	OBJ	OBJ	(Rules (21), (24), (25))

'Like' type verbs in Catalan (as well as other Romance languages like Spanish or Italian) have two internal arguments and no external argument. The experiencer argument gets dative case by principle (20i) and maps onto OBJ because of the Nominative Subject Constraint (24). The theme argument is assigned nominative case by principle (20iii), thus being compatible with both SUBJ and OBJ:[12]

(27) *Agradar* 'like'

(exp)	(th)	
< I	I	>
dat	nom	(Case assignment principles (20i, iii))
OBJ	SUBJ/OBJ	(Rules (21), (24))

Intransitive verbs, whether unergative or unaccusative, only have one *core* argument (an external and an internal argument, respectively), as exemplified in (28) for the unergative *treballar* 'work'. Case assignment principle (20iii) applies assigning nominative case. This core argument, as we have seen in section 2, alternates between SUBJ and OBJ.

(28) *Treballar* 'work'

(ag)		
< E	>	
nom		(Case assignment principle (20iii))
SUBJ/OBJ		(Rule (21))

Transitive verbs like *llegir* 'read' have an external and an internal argument. Since there is only one internal argument, dative case is not assigned;

[12] Although space does not permit us to give detailed examples of this class of verbs, only the theme argument can be expressed by the clitic *en*; the experiencer cannot, due to the case restriction in constraint (32).

accusative case is assigned to the less prominent argument (i.e., the internal argument); by principle (20iii), nominative case is assigned to the external argument. In accordance to the Nominative Subject Constraint and the 1NDO constraint, the external argument maps to the SUBJ and the internal argument to the OBJ:

(29) *Llegir* 'read'

(ag)	(th)	
< E	I >	
nom	acc	(Case assignment principles (20ii, iii))
SUBJ	OBJ	(Rule (21), (24), (25))

When the transitive verb is passivized, the linkage of the external argument to a direct grammatical function is blocked. Since there is only one internal argument, case assignment principles (20i, ii) will not be used. Then, by principle (20iii), the internal argument gets nominative case. This internal argument can map onto either SUBJ or OBJ:[13]

(30) *Llegit* 'read-PASS'

(ag)	(th)	
< E	I >	
	nom	(Rule (22), case assignment principle (20iii))
(OBL) DGF	SUBJ/OBJ	(Rules (21), (23))

From the representations in (26)-(30), which illustrate different patterns of argument realization, we can see that a clause in Catalan: **i)** *may* contain at most one SUBJ; **ii)** *need not* contain a SUBJ, and **iii)** may contain more than one OBJ. The uniqueness of the subject and the multiplicity of objects can be handled in a variety of ways (see e.g. Alsina 1996 and Patejuk and Przepiórkowski 2016). This proposal can be implemented within the standard LFG formalism by assuming that the SUBJ is single-valued and OBJ is set-valued. But we will not go into further details of this topic in this paper.

3.4 Constraints on the subject-object alternation

In section 2 we saw that the intransitive argument can alternate between SUBJ and OBJ. However, if this SUBJ/OBJ alternation were completely free, nothing would require the presence of the clitic *en* in (31), as shown by the contrast between the grammatical (15), repeated as (31a), with the clitic, and the absence of the clitic in the ungrammatical (31b):

[13] The nominative argument in a passivized clause can be either SUBJ or OBJ, showing the same behavior as the direct argument of intransitive verbs, described in section 2. We will discuss the conditions under which the subject-object alternation happens in subsection 3.4. We leave detailed issues about passivization and impersonalization for further study.

(31) a. Avui en surten molts.
today en.cl *leave*.pl *many*.pl
'Today many are leaving.'

b. * Avui surten molts.
today leave.pl *many*.pl
'Today many are leaving.'

Consider the information the *en* clitic provides: the *en* clitic corresponds to an OBJ that is pronominal and indefinite, which can either be nominative or accusative, but not dative, as illustrated in the f-structure in (32):

(32) *En*: $\begin{bmatrix} \text{OBJ} \begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{DEF} & - \\ \text{CASE} & \neg\text{DAT} \end{bmatrix} \end{bmatrix}$

The presence of this clitic indicates that it corresponds to an object, which may be expressed by an NP lacking a head N, as is the case of *molts* 'many' in (31a). However, if the core argument of a verb like *sortir* 'leave' were free to also be expressed as a subject, we would expect (31b), without the clitic *en*, to be grammatical, as this clitic cannot correspond to a subject. In order to explain the ungrammaticality of (31b), we assume that the subject-object alternation of the intransitive argument is constrained by definiteness and posit a constraint that penalizes an indefinite subject:[14]

(33) Indefinite Subject Ban:
*SUBJ [DEF -]

For an intransitive verb whose single direct argument is indefinite, constraint (33) penalizes the subject realization and favors the object realization. This explains the obligatoriness of *en* in (31). But notice that this constraint has no effect on transitive verbs, within an Optimality Theory (OT) conception (see Kuhn 2003), provided 1NDO (25) ranks higher than (33): the subject realization of the external argument of a transitive verb is the optimal candidate, even if it is indefinite and violates (33).

By contrast, when the sole argument of the intransitive verb is definite, it is the subject of the clause, like the NP *els estudiants* in (34):

(34) Avui surten els estudiants tard.
today leave.pl *the*.m.pl *student*.m.pl *late*
'Today the students are leaving late.'

The reasoning is that we also assume the Subject Condition (SC) (see Bresnan and Moshi 1990, among others), which requires every clause to have a subject, as an OT constraint: SC is a low-ranking constraint and, in particular, lower than the Indefinite Subject Ban (33) in Catalan.[15] When the intransitive

[14] According to Bartra (2009:3), Spanish and Catalan allow plural indefinites as objects of the verb but not as external subjects. The claim refers to bare NPs, a subset of indefinites, and it is also made by Espinal (2010) and Espinal and McNally (2010).

[15] Notice that, because of this ranking of constraints and because, in languages like Catalan, SC ranks below the faithfulness constraint requiring every GF to correspond to an argument, there

argument is definite and is not constrained by (33), the SC will penalize the candidate that lacks a subject and select the one in which the argument maps onto the subject.[16]

An additional fact that needs to be considered is that the *en* clitic cannot be licensed by a preverbal NP, even if this NP is indefinite:

(35) a. Ja n'han sortit quatre de l'ou.
already en.cl-*have*.pl *leave*.pp *four from the-egg*

b. Quatre ja (*n') han sortit de l'ou.
four already en.cl *have*.pl *leave*.pp *of the-egg*
'Four of them have already come out of the egg.'

(based on GLC 2016:699)

We adopt the assumption in Vallduví (2002) that preverbal NPs in Catalan (such as *quatre* in (35b)) are topics (not subjects) anaphorically related to an in-clause GF. Since the topic is the antecedent of an anaphoric pronoun (possibly null, as with null subjects) and anaphoric pronouns must be definite, it follows that topics cannot be related to the clitic *en*, because the lexical information of the *en* clitic specifies that it corresponds to an indefinite object. This makes it incompatible with its being an anaphoric pronoun dependent on the preverbal topic, thus explaining the ungrammaticality of the *en* clitic in (35b).

At this point, one may ask if it is possible to use a definite object clitic in place of the indefinite *en*, as it would qualify as a topic-anaphoric pronoun; the fact is that the definite object clitics *el/la/els/les* are incompatible with intransitive verbs:

(36) * Avui els surt/surten tard.
today them.obj.m.pl *leave*.sg/*leave*.pl *late*
'Today they are leaving late.'

Whichever agreement form of the verb is chosen, the core argument of the intransitive verb in (36) cannot be expressed by means of *els*. According to our analysis of (34), a definite argument of an intransitive verb is the subject. Since clitics like *el, la, els,* and *les* are (non-dative) *object* pronouns, they cannot be used as subjects, which explains the ungrammaticality of (36).[17]

---

are no expletive subjects in Catalan and there is no subject in a sentence like (31a). Languages with expletive subjects, such as French and English, have the opposite ranking of SC and this faithfulness constraint.

[16] Although we do not have space to give a detailed OT analysis of the phenomena considered here, the following ranking of constraints is assumed for Catalan: (25) » (33) » SC.

[17] The argument realization theory presented in this section has been developed in order to account for the facts of Catalan. It is beyond the scope of this paper to include this theory in a general theory of argument realization, defining the parameters of variation needed to account for cross-linguistic variation in this area. The approach may have points in common with Kiparsky's (1987) linking theory, but space limitations prevent us from making a detailed comparison.

4 Verbal agreement

In order to account for the idea that a verb can agree with either a subject or an object, if nominative, we follow Haug and Nikitina (2012, 2016), and Alsina and Vigo (2014, 2017), among others, in assuming that verbal agreement is mediated by the feature bundle AGR, which contains the agreement features encoded by the verb. Two general constraints, adopted from Alsina and Vigo (2014, 2017), are relevant to account for the agreement of the verb with one of its dependent GFs: the requirement that the clausal AGR feature be shared with that of a dependent GF (AGRSHARE (37a)), and the requirement that the agreeing GF be nominative (*AGRCASE (37b)):

(37) a. AGRSHARE: $\left[\begin{array}{l}\text{AGR } \boxed{1} \\ \text{DGF}\left[\text{AGR } \boxed{1}\right]\end{array}\right]_f$

For f-structure f that maps to a constituent of category V

b. * AGRCASE: $*\left[\begin{array}{ll}\text{AGR} & \boxed{1} \\ \text{GF} & \left[\begin{array}{ll}\text{AGR} & \boxed{1} \\ \text{CASE} & \neg\text{NOM}\end{array}\right]\end{array}\right]_f$

For f-structure f that maps to a constituent of category V

Thus, verbal agreement with a subject and with an object is represented as in (38a) and (38b), respectively:

(38) a. Avui surten els estudiants tard.
today leave.pl *the*.pl *student*.pl *late*
'Today the students are leaving late.'

$$\left[\begin{array}{ll}\text{PRED} & \text{`}\textit{leave}\text{ <Arg}_1\text{>'} \\ \text{AGR } \boxed{1} & \left[\begin{array}{ll}\text{PERS} & 3 \\ \text{NUM} & \text{PL}\end{array}\right] \\ \text{SUBJ} & \left[\begin{array}{ll}\text{PRED} & \text{`student'} \\ \text{DEF} & + \\ \text{AGR} & \boxed{1} \\ \text{CASE} & \text{NOM}\end{array}\right]_1\end{array}\right]$$

b. Avui en surten molts.
today en.cl *leave*.pl *many*.pl
'Today many are leaving.'

$$\left[\begin{array}{ll}\text{PRED} & \text{`}\textit{leave}\text{ <Arg}_1\text{>'} \\ \text{AGR } \boxed{1} & \left[\begin{array}{ll}\text{PERS} & 3 \\ \text{NUM} & \text{PL}\end{array}\right] \\ \text{OBJ} & \left[\begin{array}{ll}\text{PRED} & \text{`PRO'} \\ \text{DEF} & - \\ \text{AGR} & \boxed{1} \\ \text{QUANT} & \text{`many'} \\ \text{CASE} & \text{NOM}\end{array}\right]_1\end{array}\right]$$

In Catalan, a raising verb like *semblar* 'seem' can agree with the nominative object of the embedded clause:

(39) Semblen [arribar-ne molts.]
seem.3p.pl *arrive*.inf-*en*.cl *many*.pl
'Many seem to arrive.'

This is an instance of (apparent) long-distance agreement, as the inflected verb form *semblen* 'seem' in (39) doesn't seem to agree with any of its dependent GFs, but with the object *molts* 'many' in the infinitival complement clause. The only GF in the f-structure of *semblen* 'seem' that this verb could agree with is its complement clause, but, if the verb were to agree with it, it would

have to be in the third person singular form on the assumption that clauses agree in the third person singular. To solve this problem, we assume that long-distance agreement like the one in (39) is a combination of two local agreement relations, as in Alsina and Vigo (2017): **i)** the sharing of the AGR of the raising clause with the AGR of its infinitival complement, and **ii)** the sharing of this AGR with that of the object of the infinitive.

But not all verbs allow AGR sharing with the AGR of their embedded clause: only raising verbs do. To be formal, we assume a constraint, i.e., Clausal Opacity, which blocks the sharing of either AGR or GF in a given clause with either the AGR or a GF of its embedded clause. Raising verbs include a lexical specification overriding Clausal Opacity.

(40) Clausal Opacity:

$$*\begin{bmatrix} G & \boxed{1} \\ GF & [F\ \boxed{1}]_f \end{bmatrix}_g$$

For f-structures *f, g* that map to constituents of category V, and F, G = {DGF, AGR}

The cross-clausal agreement in (39) is possible because *semblar* 'seem' is a raising verb; thus, Clausal Opacity does not apply to f-structures whose PRED belongs to this verb, allowing both the structure-sharing of its subject with the subject of its infinitival complement (raising, as standardly understood) and the structure-sharing of its AGR with that of its infinitival complement ("raising" of the agreement features). So, the f-structure of (39) can be represented as:

(41)
$$\begin{bmatrix} \text{PRED} & \text{'}\textit{seem}\text{ <Arg1>'} \\ \text{AGR} & \boxed{1} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'}\textit{arrive}\text{ <Arg2>'} \\ \text{AGR} & \boxed{1}\begin{bmatrix} \text{PERS} & 3 \\ \text{NUM} & \text{PL} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{DEF} & - \\ \text{AGR} & \boxed{1} \\ \text{QUANT} & \text{'many'} \\ \text{CASE} & \text{NOM} \end{bmatrix}_2 \end{bmatrix}_1 \end{bmatrix}$$

Just like the raising of a subject is unbounded and can cross as many clauses as contain a raising verb, the raising of the agreement features is likewise potentially unbounded. All that is required is for there to be a chain of raising verbs overriding Clausal Opacity, as can be seen in the following example, where both *semblen* 'seem' and *tendir* 'tend' are raising verbs:

(42) Semblen [tendir [a arribar-ne molts]].
seem.3p.pl *tend*.inf *to arrive*.inf-*en*.cl *many*
'Many seem to tend to arrive.'

Once we have assumed that an intransitive argument can be a nominative object, we can explain the agreement facts, namely, the observation that the

verb agrees with its object and can be involved in long-distance agreement, adopting the agreement theory of Alsina and Vigo (2014, 2017) without additional assumptions.

5 Conclusion

This paper has argued for the claim that the single direct argument of an intransitive verb in Catalan can be a nominative object. This argument shows a subject-object alternation, but is invariably in the *nominative* case. The alternation is constrained by definiteness, so that the argument is a subject if it is definite and is an object if it is indefinite. As a subject, it displays the expected properties of a subject, including the possibility of pro-drop; as an object, it displays the expected properties of an object, including expression by means of the object clitic *en*. The claim that it is a nominative expression explains the observation that it agrees with the verb, even when it is an object, applying a theory of verbal agreement proposed independently of the facts of Catalan.

The theory of argument realization proposed in this paper is a simple one, as it assumes only two argument-to-GF mapping rules, three case assignment principles, and a small set of constraints restricting the GF assignment on the basis of the case features and definiteness of the arguments, in addition to morphosyntactic operations such as passivization.

References

Alsina, Alex. 1986. Assaig de definició de les funcions del pronom "en". In *Estudis de Llengua i Literatura Catalanes XII, Miscel·lània Antoni M. Badia i Margarit*, vol. 4, pp. 95-121. Barcelona: Publicacions de l'Abadia de Montserrat.

Alsina, Alex. 1995. The fall of function-argument biuniqueness. In Glyn Morrill and Richard Oehrle (Eds.) *Formal Grammar: Proceedings of the Conference of the European Summer School in Logic, Language and Information*, pp. 1-16. Barcelona: Universitat Politècnica de Catalunya.

Alsina, Alex. 1996. *The Role of Argument Structure in Grammar: Evidence from Romance*. Stanford: CSLI Publications.

Alsina, Alex. 2001. On the nonsemantic nature of argument structure. *Language Sciences*, 23, pp. 355-389.

Alsina, Alex and Sam Mchombo. 1990. The syntax of applicative in Chichewa: problems for a theta theoretic asymmetry. *Natural Language and Linguistic Theory*, 8(4), pp. 493-506.

Alsina, Alex and Sam Mchombo. 1993. Object asymmetries and the Chichewa applicative construction. In Sam Mchombo (Ed.), *Theoretical Aspects of*

Bantu Grammar, pp. 17-45. Stanford: CSLI publications.

Alsina, Alex and Eugenio M. Vigo. 2014. Copular inversion and non-subject agreement. In Miriam Butt and Tracy Holloway King (Eds.), *Proceedings of the LFG14 Conference*, pp. 5-25. Stanford: CSLI Publications.

Alsina, Alex and Eugenio M. Vigo. 2017. Agreement: interactions with case and raising. In Miriam Butt and Tracy Holloway King (Eds.), *Proceedings of the LFG17 Conference*, pp. 3-23. Stanford: CSLI Publications.

Bartra, Anna. 2009. Some remarks about the grammaticalization process of the DP functional domain in Old Romance. In Teresa Espinal, Manuel Leonetti, and Louise McNally (Eds.), *Proceedings of the IV Nereus International Workshop: Definiteness and DP Structure in Romance Languages*, pp. 1-26. Universität Konstanz.

Bresnan, Joan and Jonni Kanerva. 1989. Locative inversion in Chichewa: a case study of factorization in grammar. *Linguistic Inquiry* 20 (1), pp. 1-50.

Bresnan, Joan and Lioba Moshi. 1990. Object asymmetries in comparative Bantu syntax. *Linguistic Inquiry* 21 (2), pp. 147-186.

Burzio, Luigi. 1986. *Italian Syntax: a Government-Binding Approach*. Dordrecht: Reidel.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge MA: The MIT Press.

Cortés, Corinne. 1995. Subject-object asymmetries and verb classes. *Linguistica Atlantica* 17, pp. 63-78.

Cortés, Corinne and Anna Gavarró. 1997. Subject-object asymmetries and the clitic *en*. In James. R. Black and Virginia Motapanyane (Eds.), *Clitics, Pronouns and Movement*, pp. 39-62. Amsterdam: John Benjamins.

Espinal, Teresa. 2010. Bare nominals in Catalan and Spanish. Their structure and meaning. *Lingua* 120(4), pp. 984-1009.

Espinal, Teresa and Louise McNally. 2010. Bare nominals and incorporating verbs in Spanish and Catalan. *Journal of Linguistics* 46, pp. 1-42.

Fabra, Pompeu. 1912. *Gramática de la Lengua Catalana*. Barcelona: L'Avenç.

Findlay, Jamie. 2016. Mapping theory without argument structure. *Journal of Language Modelling*, 4(2), pp. 293–338.

Givón, Talmy. 1984. *Syntax: a Functional-Typological Introduction*. Amsterdam: John Benjamins.

Gràcia, Lluïsa. 1989. *Els Verbs Ergatius en Català*. Maó, Menorca: Institut Menorquí d'Estudis.

Haug, Dag and Tanya Nikitina. 2012. The many cases of non-finite subjects: the challenge of "dominant" participles. In Miriam Butt and Tracy Holloway King (Eds.), *Proceedings of the LFG12 Conference*, pp. 292-331. Stanford:

CSLI Publications.

Haug, Dag and Tanya Nikitina. 2016. Feature sharing in agreement. *Natural Language and Linguistic Theory*, 34(3), pp. 865-910.

Institut d'Estudis Catalans. 2016. *Gramàtica de la Llengua Catalana* (GLC), pp. 697-699. Barcelona: Institut d'Estudis Catalans.

Kibort, Anna. 2001. The Polish passive and impersonal in Lexical Mapping Theory. In Miriam Butt and Tracy Holloway King (Eds.), *Proceedings of the LFG01 Conference*, pp. 163-183. Stanford: CSLI Publications.

Kibort, Anna. 2009. Intermediary agents and unexpressed pronouns. In Miriam Butt and Tracy Holloway King (Eds.), *Proceedings of the LFG09 Conference*, pp. 378-398. Stanford: CSLI Publications.

Kiparsky, Paul. 1987. *Morphology and Grammatical Relations*, ms. Stanford CA: Stanford University.

Kuhn, Jonas. 2003. *Optimality-Theoretic Syntax: a Declarative Approach*. Stanford: CSLI Publications.

Levin, Lorraine. 1986. *Operations on Lexical Forms: Unaccusative Rules in Germanic Languages*. PhD dissertation. MIT.

Lødrup, Helge. 1999. Linking and optimality in the Norwegian presentational focus construction. *Nordic Journal of Linguistics*, 22, pp. 205-230.

Mohanan, Tara. 1994. *Argument Structure in Hindi*. Stanford: CSLI Publications.

Patejuk, Agnieszka and Adam Przepiórkowski. 2016. Reducing grammatical functions in LFG. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King and Stefan Müller (Eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar*, pp. 541-559. Stanford: CSLI Publications.

Saccon, Graziella. 1995. *Ne*-cliticization does not support the unaccusative/intransitive split. In Glyn Morrill and Richard Oehrle (Eds.) *Formal Grammar: Proceedings of the Conference of the European Summer School in Logic, Language, and Information*, pp. 227-238. Barcelona: Universitat Politècnica de Catalunya.

Sigurðsson, Halldór Á. 2004. Icelandic non-nominative subjects: facts and implications. In Peri Bhaskararao and Karumuri Venkata Subbarao (Eds.), *Non-nominative Subjects*, vol. 2, pp. 137-159. Amsterdam: John Benjamins.

Vallduví, Enric. 2002. L'oració com a unitat informativa. In Joan Solà, Maria Rosa Lloret, Joan Mascaró, and Manuel Pérez Saldanya (Eds.), *Gramàtica del Català Contemporani*, vol. 2, pp. 1221-1279. Barcelona: Editorial Empúries.

Romanian Object Clitics: Grammaticalization, agreement and lexical splits

Roxana-Maria Barbu
Carleton University

Ida Toivonen
Carleton University

Proceedings of the LFG'18 Conference

University of Vienna

Miriam Butt, Tracy Holloway King (Editors)

2018

CSLI Publications

pages 67–87

http://csli-publications.stanford.edu/LFG/2018

Keywords: Romanian, clitics, agreement marking, pro-drop

Barbu, Roxana-Maria, & Toivonen, Ida. 2018. Romanian Object Clitics: Grammaticalization, agreement and lexical splits. In Butt, Miriam, & King, Tracy Holloway (Eds.), *Proceedings of the LFG'18 Conference, University of Vienna*, 67–87. Stanford, CA: CSLI Publications.

Abstract

Direct object clitics in Modern Standard Romanian display different properties depending on whether or not they double an object. We propose a dual analysis for the clitics: they function as agreement makers when they double an object and as pronouns when they do not. Furthermore, the lexical entries differ beyond the presence or absence of pronominal referential features, and this accounts for the split behavior. The analysis is placed in its historical context and extended to other varieties of Romanian. Finally, we argue that the Romanian lexical split is not an isolated phenomenon: multiple similar splits can be found in the typology of agreement marking.

1 Introduction

Romanian object clitics can occur with or without an object double.[1] This is illustrated with the third person masculine clitic *l-* in (1–3):[2]

(1) L-am văzut.
3SG.M.ACC-have.1SG seen
'I saw him/it.'

(2) L-am văzut pe el.
3SG.M.ACC-have.1SG seen ACC him
'I saw him.'

(3) L-am văzut pe băiat.
3SG.M.ACC-have.1SG seen ACC boy
'I saw the boy.'

Example (1) marks the object with a clitic only; there is no independent object nominal. In (2), the clitic doubles a pronominal object, and in (3), it doubles a non-pronominal noun.

This phenomenon is often referred to as "*pro*-drop" in the literature: the overt pronoun is "dropped" or phonologically unrealized. We will make use of the traditional term *pro*-drop, but we use it in a theory-neutral way to refer to the basic data pattern, and not as a term that implies that something has actually been dropped.

The analysis of Romanian object clitics presented in this paper builds on the standard LFG analysis of *pro*-drop, as spelled out in Fassi Fehri (1984); Bresnan & Mchombo (1987); Bresnan et al. (2016, Chapter 8), and elsewhere. In line with previous LFG analyses, we propose that the Romanian clitics have a dual nature: they are ambiguous between agreement markers and referential pronouns.

[1]We want to thank the audience at LFG18 for insightful questions and comments. We would also like to thank Raj Singh for his feedback. We are very grateful to Aurelia Barbu, Octavian Barbu, and Lenuţa Focşa for discussing the Romanian data with us and providing acceptability judgements. This paper has been greatly improved thanks to comments from the editors and an anonymous reviewer.

[2]The following abbreviations are used in this paper: SG = singular, PL = plural, M = masculine, F = feminine, ACC = accusative, PTCPL = participle, OBJ = object, NUM = number, GEND = gender, PERS = person, DEF = definite, PST = past, and IMP = imperative.

In examples such as (1), where the clitic is not accompanied by an independent object, the clitic is a pronoun. In (2) and (3), the clitic is an agreement marker.

A uniform analysis where the clitic is consistently either an agreement marker or a pronoun may seem like a better analysis *a priori*, but there are empirical arguments against this position. The argumentation builds on the fact that the agreement-marking clitic is more restricted in its distribution than the pronominal clitic. The clitics thus do not only differ in pronominal status. In LFG terms, they differ beyond the presence or absence of the [PRED 'pro'] feature.

The proposed analysis of Romanian clitics will be placed in a historical context. The "lexical split" in Romanian clitics might seem like an unusual quirk, but it in fact follows naturally from commonly assumed grammaticalization processes that such a split would occur. We will also consider dialectal variation within Romanian. The paper finally mentions a number of examples from a variety of languages that illustrate that the Romanian clitic system is not so exotic after all.

2 Object clitics in Modern Standard Romanian

2.1 A brief introduction to the clitics

The object clitic forms in Modern Standard Romanian (MSR)[3] are given in (4):

(4) Direct object clitic pronouns

	SG	PL
1	*mă/m*	*ne*
2	*te*	*vă/v*
3M	*îl/l*	*îi/i*
3F	*o*	*le*

The morphophonological status of Romanian clitics is controversial (Dobrovie-Sorin, 1994; Monachesi, 1998; Popescu, 2000; Luís, 2004): Are they clitics or bound morphemes? This paper will follow the majority view and treat them as clitics (i.e., non-projecting, phonologically dependent words) and not bound morphemes. However, nothing in our analysis hinges on this decision. Since the LFG architecture allows for 'mismatches' between levels of grammatical structure, the morphophonological status of the clitics does not dictate whether they are agreement markers or pronouns. We return to this later; for now it suffices to note that our analysis can be translated into one that treats the clitics as bound morphemes.[4]

The phrase-structural realization of the clitics is not central to our analysis, but we outline our assumptions here for concreteness. The $\hat{\text{D}}$ is a non-projecting

[3]There is variation in Romanian regarding the distribution of object clitics. Some of that variation will be discussed in Section 5. The dialect we present in this section and the next is widespread and consistent with most grammar books. Still, MSR might not be the best label to use, and we do not mean for it to have any special significance.

[4]In fact, we expect that a careful analysis of different varieties of Romanian might reveal that the morphemes are phonologically bound in some dialects and true clitics in others.

D node (Toivonen, 2003). Its distribution is restricted by the Romanian phrase structure rules in (5) and (7) below:

(5) $\mathrm{I}^{\mathrm{o}} \longrightarrow \underset{(\uparrow \mathrm{OBJ})=\downarrow}{\hat{\mathrm{D}}} \quad \underset{\uparrow=\downarrow}{\mathrm{I}^{\mathrm{o}}}$

The c-structure for (1–3) is given in Figure 1:

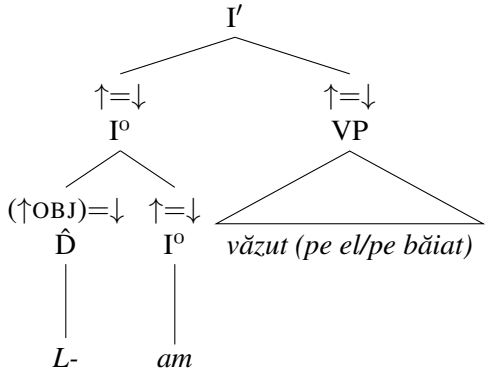


Figure 1: C-structure

The rule in (5) accounts for all clitics except the third person singular feminine clitic *-o*, which can cliticize to past participles as in (6):

(6) Ai păcălit-o.
have.2SG tricked-3SG.F
'You tricked her.'

Examples such as (6) are allowed by the following rule:

(7) $\mathrm{V}^{\mathrm{o}} \longrightarrow \underset{\uparrow=\downarrow}{\mathrm{V}^{\mathrm{o}}} \quad \underset{\substack{(\uparrow \mathrm{OBJ})=\downarrow \\ (\uparrow \mathrm{OBJ\ NUM}) = \mathrm{SG} \\ (\uparrow \mathrm{OBJ\ GEND}) = \mathrm{FEM} \\ (\uparrow \mathrm{OBJ\ PERS}) = 3}}{\hat{\mathrm{D}}}$
$(\uparrow \mathrm{PTCPL}) = \mathrm{PAST}$

We assume that all Modern Romanian object clitics are of the category $\hat{\mathrm{D}}$, and the distribution of the clitics is accounted for by the phrase structure rules that govern the distribution of $\hat{\mathrm{D}}$.[5] The word order facts are intriguing, and rules beyond (5) and (7) are needed to account for the full distribution. For example, it does not follow from the two rules above that the feminine clitic *-o* cannot precede the auxiliary in examples like (6).[6] However, we set the details of Romanian word order aside, since they are not directly relevant to our main focus here: the Romanian direct object clitics' status as agreement markers and incorporated pronouns.

[5]There might be other Romanian clitics that are of the category $\hat{\mathrm{D}}$; for example, the indirect object pronouns. If so, it is important to ensure that the right forms map onto the right $\hat{\mathrm{D}}$ nodes. This can be done with case specifications on the entries and rules, or by "constructive case" (Nordlinger, 1998).

[6]The placement of the third person feminine clitic is intricate. Zafiu et al. (2016, 2.2.3.7) describe

2.2 Uniform hypothesis 1: The clitics are agreement markers

Before the Romanian object clitics are given a formal analysis, two alternative hypotheses will be considered and rejected. The first hypothesis (presented in this subsection) is that the clitics are agreement markers whether or not an independent NP object is (overtly) present. The second hypothesis (presented in Section 2.3) is that the clitics are pronouns whether or not an independent NP object is present. Both of these hypotheses are uniform: the clitics are *not* ambiguous between agreement markers and pronouns.

The analysis that is most commonly associated with the term *pro*-drop assumes the presence of a phonologically empty pronoun ('little *pro*') in non-doubling cases. On this hypothesis, the third person plural masculine pronoun *i-* agrees with *ei* or *băieţi* in (8), and it agrees with an empty pronoun *pro* in (9):

(8) I-am văzut pe ei / pe băieţi.
3PL.M-have.1SG seen ACC them.M / ACC boys
'I saw the boys.'

(9) I-am văzut *pro*.
3PL.M-have.1SG seen
'I saw them.'

Analyzing the clitics uniformly as agreement markers is empirically problematic: the doubling clitics are governed by different restrictions than the non-doubling clitics. In MSR, clitic doubling occurs only with objects that are human, definite and marked with the preposition *pe*. The object-marking preposition *pe* can be compared to the Spanish *a*, and Romanian clitic doubling falls under "Kayne's generalization" (Kayne 1975; Aoun 1981, 275; Jaeggli 1981, 39), which states that clitics can only double prepositionally marked phrases. The non-doubling clitics refer more freely. These generalizations are illustrated below.

Non-doubling clitics can refer to animals and inanimates:

(10) Căţelul a furat cârnatul. L-am prins cu el în
dog.DEF has stolen sausage.DEF 3SG.M-have.1SG caught with it in
gură.
mouth
'The dog stole a sausage. I caught it with it in its mouth.'

(11) Tabloul este unicat. L-am cumpărat la o licitaţie.
painting.DEF is unique 3SG.M-have.1SG bought at an auction
'The painting is unique. I bought it at an auction.'

---

the distribution as follows: "In MR [Modern Romanian], *o* is always enclitic with the compound past tense, the analytic pluperfect (dialectal), the conditional, and gerundial periphrases, and proclitic with the *voi* infinitival future and with the future perfect. It can be either proclitic or enclitic with the perfect subjunctive."

However, note that the object *cârnatul* 'sausage' is not doubled by a clitic in (10). Although non-doubling clitics can refer to non-humans, clitics cannot double non-human NPs. This is further illustrated in (12):

(12) a. Am văzut melcul.
have.1SG seen snail.DEF
'I saw the snail.'

b. * L-am văzut (pe) melc.
3SG.M-have.1SG seen ACC snail

Clitics also cannot double indefinites:

(13) a. Am văzut un băiat.
have.1SG seen a boy.
'I saw a boy.'

b. * L-am văzut (pe) un băiat.
3SG.M-have.1SG seen ACC a boy

Finally, (14) illustrates that clitics do not double objects that are not *pe*-marked:

(14) a. Am văzut (*pe) băiatul.
have.1SG seen ACC boy.DEF
'I saw the boy.'

b. * L-am văzut băiatul.
3SG.M-have.1SG seen boy.DEF

In sum, the doubling clitic (which will be analyzed here as an agreement marker) is restricted in ways that the non-doubling clitic (here, a pronoun) is not.

2.3 Uniform hypothesis 2: The clitics are pronouns

A second potential uniform analysis of the clitic is that it is uniformly a pronoun (cf. Alexopoulou's 1999 on Greek clitics, and Aoun 1981 on Romanian, Hebrew and Lebanese Arabic clitics). Compare *o-* in example (6) above to example (15):

(15) Ai păcălit-o pe Ioana.
have.2SG tricked-3SG.F ACC Joanna
'You tricked Joanna.'

According to this analysis, *-o* is uniformly a pronoun, and *pe Ioana* in (15) would be an adjunct, likely a right-dislocated adjunct. The reading of the example would be something like "You tricked her, Joanna" (with *her* and *Joanna* co-referring).

However, this hypothesis is problematic. First, the restrictions on doubling are equally unexpected under a uniform pronoun hypothesis as under a uniform agreement marker hypothesis (Section 2.2). For example, it is unclear why it would in principle be impossible to right dislocate an inanimate object.

Second, on the uniform pronoun hypothesis, *pe Ioana* is an adjunct in (15). However, proper names (like *Ioana*) *must* be doubled by a clitic:

(16) * Ai păcălit (pe) Ioana.
have.2SG tricked ACC Joanna
'You tricked Joanna.'

Since the clitic is obligatory here, the only way to express proper names as objects would be indirectly through a clitic (the "real object") on this hypothesis. Categorically disallowing proper name objects seems like an unusual constraint.

Third, the example in (15) does not display the typical characteristics of right dislocation. There is no intonational break before the object, and the object is not stressed. The sentences in (17–18) are actual right dislocation examples:

(17) Ai păcălit-o pe ea, Ioana.
have.2SG tricked-3SG.F ACC her Joanna
'You tricked her, Joanna.'

(18) L-am văzut pe el, proful.
3SG.M-have.1SG seen ACC him prof.DEF
'I saw him, the prof.'

For these reasons, we reject the hypothesis that the clitics are uniformly pronouns.

3 A lexical split hypothesis

In LFG, *pro*-drop is commonly formalized with an optional pronominal PRED feature in the relevant lexical entries. The lexical entry for the third person masculine singular clitic would then look something like (19) (this entry will be revised):

(19) *îl/l-:* ((↑ PRED) = 'pro')
(↑ PERS) = 3
(↑ NUM) = SG
(↑ GEND) = MASC

Each PRED feature value is unique, and PRED features can therefore not unify, unlike other features. This means that the clitic cannot co-occur with an independent object (with its own PRED feature) when the PRED feature is present. This holds whether the object is pronominal or not, since pronominal PRED features are also unique. When PRED is absent, the clitic can co-occur with an object on the condition that the other features match. It then functions as a regular agreement marker.[7] When the PRED feature is present, the clitic *is* the object pronoun, and its PRED feature contributes to the completeness of the f-structure.

The optionality of PRED means that there are in effect two lexical entries, one with a PRED feature and one without:

[7]Except it is a slightly unusual agreement marker, since agreement markers are more typically bound morphemes than clitics.

(20)

Pronoun:			**Agreement:**		
(↑ PRED)	=	'pro'			
(↑ PERS)	=	3	(↑ PERS)	=	3
(↑ NUM)	=	SG	(↑ NUM)	=	SG
(↑ GEND)	=	MASC	(↑ GEND)	=	MASC

This "lexical split" opens up the possibility for further differences between the entries, and the Romanian clitics indeed differ beyond the PRED feature.

The observant reader will notice that the lexical entries in (20) do not account for all the generalizations listed in Section 2: doubling is only possible with *pe*-marked, animate, definite objects. Following Cornilescu (2000) and others, *pe* is analyzed here as an accusative case marker. The requirement that agreement-marking clitics only double *pe*-marked nouns is captured here with a *constraining equation* demanding accusative case: (↑ CASE) $=_c$ ACC. We assume that the pronominal clitic (the clitic with a PRED feature) is also specified for case, since the object clitics differ in form from other clitics (e.g., dative clitics). However, the pronominal clitic directly contributes the case, so the accusative feature is introduced with a regular defining equation: (↑ CASE) = ACC.

The *pe*-morpheme is an instance of *differential object marking*. There are two ways to express definite and animate NP objects in Romanian, through *pe*-marking (21) and through *-ul*-marking (22):[8]

(21) L-am văzut pe băiat.
3SG.M-have.1SG seen ACC boy
'I saw the boy.'

(22) Am văzut băiatul.
have.1SG seen boy.DEF
'I saw the boy.'

The *pe*-marker is obligatory with personal pronouns and proper names with human referents:

(23) L-am vizitat pe el / pe Ion.
3SG.M-have.1SG visited ACC him / ACC John
'I visited him/John.'

(24) * Am vizitat el / Ion.
have.1SG visited him / John

Inanimate objects and objects with an indefinite article are not *pe*-marked: examples (12–13) above cannot take *pe*. Unmodified *-ul*-marked direct objects are not *pe*-marked either:

(25) * L-am vizitat pe băiatul.
3SG.M-have.1SG visited ACC boy

[8]The morphological definiteness marker is referred to as *-ul* here, to distinguish it from *pe* which is also only compatible with definites. However, *-ul* has several allomorphs.

Onea & Hole (2017) propose that *pe*-marking is restricted to strong definites, but this proposal is difficult to reconcile with the fact that proper names are obligatorily *pe*-marked. An alternative account is provided by von Heusinger & Chiriacescu (2013), who suggest that *pe* marks discourse structuring potential; and further proposals are provided by Hill (2013); Ticio & Avram (2015), and Tigău (2015). There is no consensus in the literature about the conditions for the different types of objects, and we will not try to adjudicate between the proposals here.

Although *pe*-marking is not the topic of this paper, it is closely tied to clitic-doubling and we therefore cannot completely set it aside. Clitic doubling and *pe*-marking almost always co-occur, but there is nevertheless some evidence that the animacy requirement is tied to the *pe*-marker and not (necessarily) to the clitic itself. Clitic doubling only occurs with *pe*, but *pe* can occur without a clitic. In examples with *pe* and without a clitic, the animacy requirement remains:

(26) Petru n-a văzut pe nimeni.
Peter not-have.3SG seen ACC nobody
'Peter didn't see anybody.'

(27) Petru a văzut pe cineva / pe careva.
Peter have.3SG seen ACC somebody / ACC somebody
'Peter saw somebody.'

(28) Pe cine ai văzut?
ACC who have.2SG seen
'Who did you see?'

(29) Petru a văzut ceva. / * Petru a văzut pe ceva.
Peter have.3SG seen something
'Peter saw something.'

The objects in (26–28) are all specified [HUMAN +], and *pe* is obligatory. The object *ceva* in (29) is [HUMAN −] and cannot be *pe*-marked.

The *pe*-marked examples in (26–28) are human, but they are not definite. Clitic doubling only occurs on *pe*-marked definite objects. We propose that *pe* is specified for animacy but not definiteness, and the agreement clitic is specified for definiteness but not animacy. The revised lexical entries for the pronominal and agreement-marking clitics are given in (30):

(30)

Pronoun:			**Agreement:**		
(↑ PRED)	=	'pro'			
(↑ PERS)	=	α	(↑ PERS)	=	α
(↑ NUM)	=	β	(↑ NUM)	=	β
(↑ GEND)	=	γ	(↑ GEND)	=	γ
(↑ CASE)	=	ACC	(↑ CASE)	$=_c$	ACC
(↑ DEF)	=	+	(↑ DEF)	=	+

It is crucial for our analysis that the agreement marker is specified for definiteness. We assume that the pronominal clitic is also specified for definiteness, as personal pronouns typically have a definite interpretation.

Let us briefly return to the issue of human referents. The revised agreement-marking entry does not include a HUMAN feature, but it is in fact difficult to determine whether it should. The clitic only agrees with objects with human referents. However, this might be an indirect effect of the requirement that it can only agree with *pe*-marked NPs, and we have independent evidence that *pe*-marking is restricted to objects with human referents. We therefore take the more conservative view that the agreement clitic is not specified for animacy at all, since a HUMAN specification on the clitic is formally unnecessary. On the other hand, Hill (2013) and David (2015) show that the animacy requirement developed independently on the clitic and *pe*, and the clitics never double inanimates. If we were to assume that the agreement marker has an additional specification $(\uparrow \text{HUMAN}) = +$, there would be no empirical consequences. This would then be a further difference between the pronominal clitic and the agreement-marking clitic, as the pronominal clitic can refer freely to animals and inanimates (10–11).

The data presented so far suggest an analysis where the *pe* is specified with the feature [HUMAN +] and can therefore only mark accusative case on [HUMAN +] nouns. However, the data are a bit more complicated, as illustrated by (31):

(31) Am probat şapte paltoane. L-am cumpărat pe cel
have.1SG tried seven trench coats. 3SG.M-have.1SG bought ACC the
mai frumos.
nicest
'I tried seven trench coats. I bought the nicest.'

Example (31) has clitic doubling and *pe*, even though the superlative (without an overt head noun) refers back to a noun that is not [HUMAN +]. Whether we assume that *cel mai frumos* is nominalized (a deadjectival noun) or it contains an empty *one*-type pronominal, it is clear that it does not have the feature [HUMAN +] in (31). *Cel mai frumos* can refer back to human nouns in other examples, so the simplest analysis is that *cel mai frumos* is unspecified for HUMAN; that is, it has no HUMAN feature at all. If this is the case, then *pe* cannot be specified with a [HUMAN +] feature. Since *cel mai frumos* is unspecified for HUMAN, the [HUMAN +] feature from *pe* would simply be added to the feature structure in *pe cel mai frumos*, which does not work for (31). Instead, we propose that *pe* is lexically specified as follows:

(32) *pe:* $(\uparrow \text{CASE}) = \text{ACC}$
$(\uparrow \text{HUMAN}) \neq -$

The equation $(\uparrow \text{HUMAN}) \neq -$ ensures that *pe* does not unify with nouns specified as [HUMAN −]. Nothing prevents it from unifying with [HUMAN +] nouns, or with nouns which are unspecified for the [HUMAN] feature.

Let us summarize the main points of this section. Following previous work in LFG on "*pro*-drop", the Romanian object clitic is analyzed here as ambiguous between a pronoun (with [PRED 'pro']) and an agreement marker (without [PRED 'pro']). The agreement marker is more restricted than the pronoun and co-occurs only with definite, human, *pe*-marked objects. These generalizations are captured by the lexical entries in (30) and (32).

4 Diachronic variation

A wealth of research on the grammaticalization of pronouns and agreement marking (Givón & Li, 1976; Mithun, 1988; Hopper & Traugott, 1993) has shown that the diachronic development typically follows the path in (33):

(33) independent pronoun > weak pronoun > clitic pronoun > agreement affix > fused agreement marker

Numerous previous analyses of *pro*-drop have noted that it is quite unsurprising that pronoun/agreement ambiguities should emerge given the grammaticalization cline in (33) (e.g., Fassi Fehri 1984; Bresnan & Mchombo 1987; Toivonen 2001; Morimoto 2002; Butt 2007; Coppock & Wechsler 2010). When pronouns transition into agreement affixes, it seems natural that there could (at least in some cases) be a stage where the forms are not immediately reanalyzed as wholesale agreement, but instead are agreement markers when they double an NP and pronouns when they do not.

Since the mappings between the constituent structure, the feature structure, the lexicon, and the prosodic structure are quite flexible, the ambiguous stage between pronoun and agreement marker is not tied to only one kind of morphosyntactic realization. A linguistic element can in principle be ambiguous between a pronoun and an agreement marker regardless of its status as an independent word, a clitic, a bound agglutinative morpheme, or a fused morpheme. The grammaticalization path in (33) thus conflates two common sequences of changes that are often parallel. The first sequence concerns prosody and c-structural realization:

(34) independent word > small word > true clitic > affix > fused affix

A "small word" can be a word that does not project a phrase but is phonologically independent *or* a word that projects a phrase but is phonologically dependent on a syntactic host. By 'true clitic', we mean a form that does not project a phrase and is phonologically dependent on a host, but is not a bound morpheme. The table in (35) from Toivonen (2003, 45) serves to illustrate the distinctions:

(35)

	NON-PROJECTING	PROJECTING
PHON. DEP.	French 'clitic' pronouns Serbo-Croatian pronouns and auxiliaries	Kwakwala & Yagua determiners English reduced auxiliaries Swedish genitive marker Finnish, Russian and Bulgarian question particles
PHON. INDEP.	Swedish verbal particles Yoruba weak pronouns	English lexical verbs Yoruba strong pronouns Estonian question particles

The second scale has to do with referential capacity:

(36) noun > pronoun > agreement marker

This scale does not concern the prosodic or phrase-structural realization of a linguistic entity. Nouns, pronouns and agreement markers typically have similar features (e.g., person, animacy, and case), but they differ in their PRED features. Nouns have a contentful nominal PRED feature, pronouns have the PRED feature 'pro', and agreement markers have no PRED feature at all.

Even though changes along the cline in (34) tend to be closely tied to changes along (36), the two scales are not intrinsically connected. This disconnect is carefully investigated in van Rijn (2016), who concludes: "loss of referentiality correlates with a loss in form, but in a relative rather than an absolute sense [...] function and form evolve in the same direction, but need not evolve at the same pace".

Before presenting some historical data from Romanian, we want to clarify two points. First, this paper focuses on the scale in (36), and not (34). As mentioned above, we assume that the relevant elements are true clitics, but they may in fact be bound affixes. It is also possible that some of the Romanian forms are clitics and some are affixes, and there might well be differences between dialects concerning the prosodic and morphological status of the "clitics". We think this question is important, but the LFG architecture allows us to focus on the pronoun/agreement marker status without taking a stand on the clitic/bound morpheme status.

Second, although we agree with the claims from the historical linguistics literature that certain kinds of changes often follow specific grammaticalization scales, we do not argue that change *must* happen this way. We furthermore do not claim that pronouns changing into agreement markers *must* go through a stage of optionality exactly as proposed here. The point is rather that the optionality of Romanian object clitics seems quite natural when considering the grammaticalization path that has been argued to be the origin of agreement marking in many different languages.

We propose that Romanian object clitics are following a succession of stages as outlined in (37):

(37)

	clitic	NP
STAGE 1:	pronoun	adjunct
STAGE 2:	pronoun	
or	agreement marker	object
STAGE 3:	agreement marker	object

The current stage is Stage 2. During a previous stage, the clitics were unambiguous pronouns. When doubled, the doubling NP was an adjunct, not an object. The next natural stage would be Stage 3, where the clitics are unambiguous agreement markers. It is of course not possible to say whether the clitics will reach that stage. Again, certain changes along the scale in (34) tend to occur more or less concurrently with the pronoun to agreement marker shifts, but we set this aside here.

On this view, the clitic would have been optional during Stage 1, with doubling under certain discourse conditions. Until recently, the clitic was indeed optional (see, e.g., *The Grammar of the Romanian Academy* 1963; Zafiu et al. 2016, Section 2.4.2.5), and examples such as (38–39) (with no object clitic) were possible:

(38) Pre tine vădzuiu.
ACC you.SG seen.1SG
'I saw you.' (Zafiu et al. 2016, Section 2.4.2.5, [1683])

(39) Domnul Domnezeu făcu pre om.
Lord.DEF God made ACC man
'The Lord God made the man.' (Zafiu et al. 2016, Section 2.4.2.2, [1582])

Equivalent examples in MSR are unacceptable. In (38–39), *pre tine* and *pre om* (*pre* is an old form of *pe*) are regular objects, and not adjuncts added as afterthoughts.

The fact that the clitic used to be optional is consistent with our proposal. Another important part of the puzzle would be to show that the NPs were adjuncts in doubling examples. However, distinguishing arguments from adjuncts is in general not easy (see, e.g., Whaley 1993; Needham & Toivonen 2011), and it is especially difficult when the evidence is restricted to written records. Evidence from intonation and native speakers' grammaticality judgements are not available.

One potential source of evidence for adjuncthood is word order: arguments tend to be more restricted in terms of phrase-structural position, and adjuncts are often ordered after arguments (Jackendoff, 1977; Pollard & Sag, 1987, i.a.). However, word order in Old Romanian is quite free, even freer than in Modern Romanian (Nicolae, 2016), so it is difficult to draw any conclusions based on ordering.[9]

Another indication that the doubled NPs were indeed adjuncts in Old Romanian is that doubling was much less restricted than it is now. In fact, it seems like it was syntactically quite unrestricted: Zafiu et al. (2016, Section 2.4.2.5) claim that there

[9]Clitics have a higher degree of freedom in their placement relative to the verb in Old Romanian than in Modern Romanian (Zafiu et al., 2016, Section 2.2.2.1). This indicates that the Old Romanian "clitics" were in fact free-standing words. Since the changes in (34) and (36) tend to go hand-in-hand, it is not surprising to find indications that the clitics were until recently independent words.

were "no proper rules of clitic doubling" in Old Romanian. Repetition of discourse participants through adjunction (in phenomena such as right or left dislocation) is restricted through discourse considerations such as emphasis and afterthoughts. The restrictions on true agreement are clearer and easier to detect in a text, since the restrictions are based on the syntax within the clause. In order to understand discourse constraints, we need access to larger corpora, and to a certain extent speaker/writer intentions. It seems that there were no *syntactic constraints* on clitic doubling in Old Romanian: clitic doubling was possible with non-human objects, *-ul*-marked objects, and it did not have to co-occur with *p(r)e*. For example, the object *flămândzii* in (40) is *-ul*-marked and there is no *p(r)e*:

(40) flămândzii săturaţi-i
hungry.PL.DEF.ACC feed.IMP.2PL-3PL.M
'feed the hungry' (Zafiu et al. 2016, Section 2.4.2.5, [1601])

The lack of syntactic constraints on clitic doubling in Old Romanian is consistent with an adjunction hypothesis and problematic on an agreement hypothesis.

We suggest that the first change was that the PRED feature became optional:

(41)

Stage 1:			**Stage 2:**		
(↑ PRED)	=	'pro'	((↑ PRED)	=	'pro')
(↑ PERS)	=	α	(↑ PERS)	=	α
(↑ NUM)	=	β	(↑ NUM)	=	β
(↑ GEND)	=	γ	(↑ GEND)	=	γ
(↑ CASE)	=	ACC	(↑ CASE)	=	ACC
(↑ DEF)	=	+	(↑ DEF)	=	+

Note that "Stage 2" in (41) is very close to our proposed lexical entry for MSR. One further change occurred that led to the current stage: the accusative case feature morphed into a requirement for an overtly case marked (*pe*-marked) NP object. This change was formalized above as a constraining equation, and it occurred in the agreement-marking clitic but not the pronominal clitic.

The literature on the historical development of clitic doubling and *pe* sheds light on the emergence of the agreement marker's *pe* requirement. Clitic doubling and *pe*-marking developed separately, and the doubling developed after *pe* (Chiriacescu, 2007; von Heusinger & Onea Gáspár, 2008; Hill, 2013; Tigău, 2014; David, 2015). Even though clitic doubling and *pe*-marking are not intrinsically connected, several authors have argued that they serve overlapping (though distinct) discourse functions (Leonetti, 2008; Hill, 2013; Tigău, 2014). It is still unclear exactly what those discourse functions are, but it seems that discourse prominence and high referentiality are relevant notions (Leonetti, 2008; Hill, 2013; Tigău, 2014). Since the functions are overlapping, *pe* and doubling often co-occurred even before this was a grammatical requirement. We propose that the frequent co-occurrence of the forms led to a reanalysis where the co-occurrence became a morphosyntactic requirement: Modern Romanian agreement clitics require *pe*.

5 Synchronic variation

This section will present two additional documented varieties of current Romanian. The variation can be modelled quite readily with the lexical entries in (30) above as a starting point. The varieties involve only minor featural differences.

The first variety is presented by Tigău (2010, 2014). Tigău reports that some speakers of Romanian allow clitic doubling with indefinites:

(42) Petru (l-)a vizitat pe un prieten.
Peter 3SG.M-have.3SG visited ACC a friend
'Peter visited a friend.'

Even the speakers who allow doubling with indefinite objects allow it only sometimes. Tigău (2010, 2014) argues that doubled indefinite objects get a *specific* interpretation (see also Aoun 1981, Chapter 3).

The difference between MSR and the indefinite-doubling dialect described by Tigău is captured with minimally different lexical entries. Recall that both the pronominal clitic and the agreement marker are specified as [DEFINITE +] in MSR (30). In the dialect considered here, the pronoun is the same as in MSR, but the agreement marker differs by being marked for specificity instead of definiteness:

(43)

Pronoun:			**Agreement:**		
(↑ PRED)	=	'pro'			
(↑ PERS)	=	α	(↑ PERS)	=	α
(↑ NUM)	=	β	(↑ NUM)	=	β
(↑ GEND)	=	γ	(↑ GEND)	=	γ
(↑ CASE)	=	ACC	(↑ CASE)	$=_c$	ACC
(↑ DEF)	=	+	(↑ SPECIFIC)	=	+

In both dialects, clitics double only *pe*-marked objects, hence the constraining equation for accusative case in the agreement marker (which is the version of the clitic that doubles the object).

The Aromanian dialect (AR; spoken in Albania, Macedonia, Romania, Bulgaria, Serbia and Croatia) and the Megleno-Romanian dialect (MR; spoken in Greece and Macedonia) differ from the varieties discussed above. In AR and MR, all and only definite objects are doubled (Tomić 2006, Chapter 4; Tomić 2008, 84; Hill 2013). There is no *pe* requirement; in fact, AR and MR do not have accusative *pe*-marking at all. There is also no animacy requirement, which follows if the [HUMAN] requirement is associated with the *pe*.

Sentence (44) is an MR example. The clitic *l-* doubles the object *filmu*, which is inanimate and carries definiteness morphology, but is not *pe*-marked.

(44) L-am vizut filmu.
3SG.M-have.1SG seen film.DEF
'I saw the film.' (Tomić, 2006, 295)

The lexical entries for the pronominal and agreement-marking clitics in AR and MR (given in (45)) are identical except for the PRED feature:

(45)

Pronoun:			**Agreement:**		
(↑ PRED)	=	'pro'			
(↑ PERS)	=	α	(↑ PERS)	=	α
(↑ NUM)	=	β	(↑ NUM)	=	β
(↑ GEND)	=	γ	(↑ GEND)	=	γ
(↑ CASE)	=	ACC	(↑ CASE)	=	ACC
(↑ DEF)	=	+	(↑ DEF)	=	+

The AR/MR lexical entry for the pronoun is the same as the pronoun entry in the other dialects. However, the AR/MR agreement-marking clitic differs from both of the other dialects in that the equation for CASE is not a constraining equation. It further differs from the indefinite-doubling dialect in that it is marked for definiteness but not for specificity.

This section has extended the analysis proposed for MSR in Section 2 to other dialects of Romanian. The lexical split analysis makes it possible to consider the agreement-marking clitics separately from the pronominal clitics. The dialectal data show no indication that the pronominal clitics differ across dialects. However, the agreement markers display differences in their lexical features. These differences, which are formally minimal, have easily observable empirical effects.

6 Lexical splits cross-linguistically

The analysis of Romanian object clitics presented above adopts a lexical split hypothesis: each clitic form is associated with two lexical entries. The fact that the agreement markers and pronouns display differences beyond the PRED feature supports the analysis. Adopting standard assumptions about the grammaticalization of pronouns, it is not surprising that lexical splits like the Romanian one should emerge. If these claims are correct, then we should expect lexical splits to be quite common in agreement systems cross-linguistically. This section presents a list of examples of agreement systems with forms that seem to be ambiguous between agreement markers and pronouns, and where the difference goes beyond mere referential status.

Subject agreement in Modern Standard Arabic. Fassi Fehri (1984, 1988, 1993) carefully analyzes subject agreement in Modern Standard Arabic. He proposes that "some affixes have two different lexical entries" (Fassi Fehri, 1988, 119). The kinds of splits he describes look strikingly similar to the Romanian clitics. For example, he provides different lexical entries for the feminine affix *at*. Two entries are pronominal but differ in GENDER, NUMBER, PERSON and HUMAN features. A third entry is an agreement marker and has only one feature: [GENDER FEMININE].

Possessive suffixes in Finnish. Pronominal possession in Standard Finnish can be marked with an independent pronoun and a suffix on the possessed noun, or

a suffix alone. In first and second person, the independent pronoun is optional and expressions with or without the independent possessor have the same literal meaning (they differ in emphasis). When a third person independent pronoun is "dropped" and possession is marked by just a suffix, the possessor is necessarily bound by a subject within the minimal finite clause. Conversely, when an independent pronoun is present, the possessor *cannot* be bound by a subject. In Toivonen's (2000) analysis, the anaphoric suffix has a PRED feature and the suffix agreeing with a non-anaphoric independent pronoun does not. The entries also differ in that the agreement suffix is restricted to agreement with human personal pronouns.

Rioplatense Spanish object clitics. Varieties of Spanish display clitic systems very similar to that of Romanian (see, e.g., Mayer 2017). Andrews (1990) and Estigarribia (2013) analyze Rioplatense Spanish within an LFG framework. They both propose entries for pronominal clitics that differ from the agreement clitics beyond the PRED feature. In Estigarribia's analysis, the agreement marker has the following feature that the pronominal clitic lacks (p. 300): $\neg(\downarrow\text{SPECIFIC}) -$.

Pakin Lukunosh Mortlockese object suffixes. Odango (2014) argues that the object markers in the Micronesian language Pakin Lukunosh Mortlockese show a split. Most of the suffixes exclusively function as incorporated pronouns. However, when an independent object is present, the third person singular object suffix behaves like a general transitivity marker. Object markers often grammaticalize into transitivity markers (Lehmann, 2002; Mayer, 2017). A natural grammaticalization path is: independent object pronoun $>$ incorporated pronoun $>$ agreement marker $>$ transitivity marker.

7 Conclusion

Romanian object clitics can double direct objects, provided that the objects are animate, definite and *pe*-marked. The sections above described the clitic doubling in Modern Standard Romanian and provided an analysis in Lexical Functional Grammar. According to our analysis, the Romanian clitics are ambiguous between pronouns and agreement markers.

Baker & Kramer (2018) write: "Controversy and uncertainty have plagued the question of whether 'object markers' (OMs) are object pronouns cliticized to the verb or realizations of object agreement." They also note that "[i]t is an awkward fact that generative linguistics has had a hard time distinguishing reliably between pure agreement and clitic doubling (CD)." We believe that this awkwardness can be explained by the special challenges that pronouns and agreement marking pose for syntactic theory. Three challenges have been addressed in this paper.

First, as argued in this paper following Bresnan & Mchombo (1987) and others, it is not uncommon for the same forms to be ambiguous between agreement markers and pronouns. Second, it is often difficult to determine the morphological status of these elements because they display characteristics that are typical neither for regular, free standing, fully projecting words, nor for bound morphemes. In

other words, these elements often involve unusual mappings between p-structure and c-structure (see Section 4, and also Spencer & Luís 2012; Bögel 2015; Lowe 2016). Section 4 also mentioned the third challenge, which concerns the alignment between c-structure and f-structure. Prototypically, pronouns are some kind of "small words" or clitics, and agreement markers are bound morphemes. However, mismatches are possible and atypical alignment can occur: although it is possible to describe the typical c-structural expression of pronouns and agreement markers, these generalizations are not universal principles of grammar. Clitics can function as agreement markers, and pronouns can be morphologically incorporated.

References

Alexopoulou, Theodora. 1999. *The syntax of discourse functions in Greek: A non-configurational approach*. Edinburgh: University of Edinburgh dissertation.

Andrews, Avery. 1990. Unification and morphological blocking. *Natural Language and Linguistic Theory* 8(4). 507–558.

Aoun, Joseph. 1981. *The formal nature of anaphoric relations*. Cambridge: Massachusetts Institute of Technology dissertation.

Baker, Mark & Ruth Kramer. 2018. Doubled clitics are pronouns: Amharic objects (and beyond). *Natural Language and Linguistic Theory* 36(4). 1035–1088.

Bögel, Tina. 2015. *The syntax-prosody interface in Lexical Functional Grammar*. Konstanz: University of Konstanz dissertation.

Bresnan, Joan, Ash Asudeh, Ida Toivonen & Stephen Wechsler. 2016. *Lexical-functional syntax, second edition*. Hoboken, NJ: Wiley-Blackwell.

Bresnan, Joan & Sam Mchombo. 1987. Topic, pronoun, and agreement in Chicheŵa. *Language* 63(4). 741–782.

Butt, Miriam. 2007. The role of pronominal suffixes in Punjabi. In Annie Zaenen, Jane Simpson, Tracy Holloway King, Jane Grimshaw, Joan Maling & Christopher Manning (eds.), *Architectures, rules, and preferences: Variations on themes by Joan Bresnan*, Stanford: CSLI.

Chiriacescu, Sofiana. 2007. Pe-Markierung und Diskurz-Prominenz im Rumänischen. Magisterarbeit, University of Stuttgart.

Coppock, Elizabeth & Stephen Wechsler. 2010. Less-travelled paths from pronoun to agreement: The case of the Uralic objective conjugations. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG11 conference*, 165–185. Stanford: CSLI.

Cornilescu, Alexandra. 2000. Notes on the interpretation of the prepositional accusative. *Bucharest Working Papers in Linguistics* 2. 91–106.

David, Oana. 2015. Clitic doubling and differential object marking: A study in diachronic construction grammar. *Constructions and Frames* 7(1). 103–135.

Dobrovie-Sorin, Carmen. 1994. *The syntax of Romanian: Comparative studies in Romance*. Berlin/New York: Walter de Gruyter.

Estigarribia, Bruno. 2013. Rioplatense Spanish clitic doubling and "tripling" in Lexical-Functional Grammar. In Chad Howe, Sarah A. Blackwell & Margaret Lubbers Quesada (eds.), *Selected proceedings of the 15th Hispanic Linguistics Symposium*, 297–309. Somerville, MA: Cascadilla Proceedings Project.

Fassi Fehri, Abdelkader. 1984. Agreement in Arabic, binding and coherence. In *Conference on agreement in natural language*, Stanford University.

Fassi Fehri, Abdelkader. 1988. Agreement in Arabic, binding and coherence: Approaches, theories, descriptions. In Michael Barlow & Charles Ferguson (eds.), *Agreement in natural language*, 107–158. Stanford: CSLI.

Fassi Fehri, Abdelkader. 1993. *Issues in the structure of Arabic clauses and words*. Dordrecht: Kluwer.

Givón, Talmy & Charles Li. 1976. Topic, pronoun and grammatical agreement in subject and topic. In Charles Li (ed.), *Subject and topic*, 1551–188. New York: Academic Press.

von Heusinger, Klaus & Edgar Onea Gáspár. 2008. Triggering and blocking effects in the diachronic development of DOM in Romanian. *PROBUS* 20(1). 67–110.

Hill, Virginia. 2013. The direct object marker in Romanian: a historical perspective. *Australian Journal of Linguistics* 33(2). 140–151.

Hopper, Paul & Elizabeth Closs Traugott. 1993. *Grammaticalization*. Cambridge, UK: Cambridge University Press.

Jackendoff, Ray. 1977. *X-bar syntax: A study of phrase structure*. Cambridge, MA: MIT Press.

Jaeggli, Osvaldo. 1981. *On some phonologically null elements in syntax*. Cambridge: Massachusetts Institute of Technology dissertation.

Kayne, Richard. 1975. *French syntax: The transformational cycle*. Cambridge, MA: MIT Press.

Lehmann, Christian. 2002. *Thoughts on grammaticalization, 2nd edition*. Erfurt: Arbeitspapiere des Seminars für Sprachwissenschaft der Universität Erfurt.

Leonetti, Manuel. 2008. Specificity in clitic doubling and in differential object marking. *PROBUS* 20(1). 33–66.

Lowe, John. 2016. Clitics: Separating syntax and prosody. *Journal of Linguistics* 52(2). 375–419.

Luís, Ana Alexandra Ribeiro. 2004. *Clitics as morphology*. Essex: University of Essex dissertation.

Mayer, Elisabeth. 2017. *Spanish clitics on the move: Variation in time and space*. Berlin: De Gruyter Mouton.

Mithun, Marianne. 1988. The development of bound pronominal paradigms. In Winfred P. Lehman & Helen-Jo Jakusz Hewitt (eds.), *Language typology: Typological models in reconstruction*, 85–104. Amsterdam: John Benjamins.

Monachesi, Paola. 1998. The morphosyntax of Romanian cliticization. In P. Coppen, H. van Halteren & L. Teunissen (eds.), *Proceedings of Computational Linguistics in the Netherlands 1997*, 99–118.

Morimoto, Yukiko. 2002. Prominence mismatches and differential object marking in Bantu. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG02 conference*, 292–314. Stanford: CSLI.

Needham, Stephanie & Ida Toivonen. 2011. Derived arguments. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG11 conference*, 401–421. Stanford: CSLI.

Nicolae, Alexandru. 2016. Word order and configurationality. In Gabriela Pană Dindelegan (ed.), *The syntax of old Romanian*, Chapter 10. Oxford: Oxford Scholarship Online.

Nordlinger, Rachel. 1998. *Constructive case: Evidence from Australian languages*. Stanford: CSLI.

Odango, Emerson Lopez. 2014. The status of subject and object markets in pakin lukunosh mortlockese. *Oceanic Linguistics* 53(1). 110–135.

Onea, Edgar & Daniel Hole. 2017. Differential object marking of human definite direct objects in Romanian. *Revue roumaine de linguistique* 62(4). 359–376.

Pollard, Carl & Ivan Sag. 1987. *Information-based syntax and semantics*. Stanford: CSLI.

Popescu, Alexandra. 2000. The morphophonology of the Romanian clitic sequence. *Lingua* 110(1). 773–799.

Spencer, Andrew & Anna R. Luís. 2012. *Clitics: An introduction*. Cambridge: Cambridge University Press.

Ticio, Emma & Larisa Avram. 2015. The acquisition of differential object marking in Spanish and Romanian: Semantic scales or semantic features? *Revue roumanie de linguistique* 60(4). 383–401.

Tigău, Alina Mihaela. 2010. Towards an account of differential object marking in Romanian. *Bucharest Working Papers in Linguistics* 12(1). 137–158.

Tigău, Alina Mihaela. 2014. Specificity effects with clitic doubling and *pe* marking. *Bucharest Working Papers in Linguistics* 16(1). 43–62.

Tigău, Alina Mihaela. 2015. Romanian differential object marking and presupposition. *Cahiers de linguistique française* 32. 215–227.

Toivonen, Ida. 2000. The morphosyntax of Finnish possessives. *Natural Language and Linguistic Theory* 18(3). 579–609.

Toivonen, Ida. 2001. Language change, lexical features and Finnish possessives. In Miriam Butt & Tracy Holloway King (eds.), *Time over matter: Diachronic perspectives on morphosyntax*, 209–225. Stanford: CSLI.

Toivonen, Ida. 2003. *The phrase structure of non-projecting words: A case study of Swedish particles*. Dordrecht: Kluwer.

Tomić, Olga Mišeska. 2006. *Balkan Sprachbund morphosyntactic features*. Dordrecht: Springer.

Tomić, Olga Mišeska. 2008. Towards grammaticalization of clitic doubling: Clitic doubling in Macedonian and neighbouring languages. In Dalina Kallulli & Liliane Tasmowski (eds.), *Clitic doubling in the Balkan languages*, 65–87. Amsterdam/Philadelphia: John Benjamins.

van Rijn, Marlou. 2016. The grammaticalization of possessive person marking: A typological approach. *Transactions of the Philological Society* 114(2). 233–276.

von Heusinger, Klaus & Sofiana Chiriacescu. 2013. The discourse structuring potential of differential object marking: the case of indefinite and definite indirect objects in Romanian. *Revue roumaine de linguistique* 58(4). 439–456.

Whaley, Lindsay. 1993. *The status of obliques in linguistic theory*. Buffalo: State University of New York dissertation.

Zafiu, Rodica, Alexandru Nicolae, Dana Niculescu, Gabriela Pană Dindelegan, Irina Nicula Paraschiv, Adina Dragomirescu & Andra Vasilescu. 2016. The verb and its arguments: the root clause. In Gabriela Pană Dindelegan (ed.), *The syntax of Old Romanian*, Chapter 2. Oxford: Oxford Scholarship Online.

Vafsi Oblique Pronouns: Stress-related placement patterns

Tina Bögel
University of Konstanz

Saeed Reza Yousefi
Shahid Beheshti University

Mahinnaz Mirdehghan
Shahid Beheshti University

Proceedings of the LFG'18 Conference

University of Vienna

Miriam Butt, Tracy Holloway King (Editors)

2018

CSLI Publications

pages 88–108

http://csli-publications.stanford.edu/LFG/2018

Keywords: Vafsi, clitics, syntax-prosody interface

Bögel, Tina, Yousefi, Saeed Reza, & Mirdehghan, Mahinnaz. 2018. Vafsi Oblique Pronouns: Stress-related placement patterns. In Butt, Miriam, & King, Tracy Holloway (Eds.), *Proceedings of the LFG'18 Conference, University of Vienna*, 88–108. Stanford, CA: CSLI Publications.

Abstract

This paper discusses and analyses the distribution of oblique pronoun clitics in Vafsi. In contrast to earlier studies (Stilo, 2004a,b, 2010), we show that oblique clitics in Vafsi do not have affixal counterparts, but that all instances and forms of the clitic can be explained with reference to prosodic constraints. In cases where the prosodically deficient oblique enclitic is left without a suitable host, prosodic inversion aims to place the clitic accordingly. The erstwhile 'affixal form' corresponds to the cases where the clitic is forced to carry stress itself and consequently assumes a 'full form'. We furthermore provide a complete formal analysis of the oblique pronoun clitics at the syntax-prosody interface in LFG (Bögel, 2015).

1 Introduction

Vafsi is a Northwestern Iranian language spoken by approximately 20,000 people in two dialects. It does not have a standard written form and information on its grammar is sparse. The main source of Vafsi are recordings of folk tales by Lawrence P. Elwell-Sutton from 1958 and their transcription, translation, and short linguistic analysis by Donald L. Stilo (Stilo, 2004b). Supplemented by further field study material, Stilo also wrote two follow-up papers on coordination and ditransitives in Vafsi (Stilo, 2004a, 2010). Most material used in the following discussion comes from these sources and from Mirdehghan and Yousefi (2016), and was further confirmed by one of the co-authors, Saeed Yousefi, who is a native speaker of the language.

Vafsi is a non-rigid verb-final language where the postverbal positions are determined by information structural constraints. It largely follows a tense/aspect-based split ergative system. Furthermore, it has differential object marking that seems to depend on animacy and specifity. Vafsi distinguishes between direct and oblique case marking, which is reflected in the three pronoun realisations in Vafsi: Independent pronouns, pronoun bases[1], and pronominal clitics. The paper's main focus lies on the oblique pronominal clitics, their distribution, and their 'affixal counterparts'.

The following table shows the direct and oblique sets of clitics and their 'affixal counterparts' as they are described in Stilo (2010, with adjustments reflecting pronunciation).

[†]We would like to thank the LFG2018 audience in Vienna for invaluable comments and the AFF of the University of Konstanz for their financial support, which allowed the first author to travel to Iran and thus made this paper possible.

[1]There are two oblique pronoun bases which can serve as a host for the clitics: *hazun* and *verewn*, which are used in non-subject functions, have no semantic content, and can occur postverbally (unlike the clitics by themselves). Person must be obligatorily indicated by a pronoun clitic and the distribution seems to be related to case (Stilo, 2004b, 227), but more research is needed in this particular area.

	direct *(set 1)*		**oblique** *(set 2)*	
	enclitics (copulas)	*affixes*	*enclitics/ proclitics*	*affixes*
1SG	=im(e)	-om(e)	=om	-im-
2SG	=i	-i	=i	-i-
3SG	=e (m.)/=oæ(f.)	(V)-e / (C)-∅	=es	-is-
1PL	=am(e)	-am(e)	=owan	-iwan-
2PL	=a	-a	=ian	-ian-
3PL	=end(e)	-end(e)	=esan	-isan-

Table 1: Oblique and direct pronouns in Vafsi (Stilo, 2010)

While the direct pronominal markers are usually suffixed to the verb, the oblique pronoun clitics always occur preceding the verbal complex where they prosodically attach to a preceding host.

(1) ketab=**i** d-do-*m*
book=2S.OBL DUR-give-1S.DIR
'I'll give you a book.' (Stilo, 2010, 270)

While the direct pronominal markers most often occur as affixes and the oblique pronominal markers are usually clitics, these categorical classifications seem to change under specific circumstances. According to Stilo (and as shown in Table 1), the direct affixes sometimes can occur as clitics, and the oblique clitics can occur as affixes. The following minimal pair shows the occurence of the first person oblique pronoun as a clitic (2a) and as an affix (2b).

(2) a. an=**om** ær-góæ
that=1S.OBL DUR-want
'I want that'

b. **ìm**-ær-góæ
1S.OBL-DUR-want
'I want' (Stilo, 2010, 247)

In this paper we will show that this separation of the oblique pronominal markers into two different forms is not in fact a categorical distinction into clitics and affixes, but that the difference in form can be explained with reference to the clitic's position with respect to prosodic domains and stress distribution, i.e., the oblique 'affixes' are in fact 'clitics under stress'.[2]

2 Vafsi oblique pronoun clitics

The placement of oblique pronoun clitics follows a mostly regular pattern: The clitic is placed directly preceding the verbal complex (vc). The prosodically deficient enclitic follows a host, which is not limited to a particular word category or function as shown in the following examples.

[2]Whether a similar claim can be made for the direct pronoun distinction into affixes and clitics remains to be investigated.

(3) soan-e=ra bez-e šax**=es** [tíz=a kærdæ]$_{VC}$
file-F.OBL=with goat-F.OBL horn=3S.OBL sharp=ATTR did
'He sharpened the goat's horns with a file.' (Stilo, 2004b, 291)

(4) tani hæzíri**=m** [bǽ-diæ]$_{VC}$
he.OBL yesterday=1S.OBL PUNCT-saw
'I saw him yesterday.' (Stilo, 2010, 247)

The size or complexity of the verbal complex does not seem to have an impact on the placement of the clitics. Whether the verbal complex contains a simple verb ((5)) or a complex predicate ((6)), the clitic is placed preceding the verbal complex.

(5) ya qærri**=es** [bǽ-košdé]$_{VC}$
or witch=3S.OBL PUNCT-killed
'... or he killed the witch.' (Stilo, 2004b, 244)

(6) bǽlke hævi-án**=es** [komǽk ær-kærdæ]$_{VC}$
but all-PL.OBL=3S.OBL help DUR-did
'... but he helped everybody.' (Stilo, 2004a, 305)

However, there are also a number of occurences where the clitic is not placed according to the regular pattern. In certain circumstances, the clitic can also occur between the two members of a complex predicate, attaching itself to the first element.

(7) æ-cu ešden bǽ-vær-i ya [komǽk**=i** kær-òm]$_{VC}$
DUR-an SELF PUNCT-take-2S.OBL or help=2S.OBL do-1S.DIR
'Can you carry it yourself or should I help you?' (Stilo, 2004a, 148)

As noted in example (2), repeated in (8), the clitic can occur as an 'affix'. This 'affixal form' is furthermore not limited to the sentence-initial position as shown in (9).

(8) a. an**=om** [ær-góæ]$_{VC}$
that=1S.OBL DUR-want
'I want that'

b. [**ìm**-ær-góæ]$_{VC}$
[1S.OBL-DUR-want]$_{VC}$
'I want' (Stilo, 2010, 247)

(9) bá-waz ya [**ì-**r-koš-ome]$_{VC}$
PUNCT-tell or 2S.OBL-DUR-kill-1S.OBL
'Tell (me) or I will kill you' (Stilo, 2004b, 312)

In addition, the clitic can occur *verb-medially* in its 'affixal' form, following either the punctual marker, the negative marker, or a preverb. The following minimal pair shows the clitic preceding ((10a)) and following ((10b)) the punctual marker *bæ*. Structures like these can also occur if a host outside of the verbal complex is seemingly available ((11)).

(10) a. án=**om** [bǽ-diæ]$_{VC}$
that=1S.OBL PUNCT-saw
'I saw that'

b. [b-**ím**-diæ]$_{VC}$
PUNCT-1S.OBL-saw
'I saw' (Stilo, 2010, 247)

(11) bærzegǽr-i [v-**ís**-vattæ]$_{VC}$
farmer-OBL PUNCT-3S.OBL-said
'The farmer said ...' (Stilo, 2004b, 239)

The verb-medial occurences are not reduced to the 'affixal' form. As can be seen in the following examples, the clitic form can be placed between a preverb and the main verb stem.

(12) tinan vǽxdi=ke nahar=esan [hár=**es**-da]$_{VC}$ ...
they.OBL when=SUB lunch=3P.POSS PVB=3S.OBL=gave
'When she (=*es*) gave them (*tinan*) their (=*esan*) lunch' (Stilo, 2010, 254)

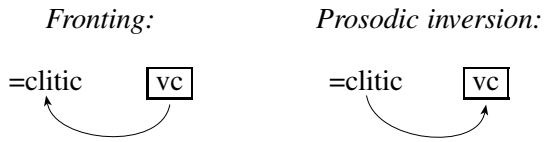
(13) bæd-æz karǽ-i ke [hár=**esan**-kærdæ]$_{VC}$
after-from things-INDEF SUB PVB=3P.OBL=did
'After the things they did ...' (Stilo, 2004a, 290)

As an intermediate conclusion it can be stated that the clitics usually occur in the position immediately preceding the verbal complex, but can also occur within the verbal complex separating members of a complex predicate and even within otherwise non-dividable parts of the verb. Furthermore the oblique pronouns appear mostly in their clitic form, but under specific circumstances occur in their 'affixal form'. The following table gives an overview.

	Position	**Examples**	**Form**
1.	preceding the verbal complex (non-initial position)	(3) – (6)	clitic
2.	between the members of a complex predicate	(7)	clitic
3.	preceding the duration marker	(8b), (9)	'affix'
4.	following the punctual marker, the negation marker, or a preverb	(10b) – (13)	'affix'

Table 2: Distribution of oblique pronoun clitics

To account for these differences, Stilo assumes that the clitic originates within the verbal complex/the verb and is 'fronted' if an adequate host is available (Stilo, 2004b, 238). However, there is no unified reason as to why the clitic would be fronted in examples (3-6), but not in examples (7-13). This paper, on the other hand, offers an explanation in prosodic terms. We claim that oblique clitics originate in the position preceding the verbal complex. If there is no suitable host available to the left, the clitics are either forced to stay in situ or are 'moved' to an adequate position via *prosodic inversion* (Halpern, 1995). In the case of prosodic inversion, the clitic is placed after a stressed host to its right.



As briefly mentioned in fn 1, the oblique clitics can never occur postverbally without a pronoun base (and then only in information-structurally marked contexts). The most likely explanation for this constraint is the possible confusion with the direct pronominal markers which are placed immediately following the verb and which have, to some extent, the same (phonological) forms as the oblique markers. Consequently, if the postverbal position is the only target position available to prosodic inversion, prosodic inversion must not apply in order to avoid confusion. However, since a prosodically deficient enclitic cannot remain in the initial position, the clitic has to assume a stressed 'full form', Stilo's former 'affixal form' as shown in Table 1. This stressed 'full form' also occurs if the clitic is moved via prosodic inversion and ends up in a position where it receives stress as the result of a postlexical phonological stress placement rule (see below).

3 Relevant aspects of Vafsi grammar

Before oblique clitics can be discussed in more detail and before the earlier claim about the clitics being subjected to prosodic inversion can be verified, further aspects of the Vafsi grammar have to be introduced. This section will therefore provide more information on some intonational patterns found in Vafsi, on the verbal complex and some of its members, and on the expression of the possessive, as they provide crucial insights into the analysis of the oblique clitic pronouns.

3.1 Intonational patterns

One of the claims made above states that if the prosodically deficient oblique enclitic is stranded in the initial position of a prosodic domain, the enclitic has to be a) placed in a suitable position via prosodic inversion, or b) assume its stressed full form. Larger prosodic domains in the prosodic hierarchy (McCawley, 1968; Selkirk, 1978) are the intonational phrase (ι, usually corresponding to a syntactic IP or CP) and the phonological phrase (φ, usually an XP) (Ladd, 1986; Selkirk, 2011).[3] So far, there is no detailed research on prosodic phrasing in Vafsi. Based on the folk tale recordings, however, Stilo (2004a,b) was able to distinguish basic patterns of Vafsi prosody, some of which will be briefly introduced in this section, as they add to the analysis of the oblique clitic pronoun and offer a fundamental explanation for the difference in form (Stilo's distinction into clitics and affixes).

As is the case in many languages, the placement of sentence stress in Vafsi is dependend on information structural constraints, but in unmarked sentences, the

[3]For Vafsi, an Accentual Phrase might be more fitting as in Sadat-Tehrani (2007) for Persian, but this is left for further research.

main pitch accent typically falls on the element directly preceding the verb (which consequently is a suitable host for a prosodically deficient enclitic).

In subordinate clauses, on the other hand, the initial subordinate conjunction usually receives the main pitch accent. Note, however, that this does not hold for the subordinate conjunction *ke*, which Stilo (2004b, 21) assumes to be a subordinating particle without any semantic content. Under the assumption that each subordinating phrase corresponds to an intontational (or at least a phonological) phrase, an oblique clitic occuring directly after the unstressed particle *ke* is thus stranded without an adequate host $_{\iota}$(ke =clitic ...)$_{\iota}$ and requires a prosodic repair mechanism, e.g., prosodic inversion. In the case of example (13), prosodic inversion then places the clitic after the first suitable host to its right, i.e., the stressed preverb *hár*.

A particular prosodic pattern in Vafsi is the so-called *sustained intonation* where the pitch level remains high and flat, with a longer duration on the sustained element and a brief, but perceptible pause following (Stilo, 2004b, 274). In terms of prosodic phonology, sustained intonation can be associated with a phonological phrase boundary (Selkirk, 1978; Nespor and Vogel, 1986; Frota, 2012). In Vafsi, such a boundary seems to occur

- after constructions connected by the coordinating conjunctions *-o* ('and') and *ya* ('or'), thus explaining examples like (7) and (9), and
- often after the subject of a sentence in Vafsi (see also Sadat-Tehrani (2007) for Persian[4]), which explains examples like (11) where the clitic cannot directly follow the subject.

Furthermore, if the prosodically deficient enclitic is placed in the initial position of a phonological/intontaional phrase as it is the case with example (8b), and cannot be moved via prosodic inversion (the only suitable position would be the postverbal position), the clitic is forced to assume its stressed full form. The question why the clitic cannot be placed after the durative marker *ær*, but certainly after the punctual marker, the negative marker, and the preverbs will become clear in the following section.

In conclusion, by assuming the oblique enclitic to be sensitive to prosodic phrasing, we can already explain a large proportion of the seemingly irregular examples: If the clitic is stranded at the beginning of a prosodic phrase without an adequately stressed host to its left, the clitic has to undergo one of two possible prosodic repair mechanisms: a) prosodic inversion, or, if this is not possible, b) assume a clitic-under-stress form.

However, not all forms can be explained with reference to prosodic boundaries. Table 3 gives an overview on the seemingly critical examples (7)-(13) with respect to prosodic boundaries (round brackets) and the applied repair mechanism; open questions are indicated by ?.

[4]In fact, adding a comma after the subject is common practice in written Persian.

pros. constraint	**example**	***prosodic inversion***	***clitic under stress***
sustained: coord	(7)	(komæk**=i** kær-òm)	
initial in IP	(8)	–	(**ìm**-ær-góæ)
sustained: coord	(9)	–	(**ì-**r-koš-ome)
initial in IP	(10b)	(b-**ím**-diæ)	?
sustained: subject	(11)	(v-**ís**-vattæ)	?
?	(12)	... hár**=es**-da	–
initial in CP	(13)	(hár**=esan**-kærdæ)	–

Table 3: An overview on examples (7)-(13) with respect to prosodic boundaries

The question why the stressed form of the clitic appears verb-medially in examples (10b) and (11), but not in (12) and (13), and why (12) requires the clitic to occur within the verbal complex at all will be discussed in the following two sections.

3.2 The verbal complex and its members

There are a number of particles in the verbal complex which shed light on the distribution patterns of the clitic and which we will therefore briefly discuss in the following section. These are:

1. The durative marker *æt*
2. The punctual marker *bć*
3. The negation marker *nć*
4. The preverbs *dć(r)-, ó(r)-*, and *há(r)-*

The **durative marker *æt-*** occurs in the present and the imperfect and is placed before the main verb ((14a), repeated from (8)). Its surface form may change depending on the phonological environment[5]. The durative marker is unstressed and is therefore unsuitable to function as a host for stranded oblique clitics ((14b)).

(14) a. an**=om** [ær-góæ]$_{VC}$
that=1S.OBL DUR-want
'I want that'

b. [**ìm**-ær-góæ]$_{VC}$
[1S.OBL-DUR-want]$_{VC}$
'I want' (Stilo, 2010, 247)

As the clitic is prohibited from moving to the postverbal position, the only remaining option is to assume a stressed form, which is also clearly visible in the following speech signal (depicting *na=san æd-dir-am* ('We (will) keep/hold them') on the left, and *isan-ær-vend-am* ('We (will) find them') on the right.)

[5]As the different phonological surface forms of the durative, but also the punctual marker, are irrelevant for the current discussion, the interested reader is referred to Stilo (2004b).

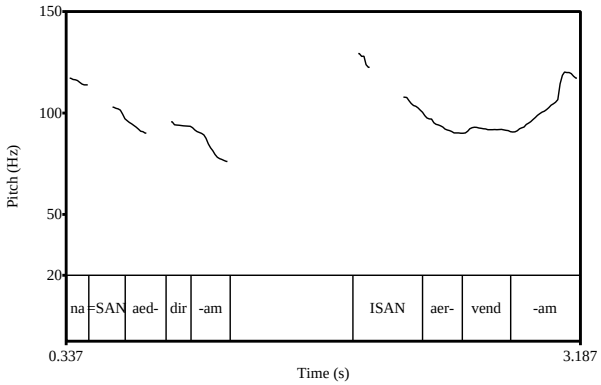


Figure 1: Pitch contrast between the clitic (left) and the stressed full form (right)

On the left side in Figure 1, the clitic precedes the durative marker and follows the element carrying stress (*na*). Both the clitic and the durative marker are clearly unstressed. On the right side, the clitic is left stranded in the initial position and thus assumes the clitic-under-stress form *isan*, which is clearly visible in the speech signal itself. The durative marker following *isan* is again unstressed.

The **punctual marker *bǽ-*** is used in the presesent subjunctive, the simple past, and all perfect tenses. Like the durative marker, the punctual marker is also placed before the verb ((15), repeated from (10a)).

(15) án=**om** [bǽ-diæ]$_{VC}$
that=1S.OBL PUNCT-saw
'I saw that' (Stilo, 2010, 247)

However, in contrast to the durative marker, the punctual marker carries stress. Furthermore, if the element following the punctual marker starts with a vowel, the punctual marker's *æ*-vowel is dropped and stress shifts to the adjacent vowel:

(16) *bǽ-* (PUNCT) + *-av* ('come') + *-e* (3S.DIR) → *báwe* ('s/he came')
(Stilo, 2004b, 15)

Since the punctual marker is stressed it can function as a suitable host for an oblique clitic stranded in the initial position, which consequently undergoes prosodic inversion and is placed in the position following the punctual marker ((17a)). However, as discussed above, if the punctual marker is followed by a vowel, the punctual marker's vowel is dropped and stress is shifted to the vowel of the following element - in that case the clitic ((17b,c)), which is then again forced to assume its clitic-under-stress form, also visible in the speech signal (Figure 2).

(17) a. *prosodic inversion*: =om bǽ-diæ → bǽ=om-diæ
b. *stress shift*: bǽ=om-diæ → b=óm-diæ → b-ím-diæ
c. [b-**ím**-diæ]$_{VC}$
PUNCT-1S.OBL-saw
'I saw'

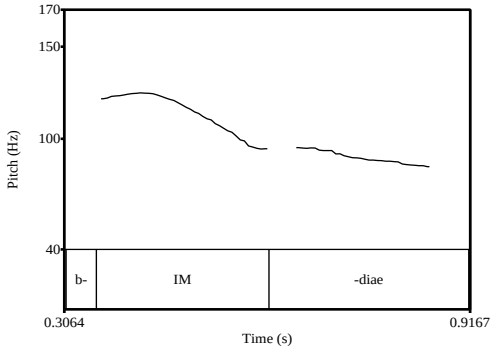


Figure 2: Speech signal depicting *b-ím-diæ* ('I saw')

The **negative marker *nǽ-*** behaves like the punctual marker *bǽ-*, in that it occurs preverbally and carries stress. Furthermore, if the following element has an initial vowel, *æ* is dropped and stress shifts to the vowel of the following element. However, if the negative and the punctual marker co-occur, the punctual marker is suppressed (*bǽ-ssim* 'I went', but *nǽ-ssim* 'I didn't go' (Stilo, 2004b, 233)). In contrast, the negative particle can co-occur with the durative marker ((18)).

(18) an ræféq-i=s [nǽ-r-vaz-e]$_{VC}$
he.DIR friend-OM=3S.POSS NEG-DUR-say-3S.DIR
'He doesn't tell his friend.' (Stilo, 2010, 259)

In cases where the negative marker co-occurs with a clitic, the process in (17) is applied. If the clitic is stranded in the initial position, it is first placed following the stressed negative marker via prosodic inversion, before assuming its full form due to the stress shift from the negative marker.

(19) [n-**ím**-ær-vaz-i?]$_{VC}$
NEG-1S.OBL-DUR-say-2S.DIR
'Won't you tell me?' (Stilo, 2010, 266)

As their name suggests, the **preverbs *dǽ(r)-, ó(r)-*** and ***há(r)-*** occur preverbally. Originally, the preverbs were directional particles; they add to the meaning of the verb from finer nuances to complete meaning changes with respect to the main verb. As the following table shows, it is almost impossible to associate each preverb with a particular meaning.

Vafsi	**English**	**Vafsi**	**English**
girætt	grab, catch	biri	interrupt, cut off
ó(r)-girætt	pick up, lift	ó(r)-biri	cut out (with scissors), curdle
há(r)-girætt	take, get, bury	há(r)-biri	shear (fleece), cut of (sheep's) head
dǽ(r)-girætt	gather up	dǽ(r)-biri	cut (general)

Table 4: The preverbs *dǽ(r)-, ó(r)-* and *há(r)-* (see Stilo, 2004b, 233)

Not all verbs have preverbs (e.g., *vin/di* 'see'), some have only a subset of preverbs, and some do not occur without any preverbs (e.g., *hár-eysi* 'make dough').[6] While the preverbs suppress the duration and the punctual marker, they are themselves suppressed by the negative marker. In this case, the exact meaning of the negated verb is not discernible and has to be determined by context.

If the preverbs co-occur with a clitic, the clitic is usually placed preceding the preverb (and the verbal complex).

(20) tæmen ketab=**es** [há-baxăa]$_{VC}$
1S.OBL book=3S.OBL PVB-gave.away
'He gave a book away to me.' (Stilo, 2010, 253)

The preverbs are all stressed, i.e., they are suitable hosts for a clitic undergoing prosodic inversion. Crucially, however, they do not follow the pattern of stress shift that occurs with the punctual and the negative marker. As a consequence, the clitic retains the clitic form and stress remains with the preverb ((21), also (13)).

(21) [hár=**om**-da]$_{VC}$ yey kelj-i
PVB=1S.OBL=gave one girl-OF
'I gave (it) to some girl.' (Stilo, 2010, 252)

From the above discussion on the unstressed durative marker, the stressed punctual marker, and the stressed negative marker, it becomes clear why the clitics assume their full form in (9), (8), (10b), and (11): with the durative marker, they have to remain in situ in the initial position of a prosodic phrase, because an adequate host is not available; with the punctual and the negative marker, prosodic inversion is possible, but the stress shift from the markers to the clitic again forces the clitic to assume its full form. Finally, with the preverbs, the clitics undergo prosodic inversion, but are not subjected to a stress shift and can thus retain their clitic form.

However, all of these cannot explain why the clitic is inverted in example (12). This question is resolved in the following section.

3.3 The possessive construction in Vafsi

In addition to other possibilities (e.g., *ezafe* (Stilo, 2004b)), the possessive in Vafsi can be expressed by the same set of clitics as the oblique pronouns. Important to note is that the possessive marker (underlined) directly follows the possessed item and is thus (in contrast to the pronoun clitic) not restricted to the preverbal position.

(22) ... šus=<u>s</u> sær=**esan** [há-biri]$_{VC}$
husband=3S.POSS head=3P.OBL PVB-cut
'... they cut off her husband's head.' (cf. Stilo, 2010, 290)

[6]This discrepancy in distribution as well as the variety in meaning suggests that the respective preverbs are closely attached to the verb to the extent that they form a single lexical entry with it.

The possessive clitic can also occur in the preverbal position where the ambiguity in form can lead to an ambiguity in understanding.

(23) a. kænizan=es báwattæ ‘*Her* servant girls said (so)’ → as possessive

b. kænizan**=es** báwattæ ‘*She* told the servant girls’ → as subject

(Stilo, 2004b, 13)

While the possessive indicator and the oblique clitic can certainly co-occur in one sentence ((22)), they never co-occur in the same position; i.e., the oblique clitic cannot attach directly to the possessive marker. In these cases, the clitic is moved via prosodic inversion which explains example (12) discussed above: The clitic moves to the position following the stressed preverb *hár*, retaining its unstressed form, because it cannot take the possessive clitic as its host. In the following example, on the other hand, the position targeted by prosodic inversion is directly after the punctual marker, which, according to the stress-shifting rules discussed above, drops its vowel and shifts the stress to the following vowel of the clitic. The clitic, now stressed, assumes its full form *ís*.

(24) bár=es [v=**ís**-værd]$_{VC}$
load=3S.POSS PUNCT=3S.OBL-took
‘He took his load.’ (Stilo, 2004b, 239)

3.4 Intermediate conclusion

As established in the previous sections, the oblique clitics have a fairly regular placement pattern preceding the verbal complex. The cases where this pattern is interrupted can be explained via prosodic means: If the prosodically deficient enclitic does not have a suitable host to its left, this violation of prosodic constraints is repaired by a) prosodic inversion, or (if this is not possible), by b) the clitic assuming its stressed full form.

pros. constraint	**example**	***prosodic inversion***	***clitic under stress***
sustained: coord	(7)	(komæk**=i** kær-òm)	
initial in IP	(8)	–	(**ìm**-ær-góæ)
sustained: coord	(9)	–	(**ì-**r-koš-ome)
initial in IP	(10b)	*(bǽ=**om**-diæ)	(b-**ím**-diæ)
sustained: subject	(11)	*(vǽ=**es**-vattæ)	(v-**ís**-vattæ)
possessive	(12)	=*poss* hár**=es**-da	–
initial in CP	(13)	(hár=**esan**-kærdæ)	–

Table 5: An overview on the respective placement constraints of examples (7)-(13)

As can be see in the table above, all of the seemingly confusing cases can be explained accordingly. It can therefore be concluded that there is no need to assume

an 'affixal form' for the oblique clitics, which would require a unified explanation as to why some structures prefer the affixal form and others do not; one would, in fact, expect the affixal form to be present at all times, as it would be considered to be part of the morphological form. With the approach presented above, on the other hand, all instances can be explained with reference to the interface between syntax and prosody, and postlexical phonology.

4 Vafsi oblique pronouns at the syntax-prosody interface

In the following section we will analyse the findings discussed above at the syntax-prosody interface as proposed in Bögel (2015), which allows for a straightforward communication at the interface itself and can furthermore account for postlexical phonological processes like the stress shift and prosodic inversion as well. The model is based on the assumption that there is a fundamental difference between two perspectives on grammar: *production* and *comprehension*.[7] *Production* refers to the process from meaning to form, i.e., from the composition of meaning in the mind to the final articulation. *Comprehension* on the other hand refers to the processing of speech/text into meaning, i.e., to the transfer from form to meaning. In between these two vanishing points are the different modules of grammar, and, depending on the process, the arrangement has a certain directionality. The following figure illustrates (adapted from Jackendoff, but see other models of speech production/comprehension, e.g., Levelt (1999)).[8]
In LFG, the syntax → prosody interface exchanges information from c-structure to p-structure and the prosody → syntax interface refers to the information transfer from p-structure to c-structure. In the following, the exemplary Vafsi oblique pronouns in (25) are discussed at the syntax–prosody interface (i.e., during production).

[7]Taking a directional perspective is crucial at the interface between the syntactic and the prosodic module. While the syntactic module provides the underlying syntactic constituency as input to prosodic constituency during production, prosodic phrasing usually does not determine syntactic phrasing. Frequent mismatches between syntactic and prosodic phrasing are a further indication that treating the interface equally in both directions is not efficient; instead, a selective 'checking' in the few cases where prosodic phrasing is relevant to syntactic phrasing is assumed. A non-directional analysis of an interface phenomenon runs the risk of providing an analysis that can never be processed in reality, whether by humans or machines, because the phenomenon itself is dependent on a pipeline view (as in the present paper). The approach presented in this paper therefore assumes a more directional view where grammar is seen as a 'map' for a particular act of language, either production or comprehension (see also Figure 3).

[8]The abstract representation and the arrangement of modules is what is generally dubbed as 'competence' as opposed to 'performance', which includes factors beyond the grammatical rules of a language (where the boundary is sometimes difficult to determine). However, it is also clear that any model of performance must be able to "incorporate the system of grammatical rules" (Chomsky, 2006, 104). The model presented here understands the arrangement of modules (including the directionality) and the associated grammatical rules as part of competence, but furthermore assumes that these modules greatly overlap and allow for backtracking during performance.

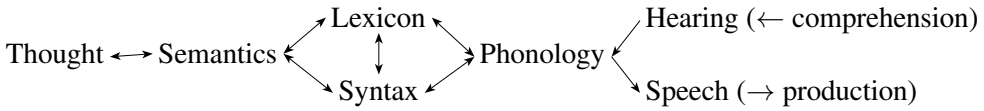


Figure 3: The language processor (cf. Jackendoff, 2002, 197, modified).

(25) a. án=**om** [bǽ-diæ]$_{VC}$
that=1S.OBL PUNCT-saw
'I saw that'

b. [b-**ím**-diæ]$_{VC}$
PUNCT-1S.OBL-saw
'I saw' (Stilo, 2010, 247)

There is, as of yet, not enough background information to determine the c-structure of Vafsi (although Belyaev and Haug (2018)'s analysis of the VP in Ossetic is a good starting point for future work on the syntactic side of Vafsi). We will therefore restrict ourselves to a flat c-structure, as it is the linear order which is relevant for the present analysis.

(26) S ⟶ XP* CL VC

For example (25a), the analysis is straightforward: The clitic prosodically attaches to an adequately stressed host to its left (NP =CL VC). In example (25b), on the other hand, the syntactic analysis does not provide a stressed element to the left of the clitic(_ =CL VC); the clitic is left stranded in the first position.[9] The correct analysis of cases like these cannot be reduced to syntax alone, but requires reference to the syntax-prosody interface and to postlexical phonology. In the following, we will analyse example (25b) at the interface as introduced in Bögel (2015).

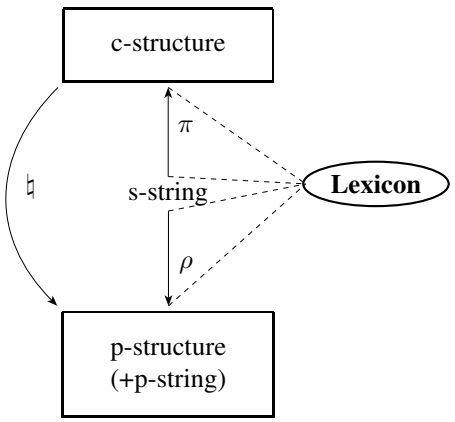


Figure 4: The syntax–prosody interface (Bögel, 2015)

Two transfer processes are assumed at the interface:

[9]Which, from a syntactic perspective, is fully justifiable, as syntax is not responsible for the creation of prosodically accurate structures.

1. The *transfer of structure* (⫮) exchanges information on larger syntactic and prosodic constituents (IP/CP/XP; ι/φ).

2. The *transfer of vocabulary* (π/ρ) associates morphosyntactic and phonological information on lexical elements and projects them to their respective (c- and p-)structures

Before describing the two transfer processes at the interface in more detail, the following section briefly introduces the representation of p-structure.

4.1 P-structure – the p-diagram (during production)

P-structure is represented via the p-diagram which allows a syllablewise linear representation of the utterrance in question. Each syllable is part of a vector (V.INDEX) which associates the syllable with relevant segmental and suprasegmental phonological information. Input to the p-diagram comes from c-structure (*transfer of structure*) and the lexicon (*transfer of vocabulary*).

PHRASING	$(_\iota = \sigma$	$(_\omega \sigma$	σ	$\sigma)_\omega)_\iota$
...	...	...	...	...
LEX.STRESS	–	prim	–	–
SEGMENTS	/om/	/bæ/	/di/	/æ/
V. INDEX	$\mathbf{S_1}$	$\mathbf{S_2}$	$\mathbf{S_3}$	$\mathbf{S_4}$

Figure 5: Syntactic input to p-structure: the initial p-diagram for example (25b)

Figure 5 represents the initial input to p-structure from the two transfer processes at the interface, which will be introduced in more detail in the following sections.

4.2 The Transfer of Vocabulary

The *transfer of vocabulary* operates at the word-level and below. It relates each element of the string to its associated morphosyntactic and phonological information in the lexicon. Each lexical entry[10] is associated with (at least) three dimensions: its semantic concept (irrelevant in the present discussion), its s(yntactic)-form which includes all the relevant morphosyntactic information, and its p(honological)-form, which includes information on segments, metrical structure, or lexical stress. The following table shows the s-form and the p-form for *bǽ-diæ* and *om*.

[10] 'Lexical entry' here is equal to the 'surface form', i.e., the morphophonologically complete word form. This does not exclude a dynamic generation of the full form *within* the lexical component.

s(yntactic)-form			**p(honological)-form**	
bǽ-diæ V	(↑ PRED)	= 'diæ⟨SUBJ⟩'	P-FORM	[bǽdiæ]
	(↑ TENSE)	= past	SEGMENTS	/b æ d i æ/
	(↑ ASPECT)	= punctual	METR. FRAME	$(\text{'}\sigma\sigma\sigma)_\omega$
	...			
om PRON	(↑ PRED)	= 'pro'	P-FORM	[om]
	(↑ PERS)	= 1	SEGMENTS	/o m/
	(↑ NUM)	= sg	METR. FRAME	$=\sigma$
	(↑ CL-TYPE)	= set2		
	...			

Table 6: Lexical entries for *om* 'I' and *bǽ-diæ* 'saw'

While the s-form represents a typical lexical entry in LFG, the p-form adds information with respect to the phonological nature of the lexical entry, in particular it encodes the number of syllables (σ), lexical stress ('), and the metrical frame (a prosodic word $()_\omega$, or a clitic ($=\sigma$ for enclitic, $\sigma=$ for proclitic)).

Each dimension of the lexicon can only be accessed by the related module (c-structure can access the s-form, p-structure p-form), which ensures modularity. However, once a dimension is accessed, the associated dimensions become available as well and information can be transferred from one module to another. The *transfer of vocabulary* works both ways (and hence has a translatory function): During comprehension, p-structure accesses the related p-form, which in turn activates the associated s-form making it available to c-structure; during production, the process is reversed from c-structure to p-structure where the information associated with the p-form is syllablewise encoded in the p-diagram.

p(honological)-form	
P-FORM	[bǽdiæ]
SEGMENTS	/b æ d i æ/
METR. FRAME	$(\text{'}\sigma\sigma\sigma)_\omega$
P-FORM	[om]
SEGMENTS	/o m/
METR. FRAME	$=\sigma$

↓ ↓ ↓

PHRASING	$=\sigma$	σ	σ	σ
LEX.STRESS	–	prim	–	–
SEGMENTS	/om/	/bæ/	/di/	/æ/
V. INDEX	$\mathbf{S_1}$	$\mathbf{S_2}$	$\mathbf{S_3}$	$\mathbf{S_4}$

Figure 6: Encoding the p-forms of example (25b) in p-structure

In addition to the transfer process at the word-level and below, the modules also

need to exchange information on prosodic and syntactic constituency, and intonation. This is accomplished via the *transfer of structure*.

4.3 The Transfer of Structure: from syntax to prosody

The *transfer of structure* directly associates c- and p-structure via the projection function ♮ (Figure 4) and exchanges information on syntactic and (higher) prosodic constituency. The assumptions made here roughly follow Selkirk (2011)'s MATCH THEORY for the higher constituents,[11] in that each IP/CP (here: S) matches an intonational phrase (ι) and each XP corresponds to a phonological phrase (φ). During production[12], the syntactic S-node will thus have the following annotation

S
(♮($T(*)$) S_{min} PHRASING) = $_\iota$(
(♮($T(*)$) S_{max} PHRASING) = $)_\iota$

which can be read as "*For all terminal nodes (T) that are daughters of the current node (*=S), take the first (S_{min}) and the last syllable (S_{max}) and for the attribute* PHRASING *add a left and a right intonational phrase boundary ($()_\iota$) at these positions.*"

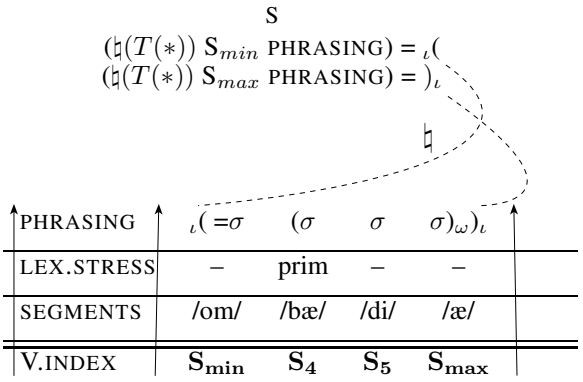


Figure 7: The *transfer of structure* for example (25b)

The p-diagram in Figure 7 depicts the result of the two transfer processes at the interface: the *transfer of structure* and the *transfer of vocabulary*. However, as noted before, the linear order predicted in syntax does not reflect the actual linear order of examples like (25b). The determination of the final linear order is based on prosodic constraints and is therefore the domain of p-structure, i.e., of (language-specific) postlexical phonology.

[11] Note, however, that the model presented here is not limited to this approach, but can easily be adjusted to fit other approaches to the interface as well.

[12] See, e.g., Butt et al. (2017) for an analysis (of Urdu polar *kya*) during comprehension.

4.4 Postlexical phonological processes

Within p-structure, the initial input to p-structure created by the two transfer processes at the interface is scrutinized in phonological/prosodic terms. As the oblique enclitic is placed in the initial position of an intonational phrase, prosodic inversion applies and places the clitic after *bǽ*. However, since the clitic begins with a vowel, stress shifts from *bǽ* to the clitic, which consequently has to assume its full form.[13]

input (=s-string):	=om bǽdiæ
prosodic inversion:	bǽ=om=diæ
stress shift:	b-ím=diæ
output (=p-string):	b-ím=diæ

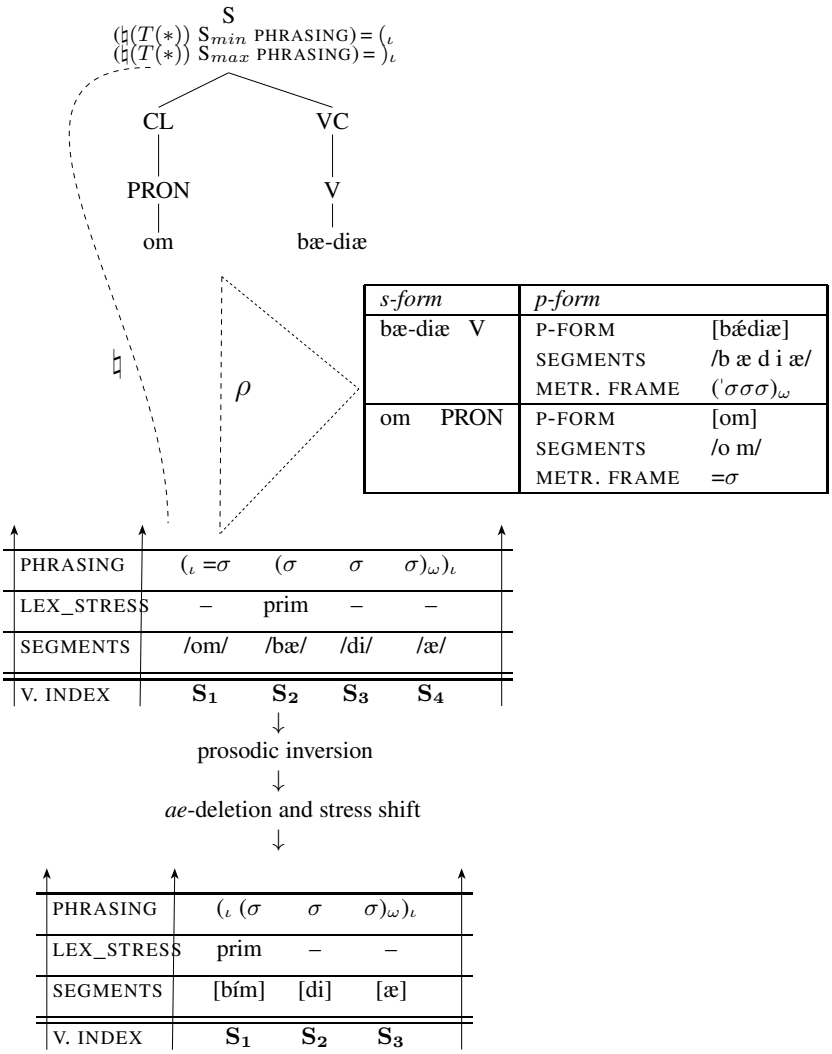
Table 7: Postlexical phonological processes in p-structure

While the input and output of p-structure are mostly congruent with respect to linear order, the approach presented here can a) account for the commonly found mismatches between syntactic and prosodic constituency in general, and b) can explain differences in the syntactic and phonological linear order, and even apparent violations of lexical integrity (if and only if the placement of an otherwise independent morphosyntactic item within another is prosodically motivated).

4.5 Vafsi oblique pronouns at the syntax-prosody interface - overview

The following figure gives a complete overview on the analysis of example (25b) from the perspective of production.

[13] While the postlexical phonological rules applied here affect the 'lexical level' of the p-diagram, they cannot affect the 'interpretation level' (to which PHRASING belongs, see Bögel (2015) for details on the different levels of the p-diagram), as this would also misplace the left intonational boundary associated with the clitic's position (thanks to Jamie Findlay for noticing this).



5 Conclusion

In this paper we showed that Vafsi oblique pronouns are clitics without an 'affixal counterpart' and that these clitics are not 'fronted' from (inside) the verb as claimed by previous research (Stilo, 2004a,b, 2010). We showed that the clitics originate at the position immediately preceding the verbal complex where they prosodically attach to a preceding host. However, if such a host is not available (either because the clitic is stranded in the sentence-initial position or is preceded by a sentence-medial prosodic boundary), the clitic can undergo prosodic inversion in which case it is placed following a suitable host to its right.

If prosodic inversion is impossible, the clitic has to remain in situ, assuming a stressed full form (the former 'affixal form') to account for its initial position within

a larger prosodic domain. The same stressed full form is assumed if the clitic undergoes prosodic inversion, but is then targeted by a postlexical phonological stress shift.

This interplay of the syntactic and prosodic module and of postlexical phonology can be modelled straightforwardly at the syntax-prosody interface as proposed in Bögel (2015). This approach not only allows for a unified analysis of Vafsi oblique pronouns as clitics, but can furthermore explain mismatches between syntactic and phonological linear order by means of prosodic constraints and postlexical phonology.

References

Belyaev, Oleg and Haug, Dag. 2018. Pronominal coreference in Ossetic correlatives and the syntax-semantics interface. In *Proceedings of LFG18*, Stanford, CA: CSLI Publications.

Bögel, Tina. 2015. *The Syntax–Prosody Interface in Lexical Functional Grammar*. Ph. D. thesis, University of Konstanz.

Butt, Miriam, Bögel, Tina and Jabeen, Farhat. 2017. Polar kya and the prosody-syntax-pragmatics interface. In *Proceedings of LFG17*, Stanford, CA: CSLI Publications.

Chomsky, Noam. 2006. *Language and Mind*. New York: Cambridge University Press, third edition.

Frota, Sónia. 2012. Prosodic structure, constituents, and their implementation. In A. C. Cohn, C. Fougeron and M. K. Huffman (eds.), *The Oxford Handbook of Laboratory Phonology*, Chapter 11, pages 255–265, Oxford University Press.

Halpern, Aaron L. 1995. *On the Placement and Morphology of Clitics*. Stanford, CA: CSLI Publications.

Jackendoff, Ray. 2002. *Foundations of Language*. New York: Oxford University Press.

Ladd, D. Robert. 1986. Intonational phrasing: the case for recursive prosodic structure. *Phonology Yearbook* 3, 311–340.

Levelt, Willem J.M. 1999. Models of word production. *Trends in Cognitive Sciences* 3(6), 223–232.

McCawley, James D. 1968. *The Phonological Component of a Grammar of Japanese*. The Hague [a.o.]: Mouton.

Mirdehghan, Mahinnaz and Yousefi, Saeed Reza. 2016. Dative case marking in Vafsi within the OT framework. *Iranian Studies* 50(1), 149–161.

Nespor, Marina and Vogel, Irene. 1986. *Prosodic Phonology*. Dordrecht: Foris.

Sadat-Tehrani, Nima. 2007. *The Intonational Grammar of Persian*. Ph. D. thesis, University of Manitoba, Winnipeg.

Selkirk, Elisabeth O. 1978. On prosodic structure and its relation to syntactic structure. In T. Fretheim (ed.), *Nordic Prosody II*, pages 111–140, Tapir.

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.

Stilo, Donald. 2010. Ditransitive constructions in Vafsi: a corpus-based study. In Andrej Malchukov, Martin Haspelmath and Bernard Comrie (eds.), *Studies in Ditransitive Constructions: A Comparative Handbook*, pages 243–276, Berlin: De Gruyter Mouton.

Stilo, Donald L. 2004a. Coordination in three Western Iranian languages. In Martin Haspelmath (ed.), *Coordinating Constructions*, Chapter 11, Amsterdam: John Benjamins.

Stilo, Donald L. 2004b. *Vafsi Folk Tales: Twenty Four Tales in the Gurchani Dialect of Vafsi as Narrated by Ghazanfar Mahmudi and Mashdi Mahdi and Collected by Lawrence P. Elwell-Sutton*. Wiesbaden: Dr. Ludwig Reichert Verlag.

Interdependencies in Chinese Noun Phrases

Kersti Börjars
The University of Manchester

Christopher Hicks
The University of Manchester

John Payne
The University of Manchester

Proceedings of the LFG'18 Conference

University of Vienna

Miriam Butt, Tracy Holloway King (Editors)

2018

CSLI Publications

pages 109–128

http://csli-publications.stanford.edu/LFG/2018

Keywords: Classifiers, Noun phrase structure, Chinese

Börjars, Kersti, Hicks, Christopher, & Payne, John. 2018. Interdependencies in Chinese Noun Phrases. In Butt, Miriam, & King, Tracy Holloway (Eds.), *Proceedings of the LFG'18 Conference, University of Vienna*, 109–128. Stanford, CA: CSLI Publications.

Abstract

In this paper, we propose a novel LFG analysis of the structure of Chinese noun phrases involving quantifiers and classifiers or measure words. The analysis accounts for the interdependencies between noun-phrase internal categories and the types of modifier they license by postulating a c-structure involving a spine of co-heads (D - Q - Class - N). This structure is more complex than the c-structure typically assumed for noun phrases in a variety of languages within LFG, but motivated specifically for Chinese both by the rigid ordering restrictions between these elements and the different categories of modifier permitted at each level. We argue, however, that the mutual interdependence of quantifiers and classifiers, and the (partial) complementary distribution between different types of classifier is a consequence of the f-structure features assigned to these. The analysis therefore exploits to the full the LFG distinction between a syntactically motivated c-structure and an independent level of f-structure.

1 Introduction

LFG generally takes a restrictive approach to functional categories in assuming that they are only warranted when a particular functional feature is associated with a structural position, such as for instance the finiteness of verb-second languages (see for instance (Kroeger, 1993) and Börjars et al., 1999). Based on these assumptions, D tends to be the only functional category used in noun phrases. However, in this paper, we will argue that the interdependencies between quantifiers and classifiers in Mandarin must be accounted for structurally through a spine of functional categories D – Q – Class – N in the noun phrase.

2 Classifiers and measure words

In Mandarin Chinese, a noun cannot combine directly with a numeral, other quantifier or demonstrative, but the noun must first combine with some element, as illustrated in (1) to (3) (grammatical examples from Her & Hsieh, 2010, 528).

(1) *yi shu
one book

(2) yi ben shu
one CL book
'one book'

(3) yi xiang shu
one MW_{BOX} book
'one box of books'

[†]We are grateful for comments made by two anonymous reviewers and the LFG2018 audience.

We follow Tai & Wang (1990, 38) and many others in making a distinction between CLASSIFIER (2) and MEASURE WORD (3) (other terms for the same distinction are Sortal classifier vs Mensural classifier (Lyons, 1977, 463) and Classifier vs Massifier (Cheng & Sybesma, 1998)). Measure words themselves fall into a number of subtypes. In addition to the container subtype illustrated in (3), there are also at least standard measures (e.g. *gongjin* 'kilo'), collections (e.g. *pian* 'group') and kinds (e.g. *lei* 'type'). For a full heuristic classification, see Li (2013). In this paper, we will limit our discussion to the container subtype, leaving an analysis of other subtypes for future work.

These elements do not occur except when there is a numeral, quantifier or demonstrative, as the ungrammaticality of the examples in (4) and (5) illustrate. Hence there is a mutual dependency.

(4) *ben shu
CL book

(5) *xiang shu
MW_{BOX} book
'one box of books'

Following Zhang (2013), we will use UNIT WORDS to refer collectively to classifiers and measure words. Measure words generally denote a quantity of the entity named by a noun and exist in all languages in some form. As noted above, though, when we refer to measure words in this paper we intend the discussion to apply specifically to the container subtype. Classifiers are elements which categorise a class of nouns by picking out some key property associated with entities named by the class of nouns. Classifiers uniquely set apart a number of Southeast Asian languages, indigenous languages of western Americas, and Sub-Saharan African languages (Nichols, 1992, 200).

The use of elements with a classifier function in Chinese dates to over 3,300 years ago (see for instance Erbaugh, 1986; Peyraube, 1991; Wang, 1994). The estimates of the number of classifiers vary greatly, partly because the distinction between measure words and classifiers is not always made, or not made along the same lines (for different estimates, see for instance Erbaugh, 1986; Hu, 1993; McEnery & Xiao, 2010). Prescriptively, there is one "correct" classifier for most nouns, for instance *zhi* with animals or *zhang* for flat things, or more specialised ones such as *pi* for horses and *ben* for books. However, usage varies greatly and *ge*, which can be used with all countable nouns, is used with increasing frequency; in a corpus study McEnery & Xiao (2010, 50) show that *ge* accounts for 38.8% of all unit word tokens in the texts they examined.

3 Interaction between classifiers and measure words

As pointed out by Her & Hsieh (2010), classifiers and measure words seem at first sight to be mutually exclusive and hence to occupy the same slot:

(6) yi ben shu
one CL book
'one book'

(7) yi xiang shu
one MW_{BOX} book
'one box of books'

(8) *yi xiang ben shu
one MW_{BOX} CL book

(9) *yi ben xiang shu
one CL MW_{BOX} book

However, a numeral immediately preceding a unit word affects the two types differently; in the case of a classifier, it counts units of the main noun itself, whereas a pre-measure word numeral independently counts units of the measure word. Hence the two can co-occur in contexts where both the noun and the measure word are counted, as in (10) (from Her & Hsieh, 2010, 536).

(10) yi xiang shi ke pingguo
one MW_{BOX} ten CL apple
'one box of ten apples'

Measure words can be stacked, as illustrated in (11), but classifiers cannot (12). If there is a classifier, there can only be one and it must be the lowest unit word; compare (10) and (13) (from Her & Hsieh, 2010, 536)

(11) yi xiang shi bao pingguo
one MW_{BOX} ten MW_{PACK} apple
'one box of ten packs of apples'

(12) *yi ge shi ke pingguo
one CL ten CL apple

(13) *yi ge shi bao pingguo
one CL ten MW_{PACK} apple

3.1 Attributive modification

In general noun-phrase internal modifiers must be marked by *de* in Chinese. In (14) we see a *de*-marked relative clause and in (15) a possessive pronoun (Li, 2013, 62).

(14) wo mai de shu
I buy DE book
'the book(s) that I bought'

(15) ta de shu
he/she DE book
'his/her book(s)'

In (16) we see *de* used with a disyllabic adjective, with co-ordinated adjectives in (17) and with a modified adjective in (18) (Zhang, 2012, 127). These are generally argued to be the adjective types that obligatorily require *de*. However, Paul (2010, 121–122) shows with reference also to earlier literature that this generalisation is not accurate; there are some more complex adjectives and adjective phrases that can also occur without *de*.

(16) congming de haizi
clever DE children
'clever children'

(17) chang erqie cu de xianglian
long and thick DE necklace
'long and thick necklace'

(18) hen chang de xianglian
very long DE necklace
'very long necklace'

Structurally, adjectives that can occur attributively without *de* can also occur with it (though see Paul (2010) for semantic implications that render some combinations infelicitous). The distinction between *de* modification and *de*-less modification will be relevant to us, but the exact membership of each type will not since only a small number of *de*-less adjectives are of relevance to us.

When a noun is preceded by a classifier, the standard position for attributive adjectives is immediately preceding the noun, whether it is *de* or *de*-less modification as in (19) and (20).

(19) yi ke da pingguo
one CL big apple
'one big apple'

(20) yi ke hen da de pingguo
one CL very big DE apple
'one very big apple'

Classifiers cannot be preceded by any kind of *de* modifier, but a restricted set of simple "dimensional" adjectives are acceptable in pre-classifier position. However, semantically, pre-classifier modifiers still modify the main noun as shown in (21) (Li, 2013).

(21) yi da ke pingguo = yi ke da pingguo
one big CL apple one CL big apple
'one big apple' 'one big apple'

Measure words behave in a similar way to classifiers structurally in that only the same small set of dimensional adjectives can precede them. However, they differ from classifiers in that the preceding adjective semantically modifies that measure

word, as it would a noun (Her & Hsieh, 2010, 537).

(22) yi da xiang pingguo ≠ yi xiang da pingguo
one big MW_{BOX} apple one MW_{BOX} big apple
'one big box of apples' 'one box of big apples'

Adjectival modifiers marked by *de*, but not simple adjectives, are also permitted in pre-quantifier position (see Li, 2013, 174). This is referred to by Zhang (2012) as the left-peripheral position. It is illustrated in (23).

(23) hen da de yi ge xiguo
very big DE one CL watermelon
'one very big watermelon'

There appears to be little, if any, semantic difference between the adnominal and left-peripheral positions.

Modifiers marked by *de* are generally assumed to be phrasal (see for instance Fan, 1958; Huang, 1989; Tang, 1990), but there is some argument about the status of *de*-less modification. Sproat & Shih (1988, 1991) and others argue that the adjective forms a compound with the noun, but Paul (2010) argues against this position, and describes the combination as phrasal. We do not take a view on *de*-less modification in general, but we will assume that the small set of adjectives that can precede unit words are non-projecting adjectives.

4 Previous analyses

Previous analyses of classifiers and measure words outside LFG are typically torn between two difficult-to-reconcile requirements. Firstly, in order to account for the fact that classifiers and measure words are mutually exclusive in basic noun phrases consisting of a numeral and a noun, it is necessary to assume that they occupy the same slot, i.e. that they form a unitary formal category of unit words. On the other hand, in order to account structurally for the transparency of classifiers, but not measure words, to modification, it is necessary to assume a split analysis in which classifiers occur in a right-branching structure while measure words occur in a left-branching structure. Her (2012) provides an extensive review of how prior proposals address (or fail to address) this basic problem, as well as the first and only LFG analysis to our knowledge. Here we briefly discuss two more recent analyses, one split and one uniformly right branching, before turning to Her's uniformly left-branching proposal.

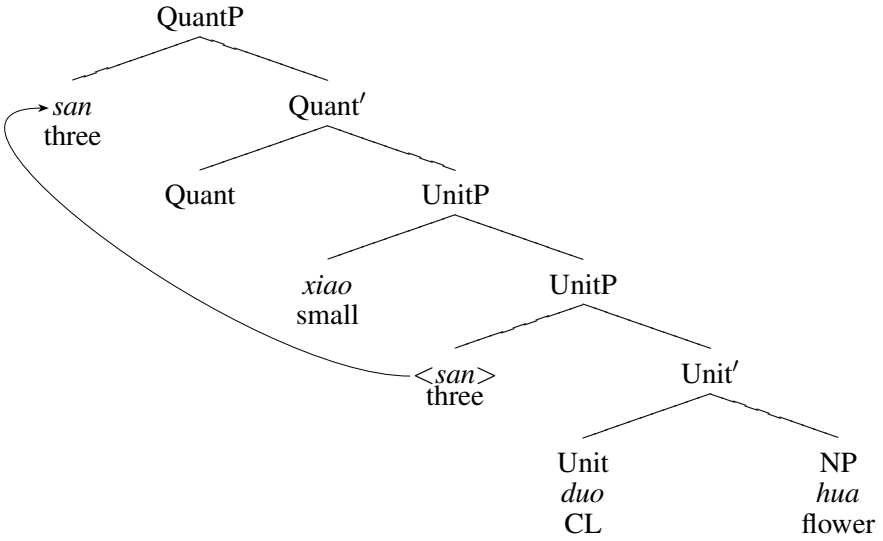
4.1 Zhang (2013)

Zhang (2013) proposes a complex split analysis which nevertheless attempts to maintain unit words as a unitary category. Classifiers and measure words are both Unit heads, although measure words start as noun heads which subsequently move

to the Unit head. Numerals originate structurally as the specifiers of the functional projection UnitP, later moving to the specifier of QuantP. In this framework UnitP then represents numerability, whether a noun can combine directly with a numeral. Dimensional adjectives which intervene between the numeral and the classifier then appear as adjuncts of UnitP.

(24) Right-branching structure for classifiers (Zhang, 2013, 233, ex (470b))

'three small flowers'

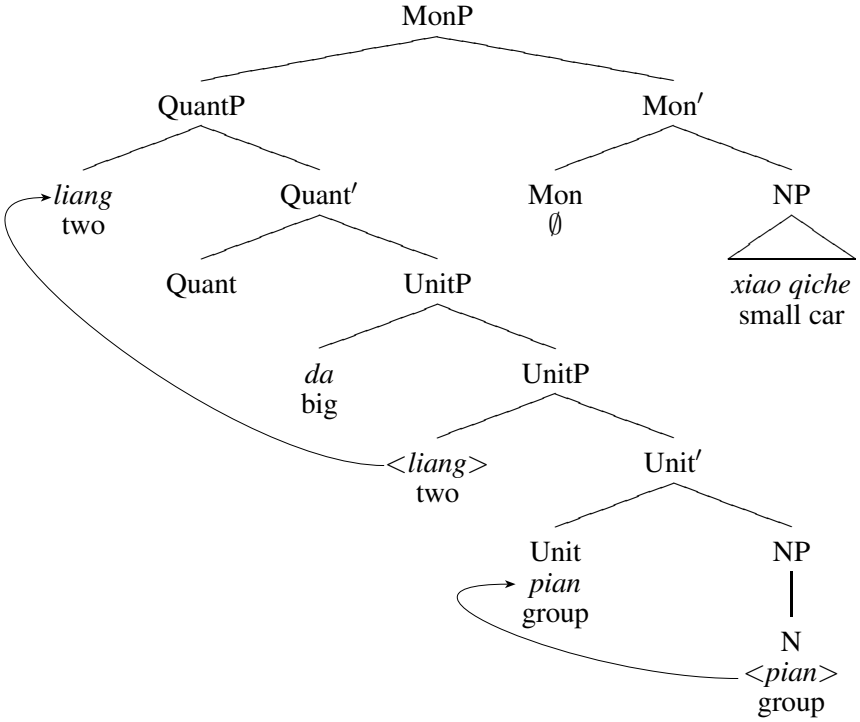


Any adjective in this structure, whether above or below the function head Unit, would c-command NP and thus scope over the NP below it.

By contrast, measure words are assumed by Zhang (2013) to occur in a (considerably more complex) left-branching structure as indicated in (25). The measure word used by Zhang to illustrate this structure belongs to the collection subtype, i.e. *pian* 'group', but container measure words are explicitly stated to have the same behaviour.

(25) Left-branching structure for measure words (Zhang, 2013, 235, ex (475b))

'two big groups of small cars'



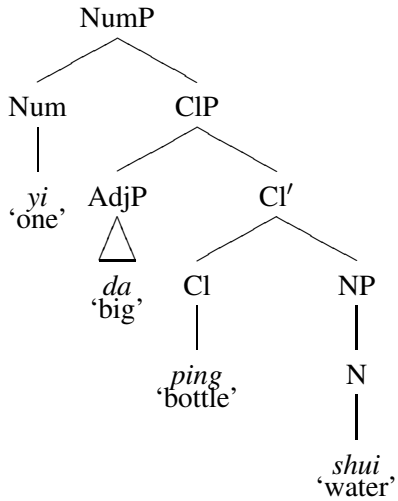
In this structure, the numeral and measure word sit inside the QuantP projection as before, but QuantP itself sits on a left branch of the main nominal projection (here represented by MonP, for "monotonicity phrase"). Ignoring the complexities of this analysis, which involves not only movement of the numeral but also the measure word, we see that any adjective within the QuantP branch of the tree is intended to apply to the measure word rather than NP. It is not clear, however, how this left-branching structure could account for examples in which measure words themselves are stacked.

4.2 Li (2013)

Li (2013) proposes that unit words belong to a unitary category called Cl, distinguishing between subcategories as [±Count, ±Measure]. In this system, classifiers are categorised as [+Count, −Measure] and the container subtype of measure word is [+Count, +Measure]. Other subtypes of measure word illustrate the remaining feature combinations, e.g. standard measures [−Count, +Measure] and kinds [−Count, −Measure]. The latter two will not concern us here, although we note that Li allows container measure words to function ambiguously as standard measure words when they denote the quantity associated with the container rather than the container per se, e.g. "six bottles of wine" when "six" does not literally count bottles, but the measure associated with bottles. For standard measures, Li adopts

a left-branching analysis. What we discuss here is the [+Count, +Measure] (container per se) reading, for which Li adopts the right-branching analysis given in (26).

(26)



Here, the projection ClP is headed by the classifier, and the ClP can itself be a complement of a higher functional category Num. Although Li does not explicitly give the tree structure for classifiers, it is claimed to be identical. That is, both classifiers and measure words belong to the same category Cl, and the structure assumed is uniformly right-branching.

Assigning classifiers and measure words to the same category has the advantage, as noted above, that these appear in complementary distribution in basic examples. As opposed to the split analyses, a uniform right-branching structure also has the advantage that in principle it might permit stacking of unit words as in examples (10) and (11). However, assuming that the NP in (26) does not branch further and cannot itself contain NumP, this structure as it stands does not permit any kind of stacking.

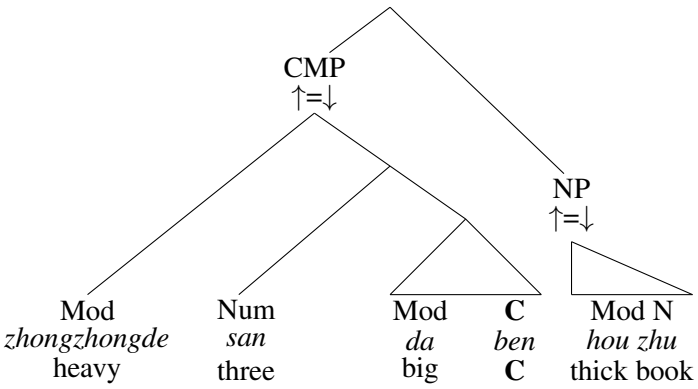
The uniform right-branching structure cannot also as it stands account structurally for the differences between classifiers and measure words with respect to adjectival modification. Li (2013, 184) addresses this issue by suggesting that the adjective in both the classifier and the measure word case applies to the constituent Cl + NP as a whole, rather than to the classifier itself. We agree that this is in principle correct. In the measure word case, Li adduces examples like *yi xiao bei putaojiu* (one small glass$_{mw}$ wine) 'a small glass of wine', where a possible interpretation is that the glass itself is big (on a scale of glass sizes) while the quantity of wine it contains is actually small (on a scale of amounts of wine). One puzzle is why Li does not think that this is a standard measure use of *putaojiu* 'glass', rather than the container use. The fact that an actual glass is involved, whatever its size, might then simply be a matter of pragmatic inference. Be that as it may, we note

that flexibility in the interpretation of scalar adjectives in container expressions is not exclusive to Chinese, but applies equally well to English which lacks classifiers and where containers are clearly denoted by nouns. The treatment of classifiers and measure words as belonging to the same category also obscures the basic distinctness of their contributions to the semantics. When a measure word is present, the scale implicit in the scalar adjective is in the first instance, in Chinese as in English, the size of the container, and in this respect measure words are clearly distinct from classifiers. Whether or not the contents of the container also form an appropriate scale depends very much on the nature of the contents: in the case of wine, there is a degree of pragmatic plausibility to the quantity of wine being an appropriate scale since units of alcohol are a prominent social concept. But a similar interpretation does not so naturally arise in examples like (26), where the content is water, or (22), where the contents are apples: 'one small box of apples' does not imply that the apples are small.

4.3 Her (2012)

Her (2012) crucially shows that within an LFG approach it is not necessary to account structurally for the differences between classifiers and measure words with respect to transparency to modification. He adopts a uniform left-branching approach in which classifiers and measure words belong to a unitary category CM. Both classifiers and measure words head CMPs which are sisters of NP, but classifiers are distinguished by being co-heads of N, while measure words have their own PRED value and head CMPs which function as an f-structure QUANTIFIER. Her does not indicate the category of the higher phrase to which CMP and NP belongs in each case. The two structures are shown schematically in (27) and (28) for classifiers and in (29) and (30) for measure words. It is the f-structure representation therefore that is split (Her, 2012, 1244-5). We would argue that the similarity between the c-structure trees in fact masks the fundamental difference between the two types: in effect, because of the significant difference in f-structure annotation (co-head vs non-co-head) CM is not really a unitary category.

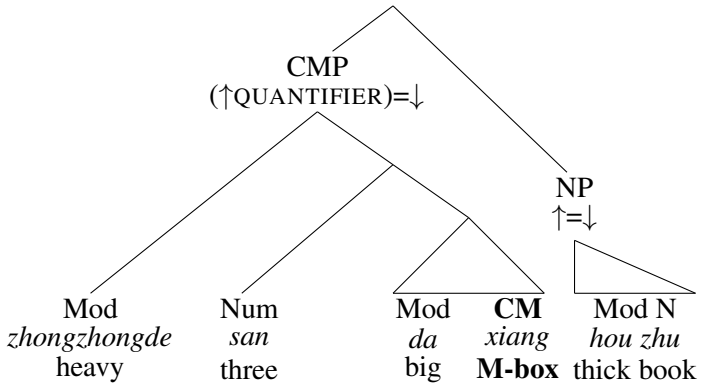
(27) Annotated left-branching tree for classifier



(28)

$$\begin{bmatrix} \text{PRED} & \text{`BOOK'} \\ \text{PROFILED} & \text{BEN} \\ \text{PROFILABLE} & \{\text{BEN, CE}\} \\ \text{CARD} & 3 \\ \text{ADJUNCTS} & \left\{ \begin{bmatrix} \text{“HEAVY”} \end{bmatrix} \begin{bmatrix} \text{“ BIG”} \end{bmatrix} \begin{bmatrix} \text{“THICK} \end{bmatrix} \right\} \end{bmatrix}$$

(29) Annotated left-branching tree for measure word



(30)

$$\begin{bmatrix} \text{ADJUNCTS} & \{\text{“THICK”}\} & \\ \text{PRED} & \text{`BOOK'} & \\ \text{PROFILABLE} & \{\text{BEN, CE}\} & \\ \text{QUANTIFIER} & \begin{bmatrix} \text{PRED} & \text{`BOX'} \\ \text{CARD} & 3 \\ \text{ADJUNCTS} & \left\{ \begin{bmatrix} \text{“HEAVY”} \end{bmatrix} \begin{bmatrix} \text{“ BIG”} \end{bmatrix} \right\} \end{bmatrix} & \end{bmatrix}$$

It is immediately clear that any modifier in the classifier structure will be a member of the ADJUNCT set of the NP as a whole, while any modifier in the measure word structure will be in the ADJUNCT set of the QUANTIFIER. That is, ”heavy” and ”big” apply to ”book” in (a) and ”box” in (b).

We will exploit Her’s insight that classifiers should be analysed as co-heads in our own analysis below. We note however two difficulties with Her’s analysis. One

is conceptual: the fact that classifiers and measure words belong to a single category CM and both project to CMP requires the use of complicated implications to enforce a match between unit word type and CMP type. Without these there would be nothing to stop a measure word with a PRED value from occurring in a CMP with an ↑=↓ annotation and a clash of PRED values with the PRED of the head noun. The necessity for such a manoeuvre arises from the postulation of a "hybrid" category, at once functional and lexical. Ideally we would like to avoid such categories. The second difficulty is more serious, and concerns the inability of the analysis to encompass structures such as (11) in which multiple unit words occur. As already pointed out with respect to Zhang's (2012) analysis, these are difficult to reconcile with a left-branching structure. There is no place within a CMP for another CMP. And if, in order to account for examples like (11) with stacked measure words, we were to allow two CMPs each headed by a measure word on a separate branch within a single NP , this would project two QUANTIFIER attributes in the corresponding f-structure, violating the principle of functional uniqueness.

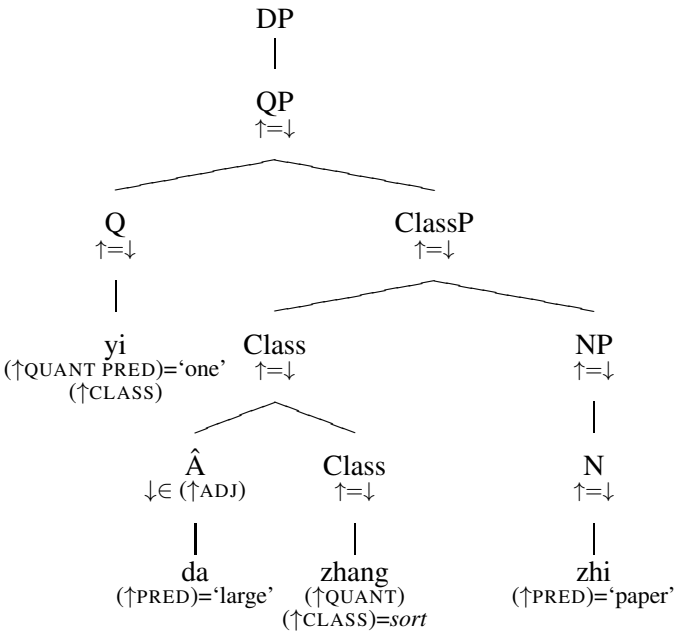
5 Our proposal

As suggested above, we adopt from Her (2012) the notion that classifiers should be treated as functional co-heads while measure words have their own PRED value. This accounts for the transparency of classifiers to modification. In order to account for unit word stacking, we adopt a uniform right-branching analysis, as proposed in a number of earlier structural analyses (for instance Cheng & Sybesma, 1998). Once the f-structure differences between classifiers and measure words are recognised, there is no fundamental barrier to a uniform right-branching analysis. Arguments for constituency which are based purely on the scoping of adjectival adjuncts lose their force. The requirement that classifiers are co-heads and measure words have their own PRED value entails however that they do not fundamentally belong to the same category. In our analysis, classifiers will belong to the functional category Class, while measure words will be (non-prototypical) nouns. The similarity between classifiers and measure words and their mutual incompatibility in basic structures will be treated as an f-structure characteristic.

5.1 Classifier c-structure

The structure we assign to classifiers is given in (31), using the example 'one large sheet of paper'.

(31)



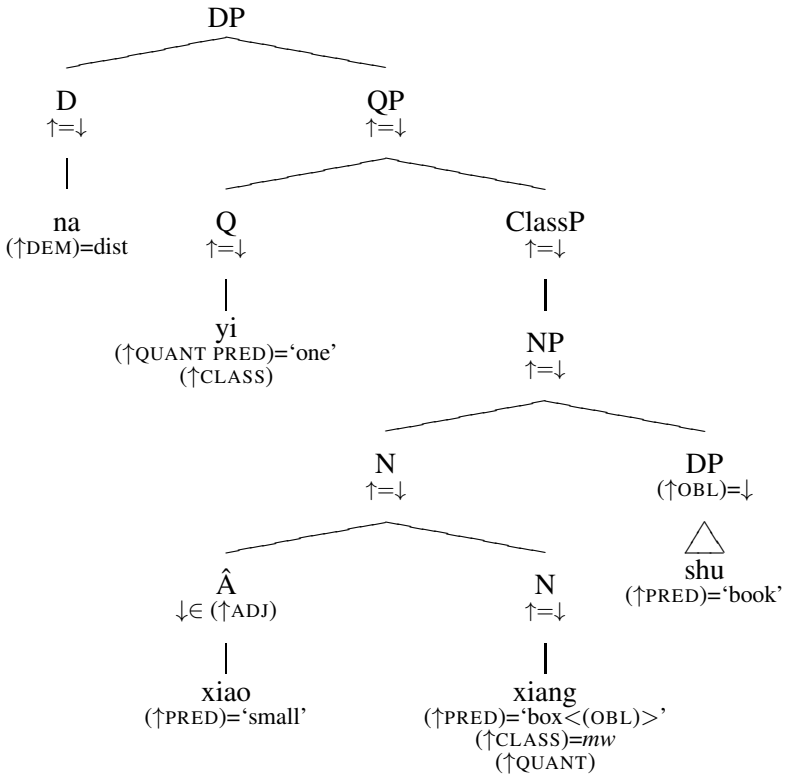
We assume for Chinese a richer hierarchy of functional categories than is typical for noun phrase analyses in LFG (and indeed for Chinese noun phrases within a minimalist framework, see Bošković (2013)). For illustrative purposes this hierarchy will include a DP projection which houses demonstratives. The hierarchy is justified not just by the strict ordering which is predicted (D - Q - Class - N), but also by the strikingly different modification possibilities at each level. As we have seen, NP modification, following the classifier, is the most varied, potentially consisting not just of single adjectives but also *de*-marked adjective phrases and relative clauses. The modification of a classifier is by contrast highly restricted: in effect it is a closed class of non-projecting mono-syllabic adjectives belonging to specific semantic subclasses. In line with the standard treatment of non-projecting categories in LFG Toivonen (2003), we assign these to the non-projecting category Â rather than the general category A (see also Sadler & Arnold, 1994). There is a further modification possibility associated with QP, as noted above and further illustrated in (23). This must be a *de*-marked adjective.

The mutual dependency between Q and CLASS is enforced by the f-structure annotations on these. The numeral is annotated (↑CLASS), an existential requirement that it occur within an f-structure where the value of CLASS is provided. The presence of a classifier satisfies this constraint, being annotated (↑CLASS) = *sort*. Conversely, a classifier, annotated (↑QUANT), requires the presence of a quantifier. Since Class (and Q) are co-heads, any modifier will be a member of the ADJUNCT set of the whole DP.

5.2 Measure word c-structure

The measure word structure is given in (32):

(32)

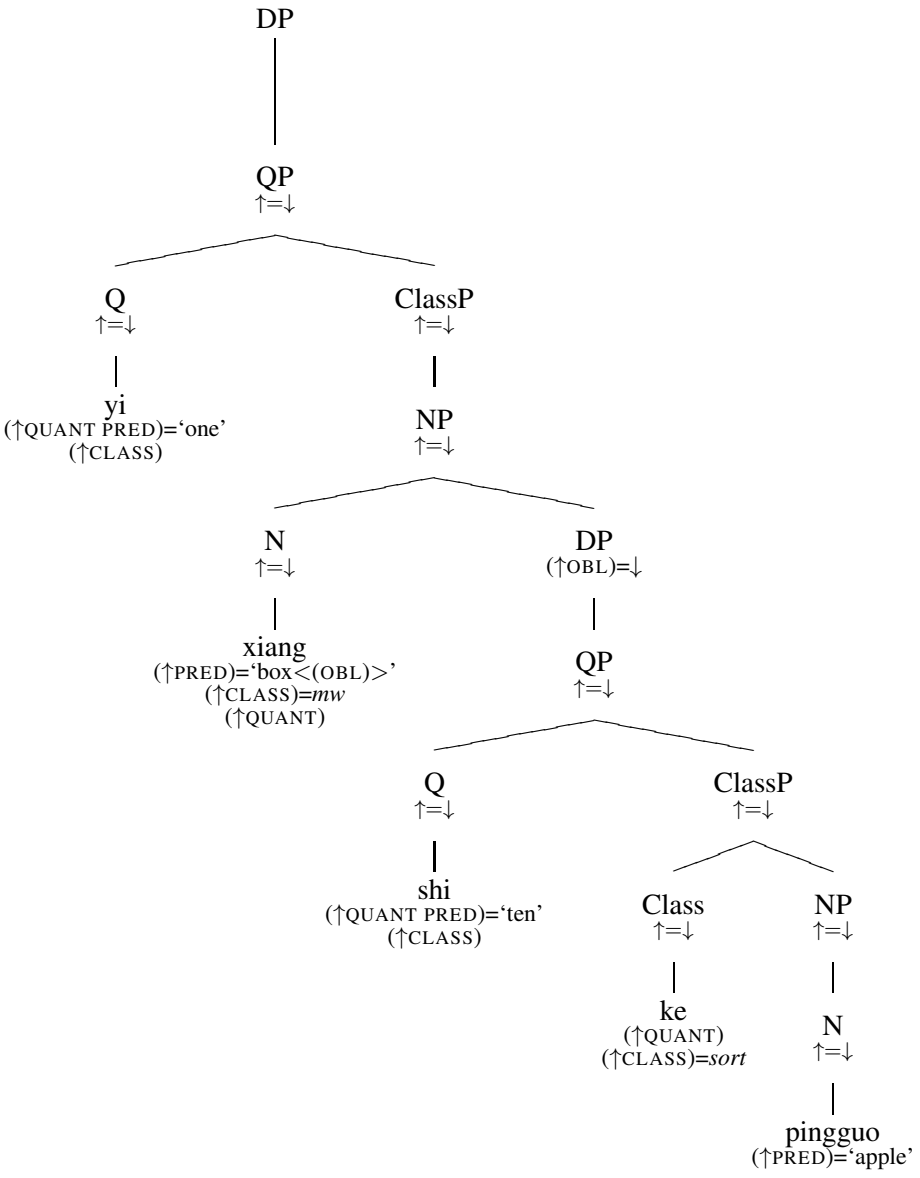


In this structure, the measure word is of category N and has its own PRED value. The similarity of measure words to classifiers is essentially captured by the f-structure annotation (↑CLASS) = *mw*. Just as with classifiers, this annotation allows quantifiers to have their requirement for a unit word satisfied, and the different values of (↑CLASS) ensure, through functional uniqueness, that classifiers and measure words cannot co-occur within the same simple f-structure. There is no longer a need for them to occupy the same structural slot for this mutual incompatibility to be enforced.

Crucially, this PRED value assigned to measure words and which allows us to distinguish them lexically also allows us to specify an argument structure. In other words, a measure word, just like any relational noun, takes an argument. We assign this argument the relation OBL, similar to that of an English *of*-PP (although it could also appropriately be assigned the specifically noun-phrase role of NCOMP (Chisarik & Payne, 2001)). Structurally, this argument is potentially a full DP (including a demonstrative), and this straightforwardly permits the stacking of unit words that we have seen in (10) and (11). If a classifier occurs in such a DP,

predictably it will be the last unit word in the structure since the DP cannot simultaneously contain a measure word with its own new argument. In (33) and (34) we provide the c- and f-structure for (10) and in (35) and (36) for (11).

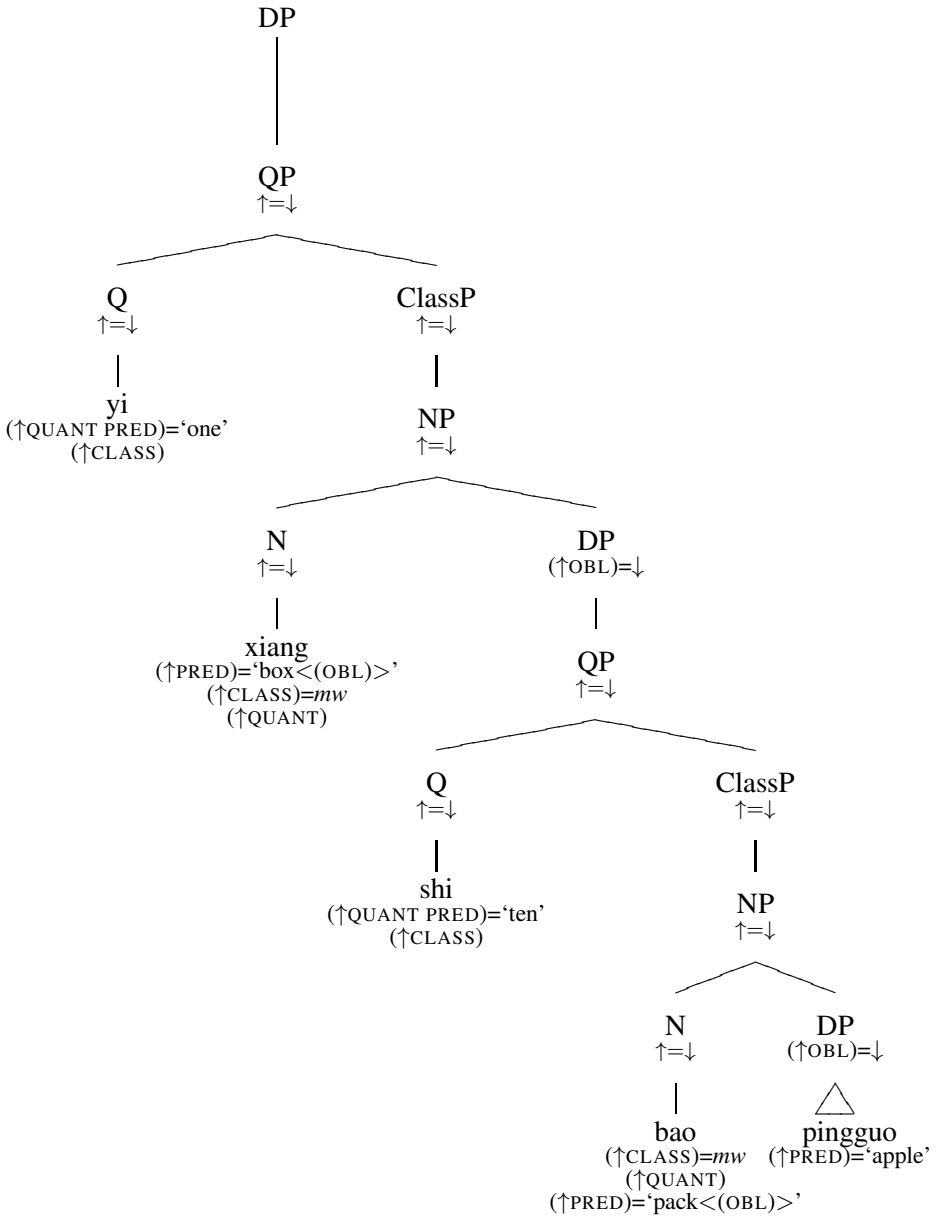
(33)



(34)

$$\begin{bmatrix} \text{PRED} & \text{‘BOX<OBL>’} \\ \text{CLASS} & \text{MW} \\ \text{QUANT} & \begin{bmatrix} \text{PRED} & \text{‘ONE’} \end{bmatrix} \\ \text{OBL} & \begin{bmatrix} \text{PRED} & \text{‘APPLE’} \\ \text{CLASS} & \text{SORT} \\ \text{QUANT} & \begin{bmatrix} \text{PRED} & \text{‘TEN’} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

(35)


DP
QP
↑=↓
Q
↑=↓
yi
(↑QUANT PRED)='one'
(↑CLASS)
ClassP
↑=↓
NP
↑=↓
N
↑=↓
xiang
(↑PRED)='box<(OBL)>'
(↑CLASS)=mw
(↑QUANT)
DP
(↑OBL)=↓
QP
↑=↓
Q
↑=↓
shi
(↑QUANT PRED)='ten'
(↑CLASS)
ClassP
↑=↓
NP
↑=↓
N
↑=↓
bao
(↑CLASS)=mw
(↑QUANT)
(↑PRED)='pack<(OBL)>'
DP
(↑OBL)=↓
pingguo
(↑PRED)='apple'

(36)

$$\begin{bmatrix} \text{PRED} & \text{`BOX<OBL>'} \\ \text{CLASS} & \text{MW} \\ \text{QUANT} & \begin{bmatrix} \text{PRED} & \text{`ONE'} \end{bmatrix} \\ \text{OBL} & \begin{bmatrix} \text{PRED} & \text{`PACK<OBL>} \\ \text{CLASS} & \text{MW} \\ \text{QUANT} & \begin{bmatrix} \text{PRED} & \text{`TEN'} \end{bmatrix} \\ \text{OBL} & \begin{bmatrix} \text{PRED} & \text{`APPLE'} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

6 Conclusion

In this paper we have provided a new analysis of Chinese unit words which exploits the LFG distinction between c-structure and f-structure. The order of categories and the range of modification permitted by each category is, with one exception, essentially syntactic and accounted for by c-structure rules. On the other hand, the mutual dependence between unit words and quantifiers, and the mutual incompatibility of these is accounted for in f-structure. The primary differences between classifiers and measure words are also explained by f-structure: the transparency of classifiers to modification follows from their treatment as coheads, while the opaqueness of measure words follows immediately from their PRED structure. These features enable us to treat unit word structures as uniformly right-branching, a sine qua non for the analysis of more complex structures with stacked unit words.

Rather than assume a single category for unit words in Chinese, we have assigned classifiers to a distinct functional category and measure words to a lexical category. The fact that measure words have nominal meanings, are opaque to modification, and have their own argument structure clearly points to a fundamental difference between measure words and classifiers. Measure words are nouns. Nevertheless, there is one aspect of measure words which does not follow from our analysis as it stands. It is usually claimed that measure words accept only the same kind of limited modification as classifiers, a fact which we have captured by allowing them to be modified by Â. This is not of course a typical property of nouns, which generally allow full AP modification. From a historical point of view, it looks as though measure words, while maintaining most of their lexical characteristics, have united with classifiers not only in their mutual interdependence with numerals, but also in their limited modification. We leave it an open question how technically to enforce the requirement that measure words like classifiers do not appear to take full AP modifiers. Possibly this can be linked with the f-structure annotation (↑CLASS). A fuller study is probably needed to check that there is genuinely no difference between measure words and classifiers in this respect.

References

Börjars, Kersti, Erika Chisarik & John Payne. 1999. On the justification for functional categories in LFG. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG99 Conference*, Stanford, CA: CSLI Publications.

Bošković, Željko. 2013. On NPs and clauses. In Günther Grewendorf & Thomas Ede Zimmermann (eds.), *Discourse and grammar: from sentence types to lexical categories*, 179–246. De Gruyter Mouton.

Cheng, Lai-Shen Lisa & Rint Sybesma. 1998. Yi-wangtang, yi-getang: classifiers and massifiers. *Tsing-Hua Journal of Chinese Studies* New Series XXVIII(3). 385–412.

Chisarik, Erika & John Payne. 2001. Modelling possessor constructions in LFG: English and Hungarian. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG01 Conference*, Stanford, CA: CSLI Publications.

Erbaugh, Mary S. 1986. Taking stock: the development of chinese noun classifiers historically and in young children. In Colette G. Craig (ed.), *Noun classes and categorization: Proceedings of a symposium on categorization and noun classification, Eugene, Oregon, October 1983*, 399–436. Amsterdam & Philadelphia: John Benjamins.

Fan, Jiyan. 1958. Xing ming zu he jian 'de' zi de yufa zuoyong [the grammatical function of *de* in the adjective-noun combination]. *Zhongguo Yuwen* 71. 213–217.

Her, One-Soon. 2012. Structure of classifiers and measure words: A lexical functional account. *Language and Linguistics* 13(6). 1211–1251.

Her, One-Soon & Chen-Tien Hsieh. 2010. On the semantic distinction between classifiers and measure words in Chinese. *Language and Linguistics* 11(3). 527–551.

Hu, Qian. 1993. *The acquisition of Chinese classifiers by young Mandarin-speaking children*: Boston University Graduate School dissertation.

Huang, Chu-Ren. 1989. *Mandarin chinese NP de – a comparative study of current grammatical theories*: Cornell University 1987 dissertation.

Kroeger, Paul R. 1993. *Phrase structure and grammatical relations in Tagalog*. Stanford, CA: CSLI Publications.

Li, XuPing. 2013. *Numeral classifiers in Chinese. The syntax-semantics interface*. Berlin, Boston: De Gruyter Mouton.

Lyons, John. 1977. *Semantics*, vol. II. Cambridge: Cambridge University Press.

McEnery, Tony & Richard Xiao. 2010. *Corpus-based contrastive studies of English and Chinese*. London: Routledge.

Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago: University of Chicago Press.

Paul, Waltraud. 2010. Adjectives in Mandarin Chinese. The rehabilitation of a much ostracized category. In Cabredo Hofherr Patricia & Ora Matushansky (eds.), *Adjectives formal analyses in syntax and semantics*, Amsterdam: John Benjamins.

Peyraube, Alain. 1991. Some remarks on the history of Chinese classifiers. In Patricia M Clancy & Sandra A Thompson (eds.), *Santa Barbara paper in linguistics, vol 3: Asian discourse and grammar*, 106–126. Department of Linguistics, University of California, Santa Barbara.

Sadler, Louisa & Douglas J Arnold. 1994. Prenominal adjectives and the phrasal/lexical distinction. *Journal of Linguistics* 30(1). 187–226.

Sproat, Richard & Chilin Shih. 1988. Prenominal adjectival ordering in English and Mandarin. In *Proceedings of the North Eastern Linguistic Society meeting 18*, 465–489. Amherst, MA : GLSA.

Sproat, Richard & Chilin Shih. 1991. The cross-linguistic distribution of adjective ordering restrictions. In Carol Georgopoulos & Roberta Ishihara (eds.), *Interdisciplinary approaches to language: essays in honor of S-Y Kuroda*, 565–593. Dordrecht: Kluwer.

Tai, James & Lianqing Wang. 1990. A semantic study of the classifier *tiao*. *Journal of the Chinese Language Teachers Association* 25(1). 35–56.

Tang, Chih-Chen Jane. 1990. *Chinese phrase structure and the extended X′-theory*: Cornell University dissertation.

Toivonen, Ida. 2003. *Non-projecting words: a case study of Swedish particles*. Dordrecht: Kluwer.

Wang, Lianqing. 1994. *Origin and development of classifiers in Chinese*: Ohio State University dissertation.

Zhang, Niina Ning. 2012. Numeral classifier structures. Manuscript. National Chung Cheng University.

Zhang, Niina Ning. 2013. *Classifier structures in Mandarin Chinese*. Berlin, Boston: De Gruyter Mouton.

Schematising (Morpho)Syntactic Change in LFG: Insights from grammaticalisation in Arabic

Maris Camilleri
University of Essex

Louisa Sadler
University of Essex

Proceedings of the LFG'18 Conference

University of Vienna

Miriam Butt, Tracy Holloway King (Editors)

2018

CSLI Publications

pages 129–149

http://csli-publications.stanford.edu/LFG/2018

Keywords: Arabic, historical change, embedding, perfect, progressive

Camilleri, Maris, & Sadler, Louisa. 2018. Schematising (Morpho)Syntactic Change in LFG: Insights from grammaticalisation in Arabic. In Butt, Miriam, & King, Tracy Holloway (Eds.), *Proceedings of the LFG'18 Conference, University of Vienna*, 129–149. Stanford, CA: CSLI Publications.

Abstract

This paper explores the grammaticalisation of two particular constructions in the Arabic vernaculars, seeking to provide an account of the relevant grammaticalisations paths and the commonality between then, using LFG as the theoretical model of morphosyntactic change. The two constructions express the PROGRESSIVE and the Universal PERFECT respectively. While their synchronic syntax has been recently analysed, here we address the task of exploring how hypotheses of reconstructed developmental paths that have led to the formation/grammaticalisation of these constructions could be accounted for by using the machinery of LFG. In particular, we observe how change does not necessarily constitute, or equate to, changes at the *c*-structure level. Alternatively, changes in function need not be accompanied by, or correlated with changes in form. While appreciating that the synchronic syntaxes of the two constructions under consideration are distinct, we observe how they share part of the developmental path that has led to their respective formation, and that is the shift from adjunction to embedding; a shift also observed in syntactic developments in Indo-European.

1 Introduction

The constructions to be discussed here are first the PROGRESSIVE construction, and the other, the Universal PERFECT construction. In each case we are concerned with the emergence of what are functionally verbal auxiliary elements and the emergence of a dedicated structure for the expression of a particular meaning. The analytical deductions presented here, as well as the hypothesised grammaticalisation trajectories are not derived from any historical evidence, given the lack of written material for the vernacular Arabic varieties. Rather, the conclusions made are constructed by microvariation observed when comparing the synchronic syntax of the different varieties, and the cues provided through whatever diachronic morphosyntactic vestiges are available within their different grammars.

Both the constructions to be considered here express ASPECTual values and involve some form of verbal auxiliation, but beyond this, they have/call for/motivate rather distinct synchronic syntactic analyses. Notwithstanding this difference, our aim here is to suggest that there are significant common aspects to the diachronic path of development in these cases. In particular, we suggest that a change from adjunction to embedding is common to both, in particular from an XADJ to an XCOMP possibly as the result of argument-extension. It is following this point (and hence from this point forward) in the grammaticalisation process that the constructions develop their distinct paths. The change from clausal adjunction to clausal embedding has been said to characterise a number of syntactic shifts that have taken

[†]This work was partially funded by a Reach High Scholars Programme - Post Doctoral Grant, part-financed by the EU, Operational Programme II - Cohesion Policy 2014 - 2020 "Investing in human capital to create more opportunities and promote the well being of society" (ESF) and by Leverhulme Major Research Fellowship MRF-2016-048. Support from these sources is gratefully acknowledged.

place through time across Indo-European (Kiparsky, 1995), but we are not aware of any previous theoretically oriented work hypothesising such a development in the literature on Arabic.

These two constructions provide fertile ground for exploring grammaticalisation from an LFG perspective: they are rich in terms of morphosyntactic diversity across the various Arabic vernaculars (giving rise to a range of form-function mismatches), and their synchronic syntax is now reasonably well-understood (they have both been the focus of some recent work). They also exemplify the two ways with which LFG deals with the analysis of auxiliaries, following Falk (2008).

Synchronically, the auxiliary *gāʕid* in the PROGRESSIVE construction in (1) is a co-head with the lexical verb, in an AUX-feature analysis. The combination of this form with the following imperfective form of the lexical verb contributes the feature ASPECT = PROG to the f-structure (alongside a TNS value) (Camilleri and Sadler, 2017). On the other hand, the auxiliary that functions as the main exponent of the universal perfect in Arabic, which in the case of the Syrian construction in (2) is *il* (in its inflectioned forms), is a PRED-taking auxiliary, and the construction behaves as a raising structure (Camilleri, 2017, under review).

(1) al-muğtama **gāʕid** i-ṭ-ṭawwar
DEF-society.SGM sit.ACT.PTCP.SGM 3SGM-REFL-develop.IMPV
The society is developing. Bahraini: Persson (2009a, 266)

(2) (muna) **il**-a ḥamst iyyām bi-l-ħabis
Muna to-3SGF.GEN five day.PL in-DEF-prison
Muna has been in prison for five days. Syrian: Hallman (2016, 77)

We first briefly say a word on grammaticalisation and work on grammaticalisation in LFG. In §3 and §4, we then discuss the grammaticalisation of the PROGRESSIVE and Universal PERFECT constructions, respectively. §5 concludes.

2 Grammaticalisation

Grammaticalisation is a mechanism that takes place time through time whereby independent lexical items start losing parts of their lexical content and eventually come to express grammatical functions and meanings (Meillet, 1912), following clines (Bybee et al., 1994a; Hopper and Traugott, 2003). These changes do not occur in a vacuum, but rather are internal to syntactic structures. Together with the grammaticalisation of the lexical items, we also find the eventual grammaticalisation of a construction itself (Hopper and Traugott, 2003). Such grammaticalisation is often linked to the notions of deinflection and loss of agreement (Lehmann, 1995).

Work on grammaticalisation in LFG features particularly in the works of Butt (1996), Barron et al. (1997), Schwarze (2001), and Camilleri and Sadler (2017).

Butt and Lahiri (2002); Butt and Geuder (2003); Seiss (2009); Butt and Lahiri (2013), for instance, have been central to the discussion of how grammaticalisation distinguishes between auxiliaries and light verbs, and how this difference is reflected at the level of theory; while the latter can form complex predicate structures, and are themselves an end on a cline, on the other hand, auxiliaries are on a distinct grammaticalisation cline, and this precludes them from forming complex predicate structures.

The overarching theme present in Vincent (2001); Vincent and Börjars (2010); Börjars et al. (2016) is that of using the architecture of LFG as a means with which to better understand grammaticalisation and change by exploiting, accounting for, and dealing with a number of form-function mismatches. Previous discussions have concentrated on how meaning shifts and change need not affect the external syntactic structure in any way. The string may well remain the same, and the observed change has to do with the functional structure. This is the case when we observe the change that occurs when shifting from clausal adjunction to embedding. In other instances, meaning shifts and changes result from changes in the a-structure with no changes in either the c- or the f-structure, as would be the case of the phase in the formation of a raising predicate once what's left is the loss of the SUBJ's thematicity.

In what follows we use LFG very much in the way that others have used LFG within the domain of grammaticalisation, i.e. both to guide the step-by-step process that we hypothesise took place, and to illustrate how the change may effect, in distinct ways, either the f-, the c-, or the a-structures, and a change at one level of syntactic structure, e.g. the f-structure, need not have an effect on the c-structure, or vice-versa.

3 Grammaticalisation of the PROGRESSIVE construction

We start with the development of the progressive constructions (illustrated in (3), with (3a) repeated from (1) above), using the active participle *gāʕid/ǧālis* (lexical meaning 'sit') in auxiliary function, with a following imperfective lexical verb. This is found across the different Arabic vernaculars, but is not found in Classical Arabic.

(3) a. al-muǧtama **gāʕid** i-ṭ-ṭawwar
DEF-society.SGM sit.ACT.PTCP.SGM 3SGM-REFL-develop.IMPV
The society is developing. Bahraini: Persson (2009a, 266)

b. **yālis** yi-bni ʕmāra
sit.ACT.PTCP.SGM 3SGM-build.IMPV building
He is building a building. Emirati: Jarad (2015, 102)

The construction in (3) is just one of a number of strategies employed to express *progressive aspect*, through which we understand that given states or actions

are in progress at a particular reference time. These include the use of the imperfective verb form itself (which also expresses HABITUAL and CONTINUOUS readings) (Mitchell and al Hassan, 1994; Camilleri and Sadler, 2017); the use of the active participial forms of the lexical verb (subject to restrictions as to lexical aktionsart and not available in all dialects);[1] the use of auxiliary forms such as: *ʕammāl* lit. 'doing' and shortened counterparts in Levantine/Mesapotamian dialects (Agius and Harrak, 1987); grammaticalisation of the copula 'be' in (certain) Anatolian dialects (Akkuş, 2016); the use of prefixes such as *bi-*$_{[\text{non-1SG}]}$/*bayn-*$_{[\text{1SG}]}$ in (Ṣaʕānī) Yemeni (Watson, 1993); *ka-/ta-* in Moroccan and Algerian (Harrell (1962); Heath (2013); Souag (2006)); and the use of an imperfective form + *fi* 'in', in the case of transitive verbs in Tunisian and Libyan (Mion (2004); Pallottino (2016); McNeil (2017); Börjars et al. (2016)). Beyond this diversity, the vernaculars all have in common the use of the auxiliaries *gāʕid/ǧālis* (and their phonological variants and/or cliticised or affixed counterparts), which precede imperfective verb-forms. These forms are morphologically inflecting active participial forms that have lexical meanings that range from 'sitting; staying; remaining' in most vernaculars to more bleached uses of 'be located; situated' and *exist* in dialects such as Chadian and Libyan (Absi and Sinaud, 1968; Rubin, 2005; Pereira, 2008). In Maltese the lexical counterpart of the form *qiegħed* has in fact become highly lexicalised, meaning 'stagnant' and 'unemployed'.

This progressive construction is given attention in a number of descriptive works e.g. Johnstone (1967); Cuvalay (1991); Brustad (2000); Mion (2004), and has also received some analytic attention, e.g. Woidich (1995); Persson (2009b); Persson (2013); Jarad (2015). Camilleri and Sadler (2017) analyse examples like (3) as involving a feature-bearing auxiliary that co-heads the structure together with the lexical predicate, arguing both against an analysis where the construction could be analysed as a complex predicate construction, with *gāʕid/ǧālis* analysed as light verb, as well as an analysis where these auxiliaries headed the construction on their own as PRED-taking auxiliaries. Building on this analysis, in this contribution we consider the possible developmental path that has led to the grammaticalisation of this construction in Arabic.[2]

The development of a progressive auxiliary from a posture verb is quite a common grammaticalisation path crosslinguistically (e.g. Bybee and Dahl (1989); Bybee et al. (1994b); Heine (1993); Heine and Kuteva (2002); Seiss (2009)). Here we suggest a possible diachrony for this development in Arabic, using LFG to formalise our hypothesis.

[1] See Borg (1988); Henkin (1992); Woidich (1995); Mughazy (2005); Procházka and Batan (2015); Camilleri (2016).

[2] The reader should keep in mind that this grammaticalisation should also be understood within the current synchronic context where in a number of dialects, the imperfective morphological form itself is still able to express a PROGRESSIVE reading. Additionally, and consistent with Deo's (2015) Imperfective cycle, this construction is broadening to express habitual and characterising readings alongside the event-in-progress reading, as discussed in Camilleri and Sadler (2017), as well as a number of more specific DURATIVITY, INCEPTIVE and CONTINUATIVE meanings in certain dialects.

The initial core meaning of the ACT.PTCP of the posture verbs involved is 'sitting', which is intransitive. In synchronic structures such as (4) the additional locative NPs and PPs are adjuncts.

(4) a. šāf walad mū **ʔāʕid**, {bi-l-ħadīʔa}
see.PFV.3SGM boy NEG sit.ACT.PTCP.SGM in-DEF-garden
He saw a boy (that is) not sitting in the garden.
Lebanese: Ghadgoud (2018, 245) - **ʔāʕid**<SUBJ> **+ PP** ADJ

b. niswān **gāʕd-īn** {hinī}
woman.PL sit.ACT.PTCP-PL here
The women are sitting here.
Gulf Arabic: Persson (2009a, 249) - **gāʕid**<SUBJ> **+ (locative) NP** ADJ

The very initial stage prior to any grammaticalisation might have involved a clausal ADJ, predicated of the matrix SUBJ. Circumstantial adjunct clauses (or *ħāl*) clauses) are very common in Arabic (Badawi et al. (2003); Ryding (2005); Persson (2009a)). They can be verbal, involving imperfective or participial forms, thus explaining why the associated synchronic verb in the progressive construction is never perfective in form or non-verbal, and either asyndetic or syndetic. The eventuality in the matrix is understood as taking place concurrently with whatever eventuality is expressed by the circumstantial clause — generally, but not always, the subject is shared. Given this we hypothesize that the initial stage is along the lines of (5), as exemplified by (6) (and many other examples) along with the f-structure associated with (6c) (note that (6c) additionally shows that circumstantials *can* have disjoint subjects).

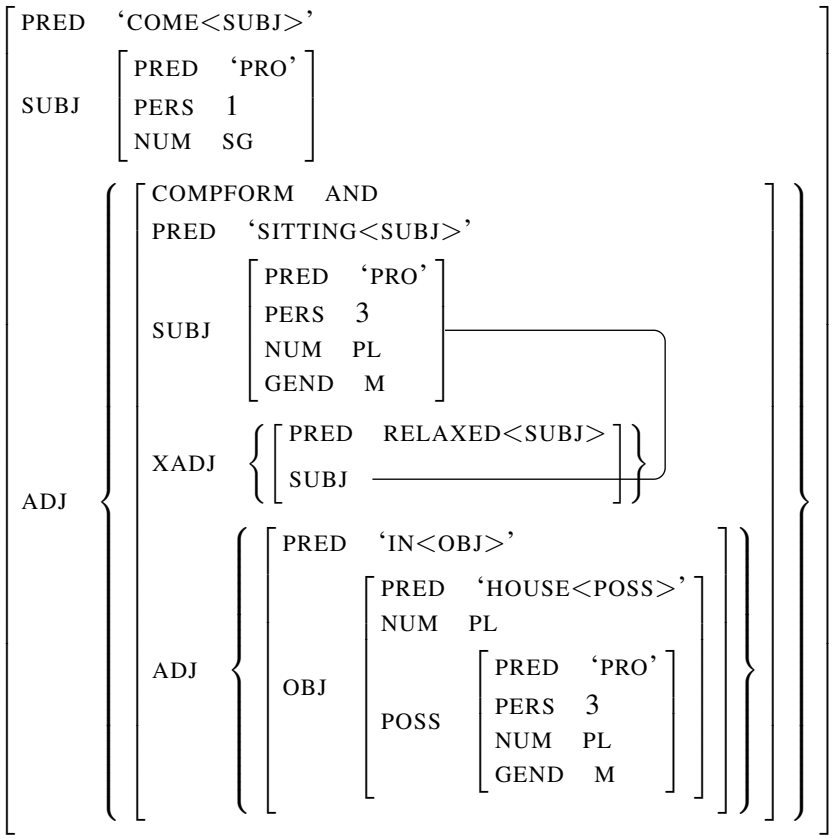
(5) Stage 0: 'sitting<SUBJ>' + XADJ, with (↑SUBJ) = (↑XADJ SUBJ)

(6) a. ana **gāʕid-a** {wa a-ɣsil aθ-θiyāb}
I sit.ACT.PTCP-SGF CONJ 1SG-wash.IMPV DEF-clothes
I am sitting (and) washing clothes.
Gulf Arabic: Persson (2009a, 250)
gāʕid<SUBJ> **+ circumstantial** XADJ **introduced by *wa* 'and'**

b. lagē-ta-h **gāʕid** {ya-smaʕ
find.PFV-1SG-3SGM.ACC sit.ACT.PTCP.SGM 3SGM-hear.IMPV
al-giṣidah}
DEF-poem
I found him sitting down listening to the poem.
Wādi Ramm Jordanian: Almashaqba et al. (2015, 162)
gāʕid<SUBJ> **+ syndetic circumstantial** XADJ

c. ği-t {wa-hum **ğālis-in** fi
come.PFV-1SG CONJ-3PLM.NOM sit.ACT.PTCP-PLM in
biyūt-hum {mu-rtāħ-in}}
house.PL-3PLM.GEN PASS.PTCP-relax-PLM

I came while they were sitting in their houses relaxed.
(Ṣaʕānī) Yemeni: Watson (1993, 380)

(7)


We hypothesise increased cohesion, and reanalysis of the XADJ as an XCOMP:

(8) Stage I: 'sitting<SUBJ, XCOMP>' where (↑SUBJ) = (↑XCOMP SUBJ)

Synchronically, there is of course an asyndetic relation between the auxiliary and the lexical verb in the progressive construction, while the circumstantial construction (see (6)) occurs with both syndetic and asyndetic linkage of the adjunct. We hypothesise the reanalysis of adjunction into embedding (as a result of increased cohesion) did not necessarily go hand-in-hand with simultaneous disappearance of the syndetic linkage (using *wa* which is synchronically the coordinating particle) at the point of functional reanalysis. The elimination of syndetic marking may have only taken place later, when the structure was understood as involving one eventuality, rather than two, although adjacency itself potentially plays

an important role as a trigger for structural analysis. Here structural analysis involves essentially argument-structure extensions rather than c-structure changes.[3]

We hypothesize that the next stage involved the semantic bleaching of 'sitting' into a wider spatial location, resulting in the SUBJ's loss of thematicity (in these contexts), giving a raising structure:

(9) Stage II: 'sitting<XCOMP>SUBJ' where (↑SUBJ) = (↑XCOMP SUBJ)

Hand in hand with this we suggest that semantic changes emerged in the lexical counterpart of the active participle, with NP/PP ADJs being reanalysed as OBJ_{loc}/OBL GFs with argument extension to '*gāʕid*<SUBJ, {OBJ_{loc}|OBL}>'. Synchronically, as well as the 'fully postural' lexical uses in (4)-(6) above, we find evidence of a 'functional split' Hopper and Traugott (2003) or 'divergence' Heine and Reh (1984), where one of the lexical meanings of *gāʕid* is (transitive) 'staying/remaining'.

(10) a. hūwa lāgi l-žeww mlīħ fa
he find.ACT.PTCP.SGM DEF-ambiance.SGM good.SGM so
gāʕəd ɣādi
stay.ACT.PTCP.SGM there
He found that the ambiance is good, so he is staying there.
Libyan: Pereira (2008, 402) - **gāʕid**<SUBJ**,** OBJ_{LOC}>

b. ʔinta **gāʕid** fi tšād walla?
you stay.ACT.PTCP.SGM in Chad INTERROG.MRKR
Are you staying in Chad?
Chadian: Absi and Sinaud (1968, 126) - **gāʕid**<SUBJ**,** OBL>

The final stage of grammaticalisation of the progressive construction involves loss of the auxiliary's PRED value, and the fusion of the bi-clausal f-structure into a mono-clausal one, in which *gāʕid* functions as an AUX-feature, while the XCOMP's PRED now functions as the (lexical) co-head in the same f-structure as *gāʕid*.

(11) Stage III: Loss of *gāʕid*'s PRED value; XCOMP PRED > matrix PRED

What is left from the (original) lexical 'sitting' is merely the temporal unboundedness of the erstwhile stative eventuality, a situation which lends itself rather easily to the development of a PROGRESSIVE (or CONTINUOUS/DURATIVE) interpretation (Kuteva, 1999). This stage accounts for the data in (3) and other presented in Camilleri and Sadler (2017). Once established, the progressive AUX+main verb construction has undergone further morphosyntactic and morphophonological changes (in some varieties) going down the grammaticalisation cline: (full verb)

[3]For the languages she looks at in her account of clause fusion, Fischer (2007, 214) couples adjacency with the presence of some sort of anaphoric relation between the clauses, in order for them to eventually result in some integrated structure. This coheres with the obligatory SUBJ structure-sharing across clauses we find in the progressive construction.

> auxiliary > clitic > affix (Hopper and Traugott, 2003, 108). The Iraqi example in (12) illustrates the full lexical form *gāʕid* (meaning 'sitting') as well as the synchronic prefix *de-* attached onto the imperfective form, realizing PROGRESSIVE ASPECT, and diachronically derived from *gāʕid*.

(12) Maryam **de**-ti-lʕab {wahiya **gāʕd-a** ʕala
Mary PROG-3SGF-play.IMPV CONJ.3SGF.NOM sit.ACT.PTCP-SGF on
l-kursī}
DEF-chair
Maryam is playing while she is sitting on the chair. Iraqi

The reconstruction of the diachronic path suggested here is largely hypothetical, because we do not have solid historical data for the spoken vernaculars, and neither do any of these synchronic varieties provide unambiguous evidence of the intermediate stage II where the auxiliary is still a PRED-taking auxiliary, involving a sense along the lines of:

(13) The clothes are lying (in some spatial location) drying/to dry.

If the argument made by Butt and Lahiri (2002), Butt and Geuder (2003) and Butt and Lahiri (2013) that light verbs are diachronic dead ends is correct, then a complex predicate construction containing a light-verb is ruled out as a diachronic precursor to the synchronic AUX-feature progressive construction. The alternative is that the AUX-feature analysis of the synchronic progressive construction has most likely developed out of a raising predicate, postulating an instance of the trajectory described by Vincent (2001, 24): "For a verb to develop into a raising verb involves the loss of theta-role assignment to one of its argument positions, a kind of semantic bleaching. **If a verb goes on to full auxiliary status** [as is the case with 'have' in PERFECT constructions, in English], the bleaching goes a step further and both subject and object arguments lose their independent thematic value".[4] Figure 1 visually represents the hypothesised diachronic developments.[5]

4 Grammaticalisation of the Universal Perfect

The perfect is often thought of (from a Eurocentric point of view) as a grammatical construction which essentially involves an auxiliary together with a participial form. We can distinguish two broad types of interpretation; the Existential/experiential perfect and the Universal/continuous perfect (McCawley, 1971,

[4]Of course, not all PRED-bearing auxiliaries are appropriately analysed as raising predicates (Falk, 2008).

[5]Though we cannot discuss this additional development here, it should be noted that at least in some varieties, *gāʕid* is also emerging (or is already established) as a copula. For these cases a similar path to that schematised in Figure 1 is additionally envisaged.

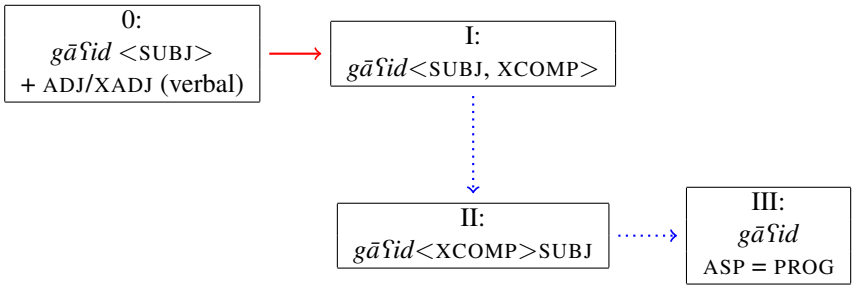


Figure 1: Grammaticalisation of gāʕid

1981; McCoard, 1978). The universal perfect conveys the meaning that the occurrence of an eventuality persists until reference time, in contrast to the existential reading, which merely asserts that the (episodic) occurrence of an eventuality remains of current relevance at reference time. This semantic distinction is conveyed in English by the presence/absence of a *for* or *since* adjunct PP (Dowty, 1979; Iatridou et al., 2001; Portner, 2003, 2011), as in the contrast in (14).

(14) a. Mary has lived in London for five years. Universal perfect

b. Mary has lived in London. Existential perfect

In (dialectal) Arabic the perfective form is ambiguous between the simple past tense and the existential perfect (Fassi-Fehri, 2003).

(15) šif-t-ha
see.PFV-1SG-3SGF.ACC
I saw it (F)/her. Past TENSE
I have seen it (F)/her. Present PERFECT

The universal/continuous perfect can be expressed by means of the construction shown in (16) for Sanʕāni Yemeni, Syrian, and Tunisian respectively.[6] These auxiliary forms have developed from prepositional predicates and we reflect this in our morphosyntactic gloss, with no intended consequence for their f-structure analysis.

(16) a. (ʕayn-i) la-hā θalāt iyyām
eye.SGF-1SG.GEN to-3SGF.GEN three day.PL
bi-t-ūžaʕ-ni
PROG-3SGF-hurt.IMPV-1SG.ACC
My eye has been hurting me for three days. Yemeni: Watson (1993, 80)

[6]This is not the only means whereby the universal perfect can be expressed. We leave fuller discussion of the range of possibilities, and whether they might be diachronically related to structures of the type shown in (16) for future work.

b. (muna$_i$) (ṣār)-l-a$_i$ ḥamst iyyām bi-l-ħabis
Muna become.PFV.3SGM-to-3SGF.GEN five day.PL in-DEF-prison
Muna has been in jail for five days. Syrian: Hallman (2016, 89)

c. ʕref-t-ek ʕind-i ʕam
know.PFV-1SG-2SG.ACC at-1SG.GEN year
I have known you for a year. Tunisian

To our knowledge Holes and Haddad (1984), Ingham (1994), Watson (1993) were the first to label this construction explicitly as a continuous perfect, in their description of Bahraini, Nejdi, and (SanʕānI) Yemeni, respectively. Hallman (2016) provides the first syntactic account of the construction (for Syrian), while Camilleri (2016) provides a distinct syntactic analysis for the Maltese counterpart to the Syrian construction. The details of these (different) syntactic analyses do not concern us here. Note however that Hallman's observation that the inflection on *il* (and/or the NP which may double it) must be the SUBJ of the construction, because we find the 3SGF pleonastic form in the context of weather verbs (see (17)), is relevant to what follows.

(17) **il-a** ḥamst iyyām mɣayym-**e**
to-3SGF.GEN five day.PL clouded-SGF
It's been cloudy for five days. Syrian: Hallman (2016, 83)

This construction is rather different from what we perceive a perfect construction to be from a Eurocentric viewpoint. However, Camilleri (2017) argues that the origin of the grammaticalisation of the universal PERFECT in Arabic parallels that for a number of Indo-European languages, particularly the Germanic, Romance and Celtic languages of Europe (according to Haspelmath (1998)). The Romance/Germanic perfect construction has been shown to develop out of a (transitive) possessive construction whose predicate is *have* (Trask, 1979; Vincent, 1982; Dahl, 1996; Drinka, 2017, *inter alia)*, with Heine and Kuteva (2006) coining the term 'possessive perfect' for such grammaticalisations, said to be rare crosslinguistically. So too in Arabic, where additionally, the possessive construction in Arabic is itself the result of a grammaticalisation out of a predicative prepositional construction. This is in fact parallel to the Celtic languages, which (excluding Welsh) have also grammaticalised a possessive perfect, but do not express possession via *have*.[7] In each case in fact, only one subtype of perfect is grammaticalised from the possessive construction. In Celtic, it is the existential perfect, while in Arabic, it is the universal perfect (our aim here is to account for why it is only the universal perfect that has grammaticalised in Arabic). (18) and (19) illustrate the goal possessive and location possessive schema (*to* X,Y $>$ X owns Y and *at* X,Y $>$ X owns Y) and their corresponding perfects, for Breton and Irish respectively.

[7]Ramchand et al. (1997) has an analysis of the Scottish Gaelic version of these constructions in terms of an AspP, even if no actual verbal form is present.

(18) a. Ur velo c'hlas **am** eus
a bike blue **to.**1SG is
I have a blue bike. Breton possession: Heine (1997, 60)

b. Kousket **am** eus
sleep.PAST.PTCP **to.**1SG is
I have slept. Breton existential perfect: Heine and Kuteva (2006, 175)

(19) a. Tá litir **agam**
is letter **at.**1SG
I have a letter. Irish possession: Heine and Kuteva (2006, 172)

b. Tá an bád díolta **aici**
is the boat sold.PTCP **at.**3SGF
She has sold the boat. Irish existential perfect: Harris (1991, 205)

In the light of these, now consider examples such as (20) and (21) which illustrate strikingly similar pairs for Palestinian and Tunisian respectively.

(20) a. kān la-mona tlat ulād
be.PFV.3SGM to-Mona three children
Mona had three children. Palestinian possession: Boneh and Sichel (2010, 4)

b. kān il-ha tlāt snēn min yōm imm-i
be.PFV.3SGM to-3SGF.GEN three year.PL from day mother-1SG.GEN
māt-et
die.PFV-3SGF
It had been three years since my mother died. Palestinian universal perfect

(21) a. ʕind-i kteb
at-1SG.GEN book
I have a book. Tunisian possession

b. ʕind-na ʕam tawa ma safer-ne-š il ħatta bled
at-1PL.GEN year now NEG travel.PFV-1PL-NEG ALL even country.SGF
oḫr-a
other-SGF
It's been a year now that we haven't travelled to another country.
Tunisian universal perfect

Before looking at the development of the universal perfect construction and its synchronic syntax, it should be observed that the possessive construction is itself the result of a grammaticalisation from a (prepositional) goal/locative structure. This (precursor) grammaticalisation of a possessive construction from a goal/locative structure can be visualised in terms of the development of (23) from (22). This involves the reconceptualisation of the *goal/locative* argument as a *possessor* and subsequent remapping to grammatical functions.[8]

(22) *la/ʕand* P: 'to'/'at'

		theme	goal/loc	
la/ʕand	<	**arg 1**	**arg 2**	>
		-o	-r	
		SUBJ	OBJ	

(23) *la/ʕand* V: 'have'

		poss(goal/loc)	theme		
la/ʕand	<	**arg 1**	**arg 2**	>	
		-o	-r		
		SUBJ	OBJ		
		(non-canonical)	ACC		case-marking

There is considerable evidence for the synchronic status of *la/ʕand* as a verb (and the grammatical function mapping in the 'have' construction, as shown in (23)). This includes the choice of the verb-appropriate form used for the expression of negation, various case and agreement facts, and so forth. This diachronic path (which may be the result of a grammaticalised topicalised locative structure, as suggested in Comrie (1991)) results synchronically in a set of non-canonical forms for the 'have' predicate which are referred to as *pseudo-verbs* in the literature on Arabic (Comrie, 2008). The term *pseudo-verb* is used to refer to lexemes which display a variety of verb-like functions, including those of auxiliaries, but are either not themselves originally verbal, or if verbal, with obsolete lexical meaning, or a completely grammaticalised meaning that is different from a concurrently existing lexical counterpart, and inflect very much in the same way as nouns or prepositions do. As a result, at the hypothesised origin of the grammaticalised possessive (universal) **perfect** construction in Arabic we have the pseudo-verbal forms of (23), illustrated in (20a) and (21a).

(24) Stage 0: V<SUBJ, OBJ> (diachronically derived from (22))

Camilleri (2017, under review) argues that two major ingredients must have been present within the possessive construction that subsequently grammaticalised into a universal perfect: (i) a theme argument (expressed by a NP) that was essentially a *temporal interval* of sorts; (ii) an XADJ whose function would have been

[8] We use *poss* atheoretically in (23) as a shorthand for whatever set of lexical entailments make the *goal/locative* argument more prominent in the hierarchy under this reconceptualisation.

similar to that which we hypothesised above as intrinsic to the development of the PROGRESSIVE construction, discussed in the previous section. (25) exemplifies the hypothesised route to the grammaticalistion of a possessive universal perfect.[9] (25a) is a straightforward possessive construction in which the *theme* argument is a temporal interval 'two free hours'. (25b) is a possessive construction with a (subject controlled) adjunct alongside a temporal adjunct as *theme*. This structure fulfills both these conditions. It is this structure which provides the initial stage for grammaticalisation, leading to the universal perfect construction in (25c).

(25) a. **la-ha** saʕt-ayn fāðy-īn
have-3SGF.GEN hour-DU free-PL
She has two free hours. possession

b. **la-ha**$_i$ saʕt-ayn fāðy-ah$_i$
have-3SGF.GEN hour-DU free-SGF
She has two hours, free. possession

c. **la-ha**$_i$ saʕt-ayn fāðy-ah$_i$
have-3SGF.GEN hour-DU free-SGF
She's been free for two hours. universal perfect - Kuwaiti

Taking this into account, a more accurate representation of Stage 0 is (26).

(26) Stage 0: V<SUBJ,OBJ$_{[\text{temporal interval}]}$>+XADJ where
(↑SUBJ) = (↑XADJ SUBJ)

Just as in the PROGRESSIVE construction, the clausal adjunct becomes more integrated with the structure and is incorporated into the subcategorisation frame of the predicate as an embedded clause, by argument-extension, bringing about a change at the following stage from XADJ > XCOMP, crucially only in cases where the *theme* is a *temporal interval*. We further hypothesise that this highly restricted type of theme (which expresses a temporal interval) is mapped as an +r argument, that is, as an OBJ$_\theta$, and thus there is a change involving OBJ > OBJ$_\theta$. The SUBJ of the pseudo-verb is structure-shared with the XCOMP SUBJ. We therefore identify Stage I as resulting in structures along the following lines:

(27) Stage I: V<SUBJ,OBJ$_\theta$,XCOMP> where (↑SUBJ) = (↑XCOMP SUBJ)

Some evidence for the thematically restricted nature of the GF associated with the *temporal interval* argument in (27) is the occurrence in the vernaculars of OBLique expressing temporal intervals, introduced by a *min* 'from' preposition

[9]Note that the Kuwaiti possessive constructions in (25a)-(25b) also occur synchronically with *ʕand* 'at', but this form cannot be used to give a universal perfect construction. In fact, *ʕand* 'at' as a possessive spread across the Arabic varieties much later, in some cases ousting *la* itself.

(or that P incorporated within the complementiser as is the case with *melli* in Tunisian), while the possessive construction is (naturally) limited to nominal arguments. Some relevant data is shown in (28), on the basis of which we hypothesise the Stage II also shown below.

(28) a. il-na$_i$ **min is-sani** **il-mādy-i** miš rayħ-ēn$_i$
to-1PL.GEN from DEF-year.SGF DEF-passed-SGF NEG go.ACT.PTCP-PL
hunak
there
It's been since last year that we haven't been there. Palestinian

b. ʕind-hum$_i$ ya-ʕerf-u$_i$ bʕaḍ-hom **mes-saɣra**
at-3PL.GEN 3-know.IMPV-PL each.other-3PL.GEN from.DEF-childhood
/ melli **huma** **sɣār**
/ from.COMP COP.3PL little.PL
They've known each other since they were children. Tunisian

(29) Stage II: V<SUBJ,{OBJ$_\theta$|OBL}, XCOMP> where (↑SUBJ) = (↑XCOMP SUBJ)

The next stage must have involved a loss of the SUBJ's thematicity (i.e. the development of a raising verb from a control predicate), thus leading to:

(30) Stage III: V<{OBJ$_\theta$|OBL}, XCOMP>SUBJ where (↑SUBJ) = (↑XCOMP SUBJ)

As a result, synchronically we find examples with a non-thematic subject, such as the 3SGF pleonastic SUBJ with weather verbs (as complements) illustrated in (31) for a number of vernaculars (and found across all varieties).[10]

(31) a. **il-a**$_i$ ḥamst iyyām mɣayym-e$_i$
to-3SGF.GEN five day.PL clouded-SGF
It's been cloudy for five days. Syrian: Hallman (2016, 83)

b. (as-sama) **(ṣār)-la-ha** (yum-eyn)
DEF-sky.SGF become.PFV.3SGM-to-3SGF.GEN day-DU
t-mattar (min yum-eyn)
3-rain.IMPV.SGF from day-DU
It's been raining for two days. Kuwaiti

c. el-mṭār ʕind-ha jemʕa wahi t-sob
DEF-rain.SGF at-3SGF.GEN week CONJ.3SGF.NOM 3-rain.IMPV.SGF
It's been raining for a week. Tunisian

[10]In (31b) and other examples we also find an optional *ṣār* the 3SGM perfective form of 'become' which serves purely as a morphophonological host for the clitic-prone *li/la*.

We find further developments from this stage in some vernaculars, though we do not have the space here to discuss them in any detail. In varieties including Iraqi and Maltese the perfect auxiliary can optionally exhibit default pleonastic 3SGM morphology, illustrated in (32) for Maltese, and other developments include the permissibility of what are putatively tensed COMP as well as XCOMP arguments, also illustrated by this example.[11]

(32) **Il-u** ∼ **il-i** żmien/sena li mor-t hemm
to-3SGM.GEN ∼ to-1SG.GEN time/year COMP go.PFV-1SG there
It's been a year that I went there. Maltese: Camilleri (2016, 167)

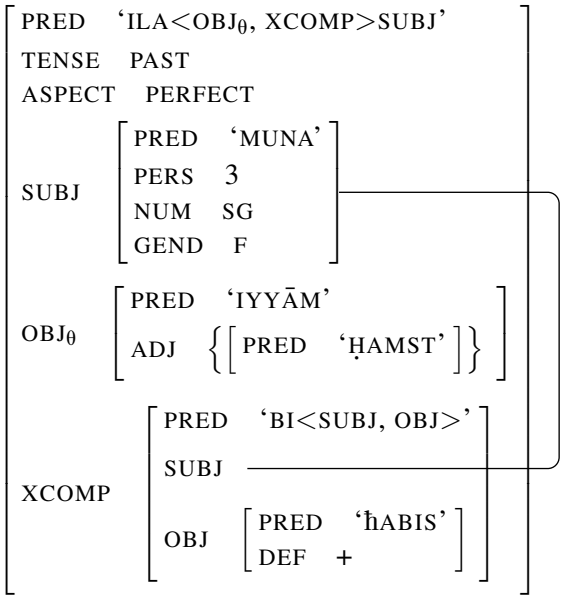
One question is whether synchronically the *il/la* element retains a PRED value or whether it is the lexical predicate that has actually become the matrix predicate, as we have argued to be the case of the PROGRESSIVE construction. Camilleri (under review) suggests that the auxiliary element does retain a PRED value in the Arabic universal perfect construction. One piece of evidence in support of this conclusion might be structures such as (33) where we seem to find the universal perfect auxiliary occurring with a COMP argument containing a pronominal co-referential with the SUBJ of the perfect auxiliary. (33) could well be an instance of copy raising which has been discussed for Arabic in Salih (1985), and accounts of Arabic within LFG in Alotaibi et al. (2013); Camilleri et al. (2014); ElSadek and Sadler (2015), and which would then provide evidence that the auxiliary within the universal perfect construction is a PRED-taking one.

(33) ʕind-**ha**$_i$ ʕam tawa [wa ma ya-ʕref-š eš
at-3SGF.GEN year now CONJ NEG 3-know.IMPV.SGM-NEG what
kāʕed sāyer-i-**l-ha**$_i$]
PROG.SGM happen.ACT.PTCP-SGM-EPENT.VWL-DAT-3SGF
It's been a year now, not knowing what's happening with her. Tunisian

Collectively, the synchronic data and the grammaticalised hypothesis render a raising structure, in association to the Universal PERFECT. We demonstrate this by providing the f-structure associated with one of Hallman's (2016) data examples from Syrian, which we analyse as a SUBJ-to-SUBJ raising structure, with the auxiliary *il*+INFL associated with an AUX PRED analysis.

(34) kān muna il-a ḥamst iyyām bi-l-ħabis
be.PFV.3SGM Muna to-3SGF.GEN five day.PL in-DEF-prison
Muna had been in prison for five days.
Syrian: Hallman (2016, 83)

[11] There are various idiosyncratic dependencies involved here, which we cannot cover here (see Camilleri (under review) for some discussion). We take the extension to a COMP to constitute a Stage IV: V$<${OBJ$_\theta$|OBL}, XCOMP|COMP$>$SUBJ where (↑SUBJ) = (↑XCOMP SUBJ).



5 Conclusion

We have discussed two instances of grammaticalisation in Arabic, using LFG to model the following mismatches:

- No change in the formal expression, but a change in function: the form *gāʕid* occurs as a lexical verb and as a featural aspectual auxiliary; and the form *li* occurs as a lexical preposition and a Aux-PRED expressing the universal perfect.
- Change in the formal expression, but no change in function: While sharing the same function of expressing an ASPECTual feature, the element *gāʕid* has a range of exponents as full, cliticised and prefixed forms in different varieties; dialects also differ in terms of whether they use *la* 'to' or *ħand* 'at' to express a universal perfect (in a common construction).
- Change in the f-structure function but no change in the c-structure: e.g. CPs introduced by *wa* 'and' can function as XADJs or XCOMPs.
- No change in the formal expression, no change in function, but change in the a- and c-structures: *la/ʕand* function as the PRED in the f-structure, yet the c-structure and a-structures differ considerably across the prepositional, possessive predicate and universal perfect uses.

We have argued that two distinct grammaticalisation paths, those leading to the development of a PROGRESSIVE construction and a possessive perfect construction expressing a universal PERFECT have both involved some sort of adjunction > embedded > matrix cline, with the constructions differing in terms of the

presence/absence of a PRED value in the latter stage. The pattern followed at the start of the grammaticalisation cline is one which has been discussed for shifts that have taken place diachronically in the development of Indo-European languages (Kiparsky, 1995), but which had never been discussed for Arabic. This commonality suggests that there may be core diachronic processes of syntactic reanalysis, structural shifts and grammaticalisations which are just as typologically widespread as instances of the lexical > grammatical item type of grammaticalisation. Further comparative work on the family of closely related Semitic languages has the potential to cast further light on the occurrence of this diachronic process of structural change.

References

Absi, Samir Abu and A. Sinaud. 1968. *Basic Chad Arabic. I The Pre-speech phase; II Comprehension texts; III The active phase*. Bloomington: Indiana University.

Agius, Dionisius A. and Amir Harrak. 1987. Auxiliary particles preceding the imperfective aspect in Arabic dialects. *Arabica* 34(2):164–180.

Akkuş, Faruk. 2016. The development of the present copula in Arabic. Ms, Yale University.

Almashaqba, Bassil Mohammad et al. 2015. *The phonology and morphology of Wadi Ramm Arabic*. Ph.D. thesis, University of Salford.

Alotaibi, Yasir, Muhammad Alzaidi, Maris Camilleri, Shaimaa ElSadek, and Louisa Sadler. 2013. Psychological Predicates and Verbal Complementation in Arabic. In M. Butt and T. H. King, eds., *Proceedings of the LFG13 Conference*, pages 6–26. Stanford, CA: CSLI Publications.

Badawi, El Said, Michael Carter, and Adrian Gully. 2003. *Modern Written Arabic: A Comprehensive Grammar*. London: Routledge.

Barron, Julia, Miriam Butt, and Tracy Holloway King. 1997. Lfg and the history of raising verbs. *LFG97, University of California at San Diego* .

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

Borg, Albert J. 1988. *Ilsienna*. Malta: Has Sajjied.

Börjars, Kersti, Khawla Ghadgouf, and John Payne. 2016. Aspectual object marking in Libyan Arabic. In D. Arnold, M. Butt, B. Crysmann, T. H. King, and S. Müller, eds., *The Proceedings of the Joint 2016 conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar*, pages 125–139. Stanford, CA: CSLI Publications.

Börjars, Kersti, Pauline Harries, and Nigel Vincent. 2016. Growing syntax: The development of a dp in north germanic. *Language* 92(1):e1–e37.

Brustad, Kristen. 2000. *The Syntax of Spoken Arabic*. Washington, DC: Georgetown University of Press.

Butt, Miriam. 1996. *The Structure of Complex Predicates in Urdu*. Dissertations in Linguistics. Stanford, CA: CSLI Publications. Revised and corrected version of 1993 Stanford University dissertation.

Butt, Miriam and Wilhelm Geuder. 2003. Light verbs in Urdu and grammaticalization. In K. v. H. Regina Eckardt and C. Schwarze, eds., *Diachronic Semantics from Different Points of View*, pages 295–349. Berlin: Mouton de Gruyter.

Butt, Miriam and Aditi Lahiri. 2002. Historical stability vs. historical change. *Unpublished Ms. http://ling. uni-konstanz. de/pages/home/butt* .

Butt, Miriam and Aditi Lahiri. 2013. Diachronic pertinacity of light verbs. *Lingua* 135:7–29.

Bybee, Joan L and Östen Dahl. 1989. *The creation of tense and aspect systems in the languages of the world*. John Benjamins.

Bybee, Joan L., Revere Perkins, and William Pagliuca. 1994a. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.

Bybee, Joan L., R. Perkins, and W. Pagliuca. 1994b. *The evolution of grammar: Tense, aspect and modality in the languages of the world*. Chicago: Chicago University Press.

Camilleri, Maris. 2016. *Temporal and aspectual auxiliaries in Maltese*. Ph.D. thesis, University of Essex.

Camilleri, Maris. 2017. Possessive perfect constructions in Arabic. Presentation at the North Atlantic Conference for Afroasiatic Linguistics, 45, Leiden, Holland, 10th June, 2017.

Camilleri, Maris. under review. The syntax and development of the universal perfect in Arabic.

Camilleri, Maris, Shaimaa ElSadek, and Louisa Sadler. 2014. Perceptual Reports in (dialects of) Arabic. In M. Butt and T. H. King, eds., *Proceedings of the LFG14 Conference*, pages 179–199. Stanford, CA: CSLI Publications.

Camilleri, Maris and Louisa Sadler. 2017. Posture Verbs and Aspect: A View from Vernacular Arabic. In M. Butt and T. H. King, eds., *Proceedings of the LFG'17 Conference, University of Konstanz*, pages 167–187. Stanford, CA: CSLI Publications.

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

Comrie, Bernard. 2008. Pseudoverb. In K. Versteegh, M. Eid, A. Elgibali, M. Woidich, and A. Zaborski, eds., *Encyclopaedia of Arabic Language and Linguistics*, vol. III, pages 739–740. Leiden: Brill: Brill.

Cuvalay, Martine. 1991. The expression of durativity in Arabic. *The Arabist, Budapest studies in Arabic* 3-4:143–158. Proceedings of the colloquium on Arabic grammar, Budapest.

Dahl, Östen. 1996. *Tense and Aspect Systems*. Oxford: Basil Blackwell.

Deo, Ashwini. 2015. The semantic and pragmatic underpinnings of grammaticalization paths: The progressive to imperfective shift. *Semantics and Pragmatics* 8:1–52.

Dowty, David R. 1979. *Word Meaning and Montague Grammar: The Semantics of Verbs and Times in Generative Semantics and in Montague's PTQ*, vol. 7 of *Synthese Language Library*. Dordrecht: D. Reidel.

Drinka, Bridget. 2017. *Language Contact in Europe: The periphrastic perfect through history*. Cambridge: Cambridge University Press.

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

Falk, Yehuda. 2008. Functional relations in the English auxiliary system. *Linguistics* 46(4):861–889.

Fassi-Fehri, Abdelkader. 2003. Arabic perfect and temporal adverbs. In A. Alexiadou, M. Rathert, and A. von Stechow, eds., *Perfect explorations*, pages 69–100. Berlin: Walter de Gruyter.

Fischer, Olga. 2007. *Morphosyntactic change: Functional and formal perspectives*, vol. 2. Oxford: Oxford University Press.

Ghadgoud, Khawla. 2018. *Negative patterns in Libyan Arabic and Modern Arabic varieties*. Ph.D. thesis, University of Manchester.

Hallman, Peter. 2016. The Universal Perfect in Syrian Arabic. *Brill's Annual of Afroasiatic Languages and Linguistics* 8(1):76–100.

Harrell, Richard S. 1962. *A Short Reference Grammar of Moroccan Arabic*. Washington: Georgetown University Press.

Harris, John. 1991. Conservatism versus substratal transfer in Irish English. *Dialects of English:*

Studies in grammatical variation pages 191–212.

Haspelmath, Martin. 1998. How young is Standard Average European? *Language Sciences* 20(3):271–287.

Heath, Jeffrey. 2013. *Jewish and Muslim dialects of Moroccan Arabic*. New York: Routledge.

Heine, Bernd. 1993. *Auxiliaries: Cognitive forces and grammaticalization*. Oxford: Oxford University Press.

Heine, Bernd. 1997. *Possession: Sources, forces and grammaticalization*. Cambridge: Cambridge University Press.

Heine, Bernd and Tania Kuteva. 2002. *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press.

Heine, Bernd and Tania Kuteva. 2006. *The changing languages of Europe*. Oxford University Press.

Heine, Bernd and Mechthild Reh. 1984. *Grammaticalization and reanalysis in African languages*. Buske Helmet Verlag Gmbh.

Henkin, Roni. 1992. The three faces of the Arabic participle in Negev Bedouin dialects: continuous, resultative, and evidential. *Bulletin of the school of Oriental and African Studies* 55(3):433–444.

Holes, Clive and S Haddad. 1984. *Colloquial Arabic of the Gulf and Saudi Arabia*. London: Routledge & Kegan Paul.

Hopper, Paul J and Elizabeth Closs Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press.

Iatridou, Sabrina, Eleni Agnastopoulou, and Rouma Izvorski. 2001. Observations about the Form and Meaning of the Perfect. In M. Kenstowicz, ed., *Ken Hale: A Life in Language*, page 1890238. Boston, MA: MIT Press.

Ingham, Bruce. 1994. *Najdi Arabic: Central Arabian*, vol. 1. Amsterdam: John Benjamins.

Jarad, Najib Ismail. 2015. From bodily posture to progressive aspect marker. *Lingua Posnaniensis* 57(1):89–111.

Johnstone, Tomas M. 1967. *Eastern Arabian dialect studies*, vol. 17. Oxford: Oxford University Press.

Kiparsky, Paul. 1995. Indo-European origins of Germanic syntax. In I. Roberts and A. Battye, eds., *Clause structure and language change*, pages 140–167. Oxford: Oxford University Press.

Kuteva, Tania. 1999. On 'sit'/'stand'/'lie' auxiliation. *Linguistics* 37/2:191–213.

Lehmann, Christian. 1995. *Thoughts on grammaticalization*. Munich: Lincom Europa.

McCawley, James D. 1971. Tense and time reference in English. In C. Fillmore and D. T. Langendoen, eds., *Studies in Linguistic Semantics*, pages 97–113. New York: Holt Rinehart.

McCawley, James D. 1981. Notes on the English present perfect. *Australian Journal of Linguistics* 1(1):81–90.

McCoard, Robert W. 1978. *The English perfect: Tense-choice and pragmatic inferences*. Amsterdam: North Holland Publishing.

McNeil, K. 2017. *Fī* ('in') as a marker of the progressive aspect in Tunisian Arabic. In V. Ritt-Benmimoun, ed., *Tunisian and Libyan Dialects: Common trends - Recent developments*, pages 161–190. Zaragoza: Universidad de Zaragoza Press.

Meillet, Antoine. 1912. L'évolution des formes grammaticales. *Scientia* 12(26.6):130–148.

Mion, Giuliano. 2004. Osservazioni sul sistema verbale dell'Arabo di Tunisi. *Rivista degli Studi Orientali* 78(1/2):243–255.

Mitchell, Terence F and Shāhir al Hassan. 1994. *Modality, Mood, and Aspect in Spoken Arabic: With Special Reference to Egypt and the Levant*. New York: Kegan Paul.

Mughazy, Mustafa. 2005. Rethinking lexical aspect in Egyptian Arabic. In *Perspectives on Arabic Linguistics XVII-XVIII: Papers from the Seventeenth and Eighteenth Annual Symposia on Arabic*

Linguistics, pages 133–172. John Benjamins.

Pallottino, Margherita. 2016. "feš taqra?" what are you reading?: Prepositional objects in Tunisian Arabic. *Brill's Annual of Afroasiatic Languages and Linguistics* 8:286–312.

Pereira, Christoph. 2008. *Le parler arabe de Tripoli (Libye): Phonologie, morphosyntaxe et categories grammaticales*. Ph.D. thesis, INALCO, Paris.

Persson, Maria. 2009a. Circumstantial Qualifiers in Gulf Arabic Dialects. In B. Isaksson, H. Kammensjö, and M. Persson, eds., *Circumstantial qualifiers in Semitic: The case of Arabic and Hebrew*, pages 206–XXX. Otto Harrassowitz Verlag.

Persson, Maria. 2009b. Circumstantial Qualifiers in Gulf Arabic Dialects. In B. Isaksson, H. Kammensjö, and M. Persson, eds., *Circumstantial qualifiers in Semitic: The case of Arabic and Hebrew*, pages 206–248. Otto Harrassowitz Verlag.

Persson, Maria. 2013. Asyndetic clause combining in Gulf Arabic dialects: Auxiliary, adverbial and discourse functions. *Zeitschrift für Arabische Linguistik* 57:5–39.

Portner, Paul. 2003. The (temporal) semantics and (modal) pragmatics of the perfect. *Linguistics and Philosophy* 26(4):459–510.

Portner, Paul. 2011. Perfect and progressive. In K. von Heusinger, C. Maienborn, and P. Portner, eds., *Semantics: An international handbook of meaning*, pages 1217–1261. Berlin: DeGruyter.

Procházka, Stephan and Ismail Batan. 2015. The functions of active participles in Šāwi bedouin dialects. In *Arabic Varieties: Far and Wide: Proceedings of the 11th International Conference of AIDA*, pages 457–466.

Ramchand, Gillian et al. 1997. *Aspect and predication: The semantics of argument structure*. Oxford University Press.

Rubin, Aaron D. 2005. *Studies in Semitic Grammaticalization*. Harvard Semitic Studies, Vol 57. Winona Lake, IN: Eisenbrauns.

Ryding, Karin. 2005. *A Reference Grammar of Modern Standard Arabic*. Cambridge: Cambridge University Press.

Salih, Mahmud. 1985. *Aspects of Clause Structure in Standard Arabic: A Study in Relational Grammar*. Ph.D. thesis, SUNY at Buffalo.

Schwarze, Christoph. 2001. Representation and variation: On the development of romance auxiliary syntax. *Time over matter. Diachronic perspectives on morphosyntax* pages 143–172.

Seiss, Melanie. 2009. On the difference between auxiliaries, serial verbs and light verbs. In M. Butt and T. H. King, eds., *Proceedings of the LFG09 Conference*, pages 501–519. CSLI Publications.

Souag, Mostafa Lameen. 2006. Explorations in the syntactic cartography of Algerian Arabic.

Trask, Robert L. 1979. On the origins of ergativity. In F. Plank, ed., *Ergativity: Towards a theory of grammatical relations*, pages 385–404. London: Academic Press.

Vincent, Nigel. 1982. The development of the auxiliaries *habere* and *esse* in Romance. *Studies in the Romance verb* pages 71–96.

Vincent, Nigel. 2001. LFG as a model of syntactic change. In Miriam Butt and Tracy Holloway King, ed., *Time over Matter: Diachronic Perspectives on Morphosyntax*, pages 1–42. Stanford, CA: CSLI Publications.

Vincent, Nigel and Kersti Börjars. 2010. Grammaticalization and models of language. In E. Traugott and G. Trousdale, eds., *Gradience, gradualness and grammaticalization*, pages 279–300. Amsterdam/Philadelphia: John Banjamins.

Watson, Janet C. E. 1993. *A Syntax of Ṣanānī Arabic*, vol. 13. Berlin: Otto Harrassowitz Verlag.

Woidich, Manfred. 1995. Some cases of grammaticalization in Egyptian Arabic. In J. Cremona and C. Holes, eds., *The Proceedings of AIDA II (1995)*, pages 259–268.

Dative Arguments in Psychological Predicates in Spanish

Paloma Carretero García
University of Essex

Proceedings of the LFG'18 Conference

University of Vienna

Miriam Butt, Tracy Holloway King (Editors)

2018

CSLI Publications

pages 150–170

http://csli-publications.stanford.edu/LFG/2018

Keywords: psychological predicates, argument structure, datives, weak pronouns, Spanish

Carretero García, Paloma. 2018. Dative Arguments in Psychological Predicates in Spanish. In Butt, Miriam, & King, Tracy Holloway (Eds.), *Proceedings of the LFG'18 Conference, University of Vienna*, 150–170. Stanford, CA: CSLI Publications.

Abstract

This paper explores the behaviour of dative arguments in the context of psychological predicates in Spanish. We focus on predicates that require a dative experiencer and a nominative stimulus. These constructions have an obligatory dative weak pronoun but also optionally allow a doubled dative NP. We are concerned with what the status of the dative is and why the unmarked order of the sentence is DAT NP + V + DAT PRN + NOM NP. We firstly examine the possibility that the dative NP is the subject but will argue through testing that the NOM NP is SUBJ. We will then propose to treat the dative argument as OBJ_θ. Finally we claim that the unexpected order stems from a mismatch between thematic and functional hierarchies and will analyse the position of the DAT NP as WEAK FOCUS, whose properties will be described in depth in the last sections of the paper.

1 Introduction

1.1 Psychological predicates

Psychological predicates are predicates whose argument structure involves an experiencer and a theme or stimulus/cause. They typically involve concepts such as *fear, enjoy, hate* or *frighten, worry, irritate....* Their arguments map differently depending on the type of predicate and they have traditionally been grouped according to their mapping pattern (Belleti & Rizzi, 1988).

In English, for instance, there is a FEAR group with the experiencer as SUBJ and a FRIGHTEN category where the experiencer is OBJ:

(1) a. I fear spiders.

b. Spiders frighten me.

Both predicates take, in principle, the same thematic roles, but differ in the way they map those roles into syntactic arguments.[1]

[†]I am indebted to Maris Camilleri for extremely valuable comments, which have improved the analysis and presentation. I thank Louisa Sadler for comments on an earlier version of this paper and Doug Arnold and the LFG research group at Essex for torturing themselves with the tests. Many thanks to the audience at the LFG18 Conference in Vienna for great discussion and comments and to the editors and reviewers for further suggestions and observations.

[1]Based on the assumption that *theme* can be considered general enough, even though more specifically we have *theme, stimulus, cause*. This is more clearly seen with *frighten*-type predicates where we get a range of readings/roles which can be more or less causative since the subject can also get an agentive reading (Grimshaw, 1990).

2 Spanish psychological predicates

Spanish shows four classes of psychological predicates, based on their subcategorisation patterns:[2]

1. Verbs that subcategorise for an accusative experiencer: *aburrir* 'to bore', *molestar* 'to disturb', *ofender* 'to offend'...:[3]

   (2) a. Los niños enfadan a sus madres
   The.M.PL child.M.PL anger.3PL.PRS ACC their mother.PL

   'Children anger their mothers.'

   b. Los niños las enfadan
   The.M.PL child.M.PL 3.F.ACC.PL anger.3PL.PRS

   'The boys anger them'

2. Verbs that behave like the English *fear*, with the experiencer as subject and the stimulus as an object NP (PP or a complement clause): *odiar* 'to hate', *temer* 'to fear', *adorar* 'to adore', *creer* 'to believe'...

   (3) Laura odia las películas románticas
   Laura hate.PRS.3SG the.F.PL film.PL romantic.F.PL

   'Laura hates romantic films.'

3. Reflexive verbs. The pattern for this group consists of an experiencer subject, a reflexive pronoun and an optional phrase such as PP. This class includes reflexive verbs that express a feeling undergone by the experiencer: *aburrirse* 'to get bored', *enfadarse* ' to get angry', *alegrarse* 'to feel happy'...

   (4) Los niños se aburren (en clase)
   The.M.PL child.M.PL REFL bore.PRS.3PL in class

   'Children get bored in class.'

[2]Vogel & Villada (1999) describe five different patterns but two of them become identical in regards to syntactic pattern, which is why we choose to reduce the grouping to four.

[3]Most of the verbs in this group can take an ACC or DAT argument since both patterns are possible. There seems to be a slight change of meaning depending on the pattern - related with volition of the SUBJ, which could be distinguished by analysing it as *cause* -with the ACC- or *stimulus* - with the DAT:

(i) A los niños les ofende la mentira
DAT the.M.PL child.M.PL DAT.3PL offend.3SG.PL the.F.SG lie
'Lies offend children = Children find lies offensive.'

Note that in configurations such as (2a) and (i), one same element *a* marks a complement as DAT or ACC. This will be further discussed in 5.1, but we can see the contrast if we compare the weak pronouns that refer to that complement, an ACC *las* pronoun in (2b) and a DAT *le* pronoun in (i). Syntactically, the configuration in (i) is the same as verbs in Type 4, which is why we do not consider them different groups.

4. This group comprises verbs that require a dative experiencer and the presence of a weak pronoun is obligatory: *gustar* 'to like', *doler* 'to hurt', *fascinar* 'to fascinate', *interesar* ' to interest'...

   (5) A Laura$_i$ le$_i$ gustan las fresas
   DAT Laura 3SG.DAT please.PRS.3PL the.F.PL strawberry.PL

   'Strawberries are pleasing to Laura.'
   ='Laura likes strawberries.'

   A schematic representation of this pattern is found in (6) below:

(6)

(DATIVE NP)	DATIVE CLITIC	V	NOMINATIVE NP
Experiencer			*Stimulus*

This paper will explore the properties of Type 4 verbs as in (5). The main questions we aim to answer are: (i) what the appropriate GFs of the different participants (experiencer and stimulus) are; (ii) how to characterise the dative argument; (iii) how to deal with doubling and the obligatory presence of the weak pronoun, and (iv) how we can account for the "unexpected" ordering. These issues will be further explored in the following sections.

3 Type 4 psychological predicates *vs.* other predicates that take a dative argument

It is also relevant to place the psychological predicates we are discussing in the context of other predicates that take a dative argument. Dative arguments -in many cases in the form of a weak pronoun- can appear in the context of all types of verbs and have many different meanings.[4] Dative arguments are found in ditransitive constructions with 'give'-type verbs:

(7) a. Juan dio un regalo a Laura
Juan give.PST.3SG a gift DAT Laura

'Juan gave a gift lo Laura'

b. Juan le dio un regalo (a Laura)
Juan 3.SG.DAT give.PST.3SG a gift DAT Laura

'Juan gave a gift lo Laura'

[4]See Cuervo (2003, pp.29-30) for a list that includes both selected and non-selected datives.

We see in (7a) that we can have the dative noun phrase without the dative weak pronoun, which is something that is not allowed with the psychological predicates at hand:

(8) *A Laura gustan las fresas
DAT Laura please.PRS.3PL the.F.PL strawberry.PL

'Strawberries are pleasing to Laura.'
='Laura likes strawberries.'

We can also front the noun phrase, which will result in a configuration that is identical to (5), repeated below as (9):

(9) A Laura$_i$ le$_i$ gustan las fresas
DAT Laura 3SG.DAT please.PRS.3PL the.F.PL strawberry.PL

'Strawberries are pleasing to Laura.'
='Laura likes strawberries.'

(10) A Laura le dio Juan un regalo
DAT Laura 3.SG.DAT give.PST.3SG Juan a gift

'Juan gave a gift lo Laura'

However, (10) is a case of Clitic Left Dislocation (CLLD), that triggers pronominal reduplication and is to be analysed as a topical element. We believe these configurations, even though similar in many aspects, are different in the issues they pose. Most importantly, we believe that the unmarked order of (9) is not the same as with other type of predicates that take dative complements, as the unmarked order for those is found in (7), even if the elements can also be fronted as in (10). This is something this paper will account for in later sections.[5]

4 Subject Issues

4.1 What's the subject?

A first obvious question to answer is to decide what the subject is. However, judging by the array of possible answers to the question, this is not a trivial matter. There have been different

[5]I thank one of the reviewers for pointing out that these predicates need to be contrasted with other predicates that also take datives. We will not be able to examine this in any further detail in this paper due to space constraints. However, it is worth mentioning that even though we find psychological predicates different enough to merit a separate treatment -or at least the problems they raise differ from other predicates, e.g. status of the subject-, the natural next step is to somehow try to extend this analysis to other predicates, especially when they interact with the pronoun *se*, or with non-selected datives as below, which seem to closely correlate with psychological predicates in regards to unmarked word order :

(ii) A Laura se le cayó un plato
DAT Laura REFL 3.SG.DAT fall.PST.3SG a plate
'Laura dropped a plate.'

proposals that we are summarising below:

Alarcos Llorach (1994) argues *a*-introduced phrases are PPs and they cannot be subjects at all so the *stimulus* NP is the SUBJ for him.

Mendívil Giró (2002) proposes a system similar to the one shown by languages with ERGATIVE and ABSOLUTIVE case. He claims psychological predicates are to be described as displaying *"lexically conditioned partial ergativity"*. According to him, the dative experiencer would be analysed as an ergative subject whereas the postposed argument would be analysed as an absolutive direct object.

Based on Zaenen et al. (1985)'s treatment of Icelandic passive constructions, Fernández Soriano (1999) or Masullo (1992) argue for a quirky dative case in Spanish: Masullo (1992) admits some differences with Icelandic and proposes an approach based on Belleti & Rizzi (1988) and points out these predicates are very similar to unaccusatives. He claims that these constituents raise to Spec (IP) and the nominative case is assigned to the postverbal NP via government rather than by specifier-head agreement. Fernández Soriano (1999) claims that this quirky case is morphological and inherent and it allows the phrase bearing it to move to case-marked positions. This is why it can move and merge as external argument where it can satisfy the EPP condition.

Landau (2010) claims that these verbs denote locative relations, the dative is actually an oblique with a null preposition and can be analysed through an extended version of locative inversion.

Cuervo (2010) proposes a specific analysis for psychological verbs which involves a specialised applicative head: "The verbal root combines with a stative *v* and takes the DP as its specifier. The experiencer is added to the structure not as an argument of the verb, but as an extra, external argument, licensed by a specialised head, the applicative Appl. The applicative head licenses the experiencer as its specifier and relates it to the *v*P it takes as a complement."(Cuervo, 2010, p. 29).

Alsina (1996) and Vanhoe (2002) claim that the dative experiencer bears an objective function.

4.2 Subjecthood tests

In order to shed some light on the subjecthood issue, we will test both the dative and nominative noun phrases to establish which one could be analysed as subject. These tests are adapted from Vogel & Villada (1999) and are not necessarily novel but will be helpful in determining how to analyse the different participants. We will use the sentence in (11) to apply the different tests. We are using animate, human participants to avoid possible interference and one singular phrase and one plural to make sure the agreement interactions are clearly noted:

(11) A Laura$_i$ le$_i$ gustan los alumnos
DAT Laura 3SG.DAT please.PRS.3PL the.M.PL student.PL
experiencer *stimulus*

'Students are pleasing to Laura'
='Laura likes the students.'

Clitics aside, and generally speaking, Spanish is SVO[6] so SUBJ appears as the first NP in an **unmarked** finite clause. This could mean that the first noun phrase in (11) is subject, in this case that would be a dative.

However, SUBJ requires NOM subject pronouns upon pronominal substitution and in (11) only *los alumnos* can be replaced by a NOM subject pronoun: *ellos* 'they'. Furthermore, the verb which typically agrees with the subject, is agreeing with the stimulus NP in person and number, and this is consistently the case if we change the person and number of the participants:

(12) a. A Laura le gust**as** **tú**
DAT Laura 3SG.DAT please.PRS.2SG 2SG.NOM

'You are pleasing to Laura'
= 'Laura likes you.'

b. A Laura le gust**o** **yo**
DAT Laura 3SG.DAT please.PRS.1SG 1SG.NOM

'I am pleasing to Laura'
= 'Laura likes me.'

This could now indicate that the stimulus participant that appears at the end of the sentence is SUBJ. We will test this further below. Namely, we will test the behaviour of these participants in control, raising and causative constructions. We will examine their binding properties in reflexive configurations and their behaviour in passive alternations. We will lastly assess their ability to be 'dropped' as this is a typical feature of subjects in Spanish.

4.2.1 Control

We now test the ability for the NPs involved in psychological verbs constructions to be controlled arguments:

(13) Los alumnos$_i$ quieren$_i$ gustarle a Laura
The.M.PL student.M.PL want.PRS.3PL please.INF=3.DAT.SG DAT Laura
stimulus *experiencer*

'The students want to be pleasing to Laura.'
='The students want Laura to like them.'

[6]See Solà i Pujols (1992), Vallduví (1993) or Vallduví (2002) for claims that we have VOS or VXS order as standard. This is not a central issue for the paper but it is worth noting that SVO order is not unanimously accepted. We thank an anonymous reviewer for pointing this out.

(14) *A Laura$_i$ (?le) quiere$_i$ gustar los alumnos
DAT Laura 3SG.DAT want.PRS.3SG please.INF the.M.PL student.M.PL
experiencer *stimulus*

(intended) 'Laura wants students to be pleasing to her.'
=(*intended*) 'Laura wants to like students.'

In (13) we see that the stimulus NP can be subject of *querer* and is therefore controlling the subject of the XCOMP psychological predicate. We cannot do the same with the dative phrase as seen in (14). In order to get the intended reading we would need to construct a sentence such as the one below in (15), but that would imply making Laura the stimulus and the students the experiencer:

(15) Laura quiere gustarles a los alumnos
Laura want.3SG.PRS like.INF=DAT.3PL DAT the.M.PL student.PL

'Laura wants to be pleasing to the students.' ='Laura wants the students to like her.'

If we want the subject of the control predicate to be the experiencer of the psychological predicate, we need to introduce a finite embedded clause, but the relationship between the two is of a different nature:

(16) Laura quiere que le gusten los alumnos (a
Laura want.PRS.3SG that 3.DAT.SG please.PRS.SBJV.3PL the.M.PL student.M.PL DAT
ella)
her

'Laura wants students to be pleasing to her.'
='Laura wants to like students.'

Based on the control tests, the stimulus participant is more likely to be SUBJECT. We now move on to raising tests.

4.2.2 Raising

In raising constructions, the SUBJECT of the embedded predicate "*raises*" to the subject position of the matrix clause. The relevant description for Spanish raising constructions is the following:

(17) a. Juan parece amable
Juan seem.PRS.3SG kind

'Juan seems kind.'

b. '*seem* <XCOMP > SUBJ'
(↑ SUBJ) = (↑ XCOMP SUBJ)

(18) a. Juan parece ser amable
Juan seem.PRS.3SG be.INF kind

'Juan seems to be kind.'

b. '*seem* <XCOMP > SUBJ'
(↑ SUBJ) = (↑ XCOMP SUBJ)

It follows, then, that the participant that can appear in the matrix clause is to be considered the subject of the psychological predicate:

(19) Los alumnos parecen gustarle a Laura
The.M.PL student.M.PL seem.PRS.3PL please.INF=(3.DAT.SG) DAT Laura

'Students seem to be pleasing to Laura.' = 'Laura seems to like the students.'

Again, it looks like the stimulus participant can do that, which points at the likelihood that it is the SUBJ. It is worth noting that (19) is not a very natural sounding sentence. The preferred alternative would be (20):

(20) A Laura parecen gustar=le los alumnos
DAT Laura seem.PRS.3PL please.INF=3.DAT.SG the.M.PL student.M.PL

'Students seem to be pleasing to Laura.' = 'Laura seems to like the students.'

Consider, however, that even though (20) shows the dative experiencer right in front of the raising predicate, *parecer* agrees with the students in number and person. We can also change the person to show this more clearly:

(21) A Laura parec**éis** gustar=le **vosotros**
DAT Laura seem.PRS.2PL please.INF=3SG.DAT 2PL.NOM

'You guys seem to be pleasing to Laura.' = 'Laura seems to like you guys.'

This does not seem to support a view of treating the experiencer as SUBJECT but rather, it points out c-structural tendencies for the dative experiencer to appear first in the sentence, and we see no f-structure differences between (19) and (20).

We can therefore conclude that the raising test favours the treatment of the stimulus as SUBJ.

4.2.3 Causatives

Vogel & Villada (1999) believe that the behaviour of the participants as possible subjects of the causative predicate *hacer* 'to make' provides data about both their syntactic and semantic properties. However, psychological verbs do not admit embedding when the participants are tested as agents of the causative verb. Neither Stimulus NPs or Experiencer NPs can be the agent of causation and the only way to convey such readings would be by introducing another clause with an added agent:

(22) *Los alumnos hicieron a Laura gustarle
The.M.PL student.M.PL make.PST.3PL DAT Laura like.INF=(DAT.SG)

'The students made Laura like them.'

(23) a. ?Los alumnos hicieron que a Laura le gustaran
The.M.PL student.M.PL make.PST.3PL that DAT Laura DAT.3SG like.PST.SBJV.3PL
(ellos / los alumnos)
(NOM.3.PL / the.M.PL student.M.PL)

'The students made Laura like them.'

(24) Los alumnos hicieron que a Laura le gustasen
The.M.PL student.M.PL make.PST.3PL that DAT Laura DAT.3SG like.PST.SBJV.3PL
las fresas
the.F.PL strawberry.PL

'The students made Laura like strawberries.'

This implies altering the sentence too much, so we consider it is not applicable for the task at hand due to obvious semantic restrictions so we will therefore discard it as a subjecthood test.

4.2.4 Binding properties in reflexive constructions

In reflexive constructions we find one single NOM NP argument that would have both the roles of experiencer and stimulus:

(25) Los alumnos se gustan
The.M.PL student.M.PL REFL like.PRS.3PL

'The students like themselves/ each other.'

We could not possibly have a similar sentence with the dative binding to the reflexive:

(26) a. *A Laura se gusta
(DAT) Laura REFL like.PRS.3SG

'Laura likes herself.'

b. Laura se gusta
Laura REFL like.PRS.3SG

'Laura likes herself.'

This test can easily be considered borderline and could merit being discarded. However, it clearly shows that a dative argument does not bind to the reflexive. The only nominative participant of a psychological predicate construction is the stimulus, which again seems to argue for its treatment as subject, even if it is not the most solid test.

4.2.5 Passivisation

Constructions with psychological predicates do not admit a passive alternation, since they do not have an agentive argument that can be suppressed.

(27) a. *Los alumnos son gustados por Laura
The.M.PL student.M.PL be.PRS.3PL like.M.PL.PART by Laura
stimulus *experiencer*

'Students are liked by Laura.'

b. *A Laura es gustada por los alumnos
DAT Laura be.PRS.3SG like.PSTPART.F by the.M.PL student.PL
experiencer *stimulus*

'To Laura is liked by the students.'

This rules out the possibility of applying a passive test to determine the subject in the constructions we are examining.

4.2.6 Ability to 'pro-drop'

It is a well known feature of Spanish that it is a language that has subject drop, so the subject of the sentence does not have to be overtly realised. This is exemplified in (28) below:

(28) a. Laura llora
Laura cry.PRS.3SG

'Laura cries.'

b. Llora
cry.PRS.3SG

'He/she cries.'

We test the ability of the participants in psychological predicates constructions to be dropped without altering the meaning of the sentence:

(29) A Laura le gustan
DAT Laura 3SG.DAT like.PRS.3PL

'They please Laura'
='Laura likes them.' [dropped stimulus]

(30) Los alumnos gustan
the.M.PL student.M.PL like.PRS.3PL

'Students please ≈ students are liked.' [dropped experiencer]

We see in (29) and (30) that both the stimulus and the experiencer can be dropped. However, the original meaning is only retained in (29). We believe that in (30) we have a different lexical operation that turns the verb into a one-place predicate with some sort of passive reading. This test indicates, therefore, that the stimulus participant is the likely subject in these constructions with psychological predicates.

4.2.7 Summary of results

Even though some tests cannot be successfully applied and there were some semantic restrictions, the results as summarised in (31) indicate that the stimulus participant has to be mapped as SUBJECT.

(31)

	CONTROL	RAISING	CAUSATIVE	BINDING	PASSIVE	PRO-DROP
CAUSE/ STIMULUS NP	✓	✓	N/A	✓?	N/A	✓
EXPERIENCER NP	X	X	N/A	X	N/A	X

Once we have argued that the stimulus is SUBJECT, we move on now to discuss how to best characterise the experiencer dative argument. We will consider treatments as OBLIQUE, OBJECT or OBJECT_θ and will ultimately argue that OBJECT_θ is the best fit for the properties that this participant displays.

5 Characterisation of the experiencer

5.1 Experiencer as OBL

We believe that 'a' is not a preposition but a grammatical marker as seen in object constructions:

- It marks human/animate objects as ACC:

  (32) a. Peino el pelo
  Comb.PRS.1SG the.M.SG hair

  'I comb the hair.'

  b. Peino **a** Marta
  Comb.PRS.1SG ACC Laura

  lit.'I comb Laura'
  'I comb Laura's hair.'

- It also marks the beneficiary/recipient in double object constructions as DAT

  (33) Doy un regalo **a** Marta
  give.PRS.1SG a gift DAT Marta

  'I give a present to Laura.'

We therefore consider the *a*-introduced phrase as a dative NP and believe it is better treated as an objective function.

5.2 Experiencer as OBJ

Vanhoe (2002) treats the experiencer dative argument as the primary object, mainly based on the idea that secondary objects need to appear with another object.

Alsina (1996) does not distinguish between types of objects and explains different case assignment through the following convention:

(34) "*Case Assignment Convention*:

a A direct function (one that has the feature [obl -]) must take the marked feature value [DAT +] if it is mapped onto an argument that is either thematically a goal or more prominent than another argument expressed as a nondative function and if it is not the expression of the external argument.

b All other direct functions take the default feature value [DAT -]".

(Alsina, 1996, p. 175)

Since we have seen that dative arguments cannot become subjects of passive constructions, but accusative objects can, he proposes to constrain the dative to ensure it does not appear as subject:

(35) "*Nondative Subject Constraint*:
*[[SUBJ +] [DAT +]]" (Alsina, 1996, p. 179)

Even though Alsina (1996)'s analysis would certainly work, we believe both objects have enough differences to merit distinct grammatical functions and we would not need any specific constraints to prevent the dative from being mapped as a subject since that follows from the properties of the OBJECT$_\theta$ as we will see in the next section.

5.3 Experiencer as OBJ$_\theta$

Following Kibort (2007), Kibort (2008) and Kibort (2013), we have a template with available slots as follows:

(36) $\langle$arg$_1$ arg$_2$ arg$_3$ arg$_4$ ... arg$_n\rangle$
[-o/-r] [-r] [+o] [-o] [-o]

The *arg*$_1$ slot is to be occupied by the SUBJ, i.e. *stimulus* NP. If we said that the next more prominent participant maps onto arg$_2$, the experiencer should be OBJ. However, if we claim the *experiencer* NP maps onto *arg*$_2$, then we are also entailing it has a [-r] feature, which will make it available to become the subject of a passive construction. With the flexibility shown by Kibort (2007)'s version of Lexical Mapping Theory, we do not necessarily need to map to all the argument slots in order: *arg*$_1$, *arg*$_2$, *arg*$_3$... Participants can be mapped onto any of the slots, provided they have the features associated with that slot.

The fact that we have a participant with distinctive morphology (dative case) and its unavailability to become subject of a passive indicate that we should map this argument onto the arg_3 slot with [+o] [+r] features. We subsequently describe our psychological predicates with a dative experiencer as follows:

(37)

		x	***b***
	gustar	⟨ arg_1	**arg_3** ⟩
		[-o]	**[+o]**
			[+r]
		SUBJ	**OBJ_θ**
		(*stimulus*)	(*experiencer*)

We have so far argued that the stimulus NP is SUBJECT and the experiencer argument maps as OBJ_θ. We will now explore the unexpected ordering by which the dative experiencer appears left fronted and the subject appears postverbally.

6 Unexpected order and doubling

As previously shown, the unmarked order of constructions with psychological predicates is as seen in (38):

(38) A Laura le gustan las manzanas
DAT Laura .3SG.DAT like.PRS.3PL the.F.PL apple.PL

Apples are pleasing to Laura.
='Laura likes apples.'

This ordering of elements is consistent with the thematic hierarchy illustrated below following Dowty (1991). However, it seems to divert from the assumption that Spanish is a SVO language with preverbal subjects.

(39)

	x	*b*
gustar	⟨ arg_1	arg_3 ⟩
	(*stimulus*)	(*experiencer*)

(40) *agent* ⟨ *instr./experiencer* ⟨ *patient* ⟨ *source/ stimulus /goal* (Dowty, 1991)

Furthermore, in this type of constructions we can have both a dative NP and a dative weak pronoun referring to the same participant, a phenomenon known as doubling. We will now assess some discourse properties that will help us deal both with doubling and the unexpected order in an elegant manner.

6.1 Weak Focus

A preliminary sensible approach to this issue would be to treat the dative noun phrase as part of information structure and give it a discourse function such as TOPIC or FOCUS. R.A.E (2010) explains that topics in Spanish can appear at the beginning of the sentence but one key feature of topics is the presence of commas in writing or the equivalent intonation in speech. An example of topic in Spanish is clitic dislocation:

(41) a. Llamé a Juan
call.PST.1SG ACC Juan

'I called Juan'

b. A Juan **lo** llamé
ACC Juan 3.M.SG.ACC call.PST.1SG

'Juan, I called'

Focus on the other hand cannot be elided since they highlight or give prominence to a particular part of the discourse. When a focus is fronted, the subject appears postverbally, in a configuration that in many cases mirrors that of interrogative or exclamative sentences. An example of focus in Spanish is contrastive focus:

(42) A JUAN llamé, no a Laura
ACC Juan call.PST.1SG NEG ACC Laura

'I called Juan, not Laura

Note that even though *a Juan* is fronted in both (41b) and (42), only (41b) requires pronominal reduplication. As noted by Leonetti & Escandell-Vidal (2009, p.157), there are however other constructions that display fronting of an element but do not fit comfortably in the description of focus or topic: "These constructions seem to have mixed properties: on the one hand, they resemble clitic dislocations in that the fronted constituent does not bear any emphatic stress; but, at the same time, like in contrastive focalisation, the construction does not include any resumptive clitic." Benincà (2004) shows evidence of a weak/unmarked focus in medieval Romance languages:

(43) a. **Autre chose** ne pot li roi trouver
another thing not can the king find

'The king cannot find any other thing.' [OLD FRENCH]

b. **Mal cosselh** donet Pilat
Bad advice gave Pilate

'Pilate gave bad advice.' [OLD PROVENÇAL]

c. **Con tanta paceença** sofria ela esta enfermidade
with so-much patience suffered her this desease

'She endured this desease with huge patience.' [OLD PORTUGUESE]

d. **Bon vin** fa l'uga negra
good wine makes the wine grape

'Black grapes make good wine.' [OLD MILANESE]

e. **Ciò** tenne il re a grande maraviglia
This has the king as a great wonder

'The king regards this as a great wonder.' [OLD FLORENTIN]

This weak focus fronting strategy is still used in Spanish (also in Sicilian and Sardinian) (Batllori & Hernanz, 2015):

(44) a. **Mucho** me temo que la crisis no ha tocado
Much 1SG.DAT fear.PRS.1SG that the crisis NEG have.PRS.3SG touch.PASTPART
fondo
bottom

'I am afraid the crisis is not over yet.'

b. **Eso mismo** pienso yo
That same think.PRS.1SG 1SG.NOM

'I think the same.'

Weak focus fronting presents the following properties:

- It involves leftward fronting of a constituent
- No intonation/prosodic prominence
- Only one weak focus allowed
- Adjacency between the fronted element and the finite verb which necessarily pushes the subject to appear in postverbal position:

(45) a. **Algo** estarán tramando estos niños
Something be.FUR.3PL plot.PRSPART this.M.PL child.PL

' These children must be up to something.'

b. ***Algo** estos niños estarán tramando

- No resumption in object fronting constructions (ruling out CLLD):

(46) a. **Algo** estarán tramando estos niños
Something be.FUR.3PL plot.PRSPART this.M.PL child.PL

' These children must be up to something.'

b. ***Algo**$_i$ **lo**$_i$ estarán tramando estos niños
Something 3SG.M.ACC be.FUR.3PL plot.PRSPART this.M.PL child.PL

' These children must be up to something.'

All of these properties fit with the behaviour we have observed for the dative noun phrase in psychological predicate constructions so we will incorporate the notion of weak focus into our analysis in the following section.

6.2 Dative NP as weak focus

If we treat **a Laura** in (38) as weak focus, it follows that we have a postverbal subject, no prosodic prominence and we are not dealing with a case of resumption so the weak pronoun must be something else. We can consider the weak pronoun as the argument the verb subcategorises for and then the dative NP is a weak focus that is linked anaphorically with the OBJ_θ, which will result in the f-structure in (47) below with the corresponding equations as in (48):

(47)
$$\begin{bmatrix} \text{PRED} & \text{`LIKE<(SUBJ)(OBJ}_\theta\text{)>'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`APPLES'} \\ \text{INDEX} & \begin{bmatrix} \text{NUM} & \text{PL} \\ \text{PERS} & 3 \end{bmatrix} \end{bmatrix} \\ \text{OBJ}_\theta & \begin{bmatrix} \text{PRED} & \text{`PRO'} \\ \text{INDEX} & \begin{bmatrix} \text{NUM} & \text{SG} \\ \text{PERS} & 3 \end{bmatrix} \\ \text{CASE} & \text{DAT} \end{bmatrix} \\ \text{WFOC} & \begin{bmatrix} \text{PRED} & \text{`LAURA'} \\ \text{INDEX} & \begin{bmatrix} \text{NUM} & \text{SG} \\ \text{PERS} & 3 \end{bmatrix} \\ \text{CASE} & \text{DAT} \end{bmatrix} \end{bmatrix}$$

(48) $(\uparrow \text{WFOC CASE}) =_c (\uparrow \text{OBJ}_\theta \text{ CASE})$
$(\uparrow \text{WFOC INDEX}) =_c (\uparrow \text{OBJ}_\theta \text{ INDEX})$

This possible analysis raises issues immediately: the WFOC has to be bound by the OBJ_θ; this relation must be local, i.e. bound by the OBJ_θ in its mother's structure and we are also introducing a new type of DF which is not necessarily ideal.

We can easily combine this problematic analysis with the idea of weak focus by adding a [WFOC +] feature in the f-structure of the GF and correlate that with a phrase structure rule that states that if the SPEC IP position is occupied by something other than the SUBJ GF, then that f-structure must have the WFOC + feature. The weak pronoun will be PRO or simply agreement[7] and we can deal with it following Bresnan (2001)'s approach to River Plate Spanish object clitics.

[7]The weak pronoun is most likely undergoing grammaticalization. We see the process is more completed with psychological predicates as the weak pronoun is obligatory, as opposed to the dative in double object constructions where, even if its presence is preferred by many speakers, it is still optional or to the accusative pronoun, which has an even more restricted distribution when the NP is present.

(49)
$$\begin{bmatrix} \text{PRED} & \text{`LIKE}\langle\ (\text{SUBJ})\ (\text{OBJ}_\theta)\rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`APPLES'} \\ \text{INDEX} & \begin{bmatrix} \text{NUM} & \text{PL} \\ \text{PERS} & 3 \end{bmatrix} \end{bmatrix} \\ \text{OBJ}_\theta & \begin{bmatrix} \text{PRED} & \text{`LAURA'} \\ \text{INDEX} & \begin{bmatrix} \text{NUM} & \text{SG} \\ \text{PERS} & 3 \end{bmatrix} \\ \text{CASE} & \text{DAT} \\ \text{WFOC} & + \end{bmatrix} \end{bmatrix}$$

C-structurally, the OBJ_θ occupies SPEC IP, which triggers the postverbal position of the subject. This is supported by the fact that if we already have a weak focus in that position, the OBJ_θ cannot appear at the beginning -unless given discourse prominence-:

(50) Mucho le gustan (a Laura) las manzanas (a Laura)
much 3SG.DAT like.PRS.3PL DAT Laura the.F.PL apple.PL DAT Laura

Apples are very pleasing to Laura.
='Laura likes apples a lot.'

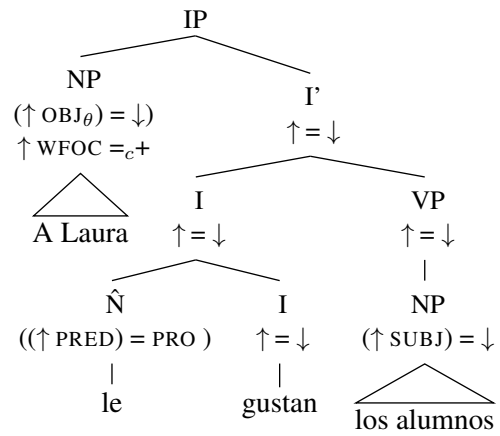
6.3 Remaining issues: postverbal SUBJ

An issue remains with the postverbal position of the subject and that is how to characterise it in the c-structure. We discard the possibility of having three branches stemming from the IP with two specifiers, which leaves us with the possibility of either adding a new S to the structure since it includes the subject or introducing a headless VP. It is not clear whether proposing a category S is actually plausible for Spanish so we will introduce a headless VP. LFG assumes that daughters of phrasal categories are optional so the head of a maximal phrase such as VP does not need to appear. This has been proposed for languages that place tensed verbs in I, which Spanish does (see Sells (2001) or King (1995) for distribution of verbs in Swedish and Russian):

(51) A Laura le gustan los alumnos
DAT Laura 3SG.DAT like.PRS.3PL the.M.PL student.PL

'Laura likes the students.'

(52)



7 Summary and conclusion

In this paper we have seen that thematic prominence, together with c-structure pre-verbal position, point at the possibility that the experiencer could be SUBJ. However, the real SUBJ is the post-verbal stimulus and the experiencer is interpreted as the logical subject. Thematic hierarchy is more prominent and reflects on c-structure, leaving the f-structure untouched. We have shown that the c-structure position of the experiencer can be ensured by adding a [WFOC +] feature to the description of the OBJ_θ. Treating the experiencer as weak focus also explains the postverbal position of the subject and why doubling is not to be considered an instance of clitic left dislocation.

References

Alarcos Llorach, Emilio. 1994. *Gramática de la Lengua Española.* chap. XXI, 266–276. Madrid: Espasa-Calpe.

Alsina, Alex. 1996. *The Role of Argument Structure in Grammar: Evidence from Romance*. Stanford, CA: CSLI Publications.

Batllori, Montserrat & Maria-Lluïsa Hernanz. 2015. Weak focus and polarity: Asymmetries between Spanish and Catalan. In Theresa Biberauer & George Walkden (eds.), *Syntax over Time: Lexical, Morphological, and Information-Structural Interactions*, chap. 17, 280–298. Oxford University Press.

Belleti, Adriana & Luigi Rizzi. 1988. Psych-verbs and theta-theory. *Natural Language and Linguistic Theory* 6(3). 291–352.

Benincà, Paola. 2004. The Left Periphery of Medieval Romance. http://www.humnet.unipi.it/slifo/2004vol2/Beninca2004.pdf.

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

Cuervo, María Cristina. 2003. *Datives at Large*: Massachusetts Institute of Technology dissertation.

Cuervo, María Cristina. 2010. Some Dative Subjects Are Born, Some Are Made. In Claudia Borgonovo et al. (ed.), *Selected Proceedings of the 12th Hispanic Linguistics Symposium*, 26–37. Somerville, MA: Cascadilla Proceedings Project.

Dowty, David. 1991. Thematic Proto-Roles and Argument Selection. *Language* 3(67). 547–619.

Fernández Soriano, Olga. 1999. Two types of impersonal sentences in Spanish: Locative and dative subjects. *Syntax* 2(2). 101–140.

Grimshaw, Jane. 1990. *Argument Structure*. MIT Press.

Kibort, Anna. 2007. Extending the Applicability of Lexical Mapping Theory. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG07 Conference*, 250–270. CSLI Publications.

Kibort, Anna. 2008. On the syntax of ditransitive constructions. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG08 Conference*, 312–332. CSLI Publications.

Kibort, Anna. 2013. Objects and Lexical Mapping Theory. Abstract for LFG13 http://web.stanford.edu/group/cslipublications/cslipublications/LFG/18/abstracts/lfg13abs-kibort.html.

King, Tracy Holloway. 1995. *Configuring Topic and Focus in Russian.* Dissertations in Linguistics. Stanford, CA: CSLI Publications.

Landau, Idan. 2010. *The Locative Syntax of Experiencers.* Linguistic Inquiry Monographs. The MIT Press.

Leonetti, Manuel & Victoria Escandell-Vidal. 2009. Fronting and verum focus in Spanish. In Andreas Dufter & Daniel Jacob (eds.), *Focus and Background in Romance Languages* (Studies in Language Companion Series 112), 155–204. John Benjamins Publishing Company.

Masullo, Pascual J. 1992. *Incorporation and Case theory in Spanish: A crosslinguistic perspective*: University of Washington, Seattle dissertation.

Mendívil Giró, José Luis. 2002. La estructura ergativa de *gustar* y otros verbos de afección psíquica en español. In *Actas del V Congreso de Lingüística General*, Universidad de León.

R.A.E. 2010. *Las funciones informativas.* chap. 40, 753–770. Manual de la Nueva gramática de la lengua española. Espasa. Real Academia Española y Asociación de Academias de la Lengua Española. edn.

Sells, Peter. 2001. *Structure, Alignment and Optimality in Swedish.* Stanford, CA: CSLI Publications.

Solà i Pujols, Jaume. 1992. *Agreement and Subjects*: Universitat Autònoma de Barcelona dissertation.

Vallduví, Enric. 1993. Catalan as VOS: Evidence from information packaging. In W. J. Ashby, M. Mithung, G. Perisinotto & E. Raposo (eds.), *Linguistic perspectives in the Romance languages*, 335–350. Benjamins.

Vallduví, Enric. 2002. L'oració com a unitat informative. In J. Solà, M.R. Lloret, J. Mascaró & M. Pérez-Saldanya (eds.), *Gramàtica del català contemporani*, vol. 2, chap. 4, 1221–1279. Empúries.

Vanhoe, Henk. 2002. Aspects of the syntax of psychological verbs in Spanish. A lexical functional analysis. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG02 Conference*, 373–389. CSLI Publications.

Vogel, Carl & Begoña Villada. 1999. An HPSG Analysis of Grammatical Relations, Syntactic Valency and Semantic Argument Structure in Spanish Psychological Predicates and other Instances of Quirky Case and Agreement. Tech. rep. Trinity College Dublin.

Zaenen, Annie, Joan Maling & Höskuldur Thráinsson. 1985. Case and Grammatical Functions: The Icelandic Passive. *Syntax and Semantics* 24. 95–136.

A Constraint-Based Analysis of the Objects of VO Verbal Compounds in Mandarin Chinese

Dewei Che
University of Vienna

Adams Bodomo
University of Vienna

Proceedings of the LFG'18 Conference

University of Vienna

Miriam Butt, Tracy Holloway King (Editors)

2018

CSLI Publications

pages 171–190

http://csli-publications.stanford.edu/LFG/2018

Keywords: VOCs, lexis, constituency, idioms, constraining equation, LFG

Che, Dewei, & Bodomo, Adams. 2018. A Constraint-Based Analysis of the Objects of VO Verbal Compounds in Mandarin Chinese. In Butt, Miriam, & King, Tracy Holloway (Eds.), *Proceedings of the LFG'18 Conference, University of Vienna*, 171–190. Stanford, CA: CSLI Publications.

Abstract

VO verbal compounds (VOCs) have become a topical issue within studies on wordhood and the syntax-semantics interface. However, the issue can become more complicated when VOCs take an extra object. Some previous analyses have often run into problems mostly because they assign the wrong grammatical function to these objects in question. This paper provides a complex predicate analysis by adopting the ideas of Ahmed et al. (2012), combined with recent findings from Zhuang et al. (2013) on the status of the O in the VOC. The description and analysis especially focus on double object realization of VOCs in Mandarin Chinese and thus provide a generalized account of the representation of their argument relations within the LFG framework.

1. INTRODUCTION

VO verbal compounds are also known as separable verbs in Mandarin Chinese given that intervening items can be placed between the verb (V) and the object (O). There have been major questions about the issue of their wordhood in Chinese and other languages as VOCs do combine to form a 'word-like unit', but at the same time they exhibit some degree of separability between the two parts (Chao 1968, Li and Thompson 1981, C.T. Huang 1984, 1988, C.R. Huang 1990, O.-S. Her 1997, 1999, Tang 2000, Zhuang et al. 2013, Che 2014, among others). As a result, the lexical status of Chinese VOCs has long been disputed among linguists. In the following examples, we may consider *jian-mian* 'to meet' as a lexical word as in (1a), while in (1b), it can appear as a syntactic phrase.

(1) a. *women mingtian jian-mian.*
we tomorrow see face
'We'll meet tomorrow.'

b. *women jian-guo liang-ci mian.*
we see-PERF two-CL face
'We've met twice.'

VOCs are commonly treated as idioms in the sense that they have non-compositional meanings, i.e. we cannot put together the literal meaning of their individual parts. For example,

†The authors thank the editors, the reviewers, and the audience at LFG18 for making helpful comments and suggestions that improved the content.

(2) a. *chi-cu*
eat vinegar
'be jealous'

b. *chao-youyu*
fry squid
'dismiss'

c. *qiao-zhugang*
hit bamboo
'blackmail'

d. *kou-maozi*
put hat
'label'

Another interesting phenomenon involving VOCs is that whenever they take an extra object, this object cannot be placed after the VOC (Li 2009, Cai 2010 and Zhuang et al. 2013). It can appear in two forms: 1) a possessive object between the V and the O whereby it becomes the possessor of the O; 2) a PP construction before the verb, as shown in (3) and (4).

(3) a. *women mingtian jian Zhangsan de*[1] *mian.* (Possessive)
we tomorrow see Zhangsan DE face
'We'll meet Zhangsan tomorrow.'

b. *women mingtian he Zhangsan jian-mian.* (PP)
we tomorrow with Zhangsan see face
'We'll meet with Zhangsan tomorrow.'

(4) a. *chao Zhangsan de youyu* (Possessive)
fry Zhangsan DE squid
'fire Zhangsan'

b. *ba Zhangsan chao-youyu* (PP)
BA Zhangsan fry squid
'fire Zhangsan'

The introduction of an extra argument renders the components of VOCs structurally discontinuous and difficult to interpret as a unit, as shown in (3a) and (4a). Thus, this paper aims to deal with double object realization within a discontinuous VOC. In Section 2, we investigate the object status of the O in the VOC and the nature of VOCs as idiom chunks. Section 3 reviews existing approaches regarding the analysis of VOCs and their objects. In Section 4, we explore the grammatical function of the extra argument and consider possible solutions to represent argument relation within Chinese VOCs. Section 5 concludes the paper.

[1] The morpheme *de* is a typical possessive marker in Mandarin Chinese.

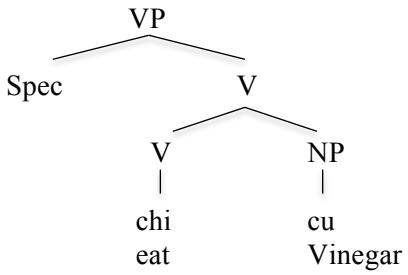
2. FORMAL TESTS

In this section, syntactic tests and arguments are provided to prove that the O is really an argument and that the VOC as a whole is an instance of an idiom chunk.

2.1 The object status of the O

Zhuang et al. (2013) present a similar insight with regard to the status of the O. Based on Her (1999), they suggest that the Os in VOCs are referential, although in a metaphorical way,[2] and this type of expression is called quasi-arguments according to Ouhalla (1999). The analysis of Zhuang et al. was conducted within GB by using θ-Criteria and Visibility Condition (Chomsky 1981). Even though it was examined in a different framework, the bottom line is that the Os in VOCs occupy a place properly governed by the Vs, as shown below in (5).

(5)



'be jealous of'

(Zhuang et al. 2013: 271)

For us, quasi-arguments seem like some fancy terminology, so we will simply prove that the O is really an argument that can be presented at f-structure in our analysis. It can be done by several syntactic tests. The main ones are topicalization, modification, passivization, and question formation. The object of the VOC *peng-dingzi* 'meet rejection' in (6) is preposed by topicalization to achieve the same effect of emphasis as a normal object in (7).

[2] Although the meaning of these VOCs is non-compositional, the original meaning of the O does seem to contribute metaphorically, for example, the sour taste of vinegar resembles the feeling of jealousy (*chi-cu* {eat-vinegar}'be jealous of') and bumping into a nail feels like receiving rejection (*peng-dingzi* {bump-nail}'meet rejection'). These make perfect sense in Chinese culture. However, it is complicated when it comes to the origin of idioms, which is beyond the scope of our current study.

(6) *dingzi ta peng duo le.*
nail s/he bump many PERF
'S/he suffered a lot of rejections.'

(7) *pingguo ta xihuan chi.*
apple s/he like eat
'S/he likes eating apples.'

As observed by Her (1999), the O in a VOC can be modified in numerous ways as an object. The modifiers are not only confined to quantitative words in (1b), but also adjective, temporal, determiner phrases and etc.

(8) *ta zhuan chi nen doufu*
s/he only eat tender tofu
'S/he flirts with the young ones only.'

(9) *zuotian de doufu hai mei chi gou ma?*
yesterday DE tofu still not eat enough Q
'Didn't you flirt enough yesterday?'

(10) *ta peng-le zhe dingzi.*
s/he bump-PERF this nail
'S/he suffered this rejection.'

Passive constructions are marked by *bei* in Mandarin Chinese and *bei* phrases occur preverbally. As suggested in Dalrymple (2001:48), passivization is one of the most widely available tests for direct-objecthood.

(11) a. *ta chi jin le shishang de doufu.*
s/he eat to the greatest extent PERF world DE tofu
'S/he has been flirting everywhere.'

b. *shishang de doufu bei ta chi jin le.*
world DE tofu BEI s/he eat to the greatest extent PERF
'S/he has been flirting everywhere.'

Mandarin Chinese is known as a typical *wh-in-situ* language. The object status of the O is also shown in the process of question formation. The context for the following sentences is two employees who are joking about their boss after s/he just fired another employee in succession.

(12) A: *ni cai laoban zui xihuan chao shenme?*
you guess boss most like fire what
'What do you think the boss likes firing most?'

B: *chao-youyu.*
fire squid
'To dismiss'

So far, it has been shown that the O can be topicalized, modified, passivized, and questioned properly. Meanwhile, there are other small tests to support our claim too, for instance, verb copying within a sentence in (13).

(13) *ta peng-dingzi peng-le bantian.*
s/he bump-nail bump-PERF half-day
'S/he has been facing rejections for quite a while.'

As shown above, it has been amply demonstrated that the O has the qualities of a syntactic object.

2.2 The VOCs as idiom chunks

A definitive feature of idiom chunks is their noncompositional semantics. According to Huang (1990), one syntactic environment to test idiom chunks is coordination which involves parallel constructions sharing a single grammatical relation to the remaining elements of the sentence. Many studies (Ackerman and Lesourd 1997, Mohanan 1997, Bodomo 1998, Bresnan and Mchombo 1995) have also used coordination as a test for unithood. In (14) two conjoined NPs are governed by the same verb, but one of the conjuncts has a literal reading, *chi pangxie* 'eat crabs'. The example illustrates that when *chi* takes a conjoined NP, the only possible reading is the literal 'to eat' reading. The data shows that the homophonous verbs of the literal reading and the idiom-chunk reading are instantiations of two different lexical predicates with different selectional restrictions and subcategorization frames.

(14) *Lisi chi pangxie gen cu*
Lisi eat crab AND vinegar
a. 'Lisi eats crabs and vinegar.'
b. *'Lisi eats crabs and is jealous.'

(Huang 1990: 269)

Wasow et al. (1983) classify English idioms, most of which are of the VO construction, into three groups: noncompositional idioms (*kick the bucket, saw log*), conventionalized metaphor (*take advantage of, spill the beans*), and compositional idioms (*pull strings*). We find it hard to understand the so-called compositional idioms. Admittedly, there is more transparency in this type than the other two. But the derived meaning of *pull strings* 'to use influence' does not correspond to the combination of its literal parts, either.

Based on our observation as well as previous studies, VOCs in Chinese usually belong to the first two groups. A close example we can think of as compositional is a VOC such as *chang-ge* {sing-song} 'sing'. However, as we can see, the meaning of the noun is basically incorporated into the verb already.

3. PREVIOUS ANALYSES

Within the framework of Lexical Functional Grammar (LFG), there are two previous works which especially address both lexical discontinuity and object realization of VOCs: C.R. Huang (1990) and O.-S. Her (1999).

3.1 Huang (1990)

Huang (1990) encapsulates the string [NP_1 DE NP_2] as the possessive-object NP construction (POBJ) and NP_2 forms a discontinuous construction with the matrix verb. The component *de* between NP_1 and NP_2 is optional. He regarded that the LFG framework has 'an edge in analyzing the POBJ construction because the distribution of this construction is closely related to the set of idiom chunks whose syntax and semantics have to be lexically marked' (Huang 1990:277). As pointed out by Nunberg et al. (1994: 510), 'positing a single underlying idiom which may be transformationally deformed is claimed to be not only parsimonious, but unavoidable'. As a result, idioms could be best analyzed by direct generation of surface structures within a framework like LFG.[3] Huang's analysis is shown below.

(15) *Sanbai chi Yunniang de cu.*
Sanbai eat Yunniang DE vinegar
'Sanbai is jealous of Yunniang.'

C-structure rules

(16) a. S → NP VP
(↑ SUBJ) = ↓ ↑ = ↓
b. VP → V NP
↑ = ↓ ↑ = ↓
c. NP → (NP) (CL) N
(↑ OBL) = ↓ ↑ = ↓ ↑ = ↓

[3] For more discussions, see Nunberg et al. (1994).

Lexical entries

(17) a. *chi* V, (↑ VMORF) = CHI

b. *cu* N, (↑ PRED) = 'BE-JEALOUS< (SUBJ) (OBL) >'

(↑ VMORF) = *c* CHI

(↑ CL) = DE

F-structure

(18)

SUBJ	[PRED 'Sanbai']
OBL	[PRED 'Yunniang']
PRED	'BE-JEALOUS < (SUBJ) (OBL) >'
CL	DE
VMORF	CHI

As seen from (16b), Huang employed the functional head equation ↑ = ↓ on both the lexical head V *chi* 'to eat', and the NP *cu* 'vinegar'. The consequence is that the N in (17b) is assigned the PRED feature and the constraining equation ↑VMORF = c CHI ensures that the idiomatic reading of 'BE-JEALOU' must co-occur with the verb *chi*.

There is a fundamental problem with Huang's analysis. He treated the N as a co-head with the V of the VP instead of an argument of the V. However, it has been proved in Section 2.1 that the O in the VOCs possesses the qualities of a syntactic object. Another problem is associated with the treatment of NP_1 as an oblique object. Indeed, the realization of this object is worth further discussions. We will revisit the issue regarding NP_1 in Section 4.1.

3.2 Her (1997, 1999)

According to Her (1999), idioms have regular syntactic structures as represented by a-structure, f-structure and c-structure in LFG framework. The literal reading and the idiomatic interpretation are determined by syntactic constraints and motivations based on metaphors, metonymies, or mental images. The concept of motivation is used in the sense of Lakoff (1987: 488):

> The relationship between A and B is motivated just in case there is an independently existing link, L, such that A-L-B "fit together". L makes sense of the relationship between A and B.

With the help of an attribute IDIOM-LINK, Her (1997) specifies the syntactic constraints in the lexical entry of the idiom's lexical head, the verb. Take the VO idiom *chi-doufu* 'to flirt with' for example.

(19) Lexical entry of *chi*

chi, V
PRED 'EAT <ag-SUBJ th-OBJ)>'

IF [SUBJ HUMAN =$_c$ +
OBJ PRED =$_c$ '*doufu*'
[IF OBJ ADJS
THEN OBJ ADJS =$_c$ {[PRED '*ruan* (tender)']}]
[IF OBJ POSS
THEN OBJ POSS HUMAN =$_c$ +]]

THEN [IDIOM-LINK = *chi-doufu* (to flirt with)]

Noticeably, a number of syntactic constraints need to be met as listed below.

(20) a. The SUBJ has to be HUMAN as specified by the constraining equation, (↑ SUBJ HUMAN) =c +.

b. The PRED of the OBJ has to be 'doufu' as required by the constraining equation, (↑ OBJ PRED) =c *'doufu'*.

c. If the OBJ has an adjunctive element, the PRED of this adjunctive element has to be 'ruan' as dictated by the constraining equation, (↑ OBJ ADJ) =c PRED '*ruan*'. Note that the OBJ can go without any adjunctive element since it is an optional requirement specified by IF...THEN.

d. If the OBJ has a POSS function, this POSS must be HUMAN as specified by the constraining equation (↑ OBJ POSS HUMAN) =c +. Again, the OBJ does not necessarily have to contain a POSS function as can be seen from the IF...THEN.

As long as these syntactic constraints are all fulfilled, the attribute IDIOM-LINK triggers the idiom interpretation mechanism just described above, whereby the idiom interpretation is linked to the 'qualified' f-structure.

However, a sentence with a VOC may intend to express a literal meaning at the same time. For example,

(21) *Sanbai chi Yunniang de doufu.*
Sanbai eat Yunniang DE tofu
'Sanbai eats Yunniang's tofu.'

The sentence fulfills all the specifications in (20) and the idiom interpretation mechanism thus must be triggered, which is clearly not intended by (21). All in all, Her's analysis is not well-formed within the LFG formalism. And different from Huang (1990), Her did not regard NP_1 in [NP_1 DE NP_2] as an object but merely an adjunct of NP_2. We will discuss the status of NP_1 in the following section soon.

4. THE PRESENT PROPOSAL

For convenience of analysis, we are following Huang (1990) by using NP_1 to refer to the extra argument that renders the components of VOCs structurally discontinuous. The original O in the VOC is termed NP_2. In Section 2.1, it has been shown that NP_2 is a grammatical object. Now we will continue to explore the grammatical function of NP_1 and its relation to NP_2.

4.1 The grammatical function of NP_1

First of all, the argument position of NP_1 is shown by the fact that it can be questioned.

(22) *lanban chao shui de youyu?*
boss fry who DE squid
'Who has the boss fired?'

At first sight, NP_1, reflecting its structural position, is easily mistaken for a usual possessor of NP_2 especially in the presence of a possessive marker *de*, although *de* is optional. However, the relation between NP_1 and NP_2 is clearly more than that of a usual possessor and possessee, as noticed by Huang (1990: 271).

(23) *wo jian-le ta de mian.*
I see-PERF s/he DE face
'I met him.'

(24) *wo jian-le zhuozi de mian.*
I see-PERF table DE face
a. * 'I met the table.'
b. 'I saw the surface of the table.'

These sentences show the relationship between the argument NP_1 and the predicate governing it. In (23), *jian-mian* {see-face} 'meet' assumes the idiom-chunk meaning. However, the literal meaning of *mian* 'face' is the only available interpretation as in (24b) because the idiom chunk imposes selectional restrictions on NP_1 and requires it to be a human object, thus the ungrammaticality of (24a), where *zhuozi* 'table' is an inanimate noun. The fact that the idiom chunk *jian-mian* 'to meet' imposes selectional restrictions on NP_1 indicates that NP_1 is an argument of the idiom chunk since predicates can impose selectional restrictions only on their arguments.

Then, passivization provides a further test to support NP_1's status as an argument. Obviously, the object status of NP_1, *yunniang*, is supported by the fact that it can be readily passivized as in (25b), since the most crucial fact is that in all known cases of passive sentences marked by *bei*, the subjects are also grammatical objects of their active counterparts in Mandarin Chinese.

(25) a. *Sanbai chi Yunniang doufu.*
Sanbai eat Yunniang tofu
'Sanbai has been flirting with Yunniang.'

b. *Yunniang bei Sanbai chi-doufu.*
Yunniang BEI Sanbai eat-tofu
'Yunniang has been flirted with by Sanbai.'

Interestingly, Huang (1990) suggests that NP_1 is an oblique object of the discontinuous VOC by comparing it to corresponding sentences with NP_1 occurring in a preverbal PP.

(26) a. *wo jian-le Zhangsan (de) mian.*
I see-PERF Zhangsan face
'I met Zhangsan.'

b. *wo gen zhangsan jian-mian.*
I with Zhangsan see-face
'I met Zhangsan'.

According to Huang, since (26a) and (26b) are synonymous, the grammatical function of NP_1 in (26a) should correspond to an oblique object in (26b) too. We find it very questionable. Similarly, we can compare it to 'double object' constructions in English.

(27) a. Mary gave John a watch.
b. Mary gave a watch to John.

Following Huang's logic, we should assign the oblique object to *John* both in (27a) and (27b). However, it would be obviously wrong. As we all know, *John* in (27a) should be the direct object (OBJ) with *watch* being the second object (OBJ_θ) and *John* in (27b) is a real oblique object (OBL) marked by the preposition *to*. With (26a), it is more sensible to argue that *Zhangsan* is a direct object too. As a matter of fact, it turns out to be true. In a canonical Chinese double object construction, the position for *Yunniang* in (25a) and *Zhangsan* in (26a) is reserved for the direct object, as shown in (28).

(28) *Zhangsan gei-le Xiaoli yi-ben shu.*
Zhangsan give-PERF Xiaoli one-CL book
'Zhangsan gave Xiaoli one book.'

As further pointed out by (Dalrymple 2001:46), 'if a Recipient appears as a full NP in a double object construction, it is the sole candidate for passivization; the second object is excluded'. This can be perfectly instantiated by the grammaticality of (25b) and the ungrammaticality of (29).

(29) **doufu bei Sanbai chi Yunniang*.
tofu BEI Sanbai eat Yunniang

In return, it also provides a good explanation why NP_2 can be passivized in a transitive VOC as in (11) but not in a ditransitive VOC as in (25a). The reason is that in ditransitive VOCs, the grammatical relation between NP_1 and NP_2 is similar to that of a direct object and a second object. However, it is not exactly the same as a double object construction [V NP_1 NP_2] given the fact that V…NP_2 forms an idiom chunk.

4.2 Bodomo et al. (2017)

Having elucidated the nature of NP_1 and NP_2, we now consider an analysis that can represent the realization of these objects involving discontinuous VOCs. The very first possible solution arises from Bodomo et al. (2017). In that paper, we provided a lexicalized analysis for the transitive VOCs. For example,

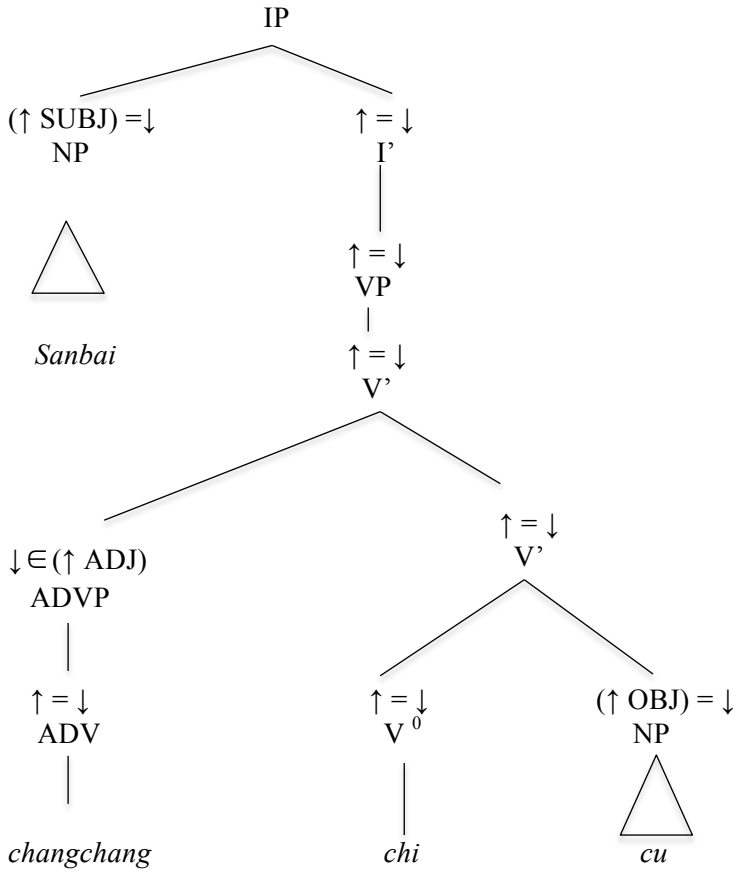
(30) *Sanbai changchang chi cu.*
Sanbai often eat vinegar
'Sanbai often gets jealous.'

Lexical entries

(31) (a) *chi* V (↑ PRED) = 'BE-JEALOUS<(↑ SUBJ)>(↑ OBJ)'
(↑ OBJ FORM) =*c* CU
(b) *cu* N (↑ FORM) = CU
(c) *sanbai* D (↑ NUM) = SG
(↑ PERS) = 3
(↑ PRED) = 'Sanbai'
(d) *changchang* ADV (↑ PRED) = 'changchang'

C-structure

(32)



We employed Bresnan's (1982) classical treatment of idiom chunks, namely, the use of semantically empty 'form'-bearing homophones with appropriate selectional restrictions. The verb *chi* 'eat' explicitly requires its

object to be *cu* 'vinegar' when it is associated with a specific meaning. The feature FORM represents a meaningless element like *cu* 'vinegar' as in the VOC *chi-cu* 'be jealous of'. Thus the O is given no PRED and placed outside the angled-brackets to show that the verb *chi* 'eat' has a non-thematic argument, as shown in (31a).

Although our analysis may apply to most of Chinese transitive VOCs, it runs into a major difficulty: the modifiability of the Os. Their status as a non-thematic object is purely motivated at the semantic level. If we assume them to be totally meaningless, how could they sometimes be modified (though not as freely as a normal object due to their idiomatic meaning) when we consider an example like (33) (also see (8) - (11))?

(33) *ta* *changchang* *chi* *gan* *cu.*
s/he always eat dry vinegar
'S/he always gets jealous to an absurd extent.

Bresnan (1982) also seems to provide a solution to our case of ditransitive VOCs. Take the idiom *keep tabs on* for example (Bresnan 1982: 46).

(34) The FBI kept tabs on John.

Thematic structure: keep-tabs-on < ag th >
| |
FBI John
Subcategorization: < S OBL > OBJ FORM TABS

This treatment was formulated within the so-called classic, i.e. pre-LMT, model of LFG and it was bound to run into difficulty within mapping theory as there is simply no way to derive the required lexical form, <S OBL> OBJ FORM TABS, and link the OBL to a theme argument.

As discussed above, we might need a different solution to continue our endeavor with ditransitive Chinese VOCs.

4.3 The reanalysis

In the literature, there are other works that have investigated similar constructions such as Lakoff (1987), Fillmore (1988), Butt (2003, 2010, 2014), Kay and Fillmore (1999), Butt et al. (2003), Kaplan and Zaenen (2003), Asudeh et al. (2008), Ahmed (2011), Megerdoomian (2012), Arnold (2015) and Findlay (2017).[4] Among them, Ahmed et al. (2012) present a best

[4] Findlay's TAG-LFG approach may work well with some cases of Chinese VOCs in which NP_2 can be treated as a direct object. However, problems occur when NP_1 takes over the function of direct object from NP_2 in a ditransitive VOC.

solution to us in terms of argument relation between the two objects. They provided the analysis of complex predicates (CPs) in the context of dependency bank development, but they kept it general enough to be applied across languages. Complex predicates can be defined as predicates which are composed of more than one grammatical element (either morphemes or words), each of which contributes a non-trivial part of the information of the complex predicate (Alsina et al. 1997). Within the framework of LFG, the pioneer work has been done by Butt (1995, 1998), Alsina (1993, 1996), Frank (1996), Bodomo (1996, 1997), Mohanan (1995), and Kaplan and Wedekind (1993).

Chinese VOCs exhibit the features of complex predicates in several ways. First of all, they are composed of two elements: the verbal and the nominal element. Second, the idiomatic meaning is derived from the combination as a whole. Third, although the VOCs as idiom chunks together impose selectional restrictions on an extra argument as discussed earlier, the special semantics within [NP_1 de NP_2] as a possessor and possessee still holds. In this sense, the relationship between NP_1 and NP_2 is closer than that of V and NP_1. In other words, we can say that NP_2 contributes NP_1 as an argument.

Following Ahmed et al. (2012), we adopt a complex predicate analysis to Chinese VOCs. The insight is especially borrowed from their treatment of N-V complex predicates in Hindi/Urdu.

(35) nAdiyah nE kahAnI yAd k-I
Nadya.F.Sg Erg story.F.Sg.Nom memory.F.Sg.Nom do-Perf.F.Sg
'Nadya remembered a/the story.'

In (35), there are altogether three arguments provided by the verb *kar* 'do': the doer, the action done, and the thing remembered. One argument *yAd* 'memory' contributes one further argument *kahAnI* 'story'. As the performed action, *memory* is encoded as an argument of the verb as part of the complex predication which is referred to as top-level PRED. The idea is demonstrated below when applied to Chinese ditransitive VOCs.

(36) *Sanbai changchang chi Yunniang (de) doufu.*
Sanbai often eat Yunniang DE tofu
'Sanbai often flirts with Yunniang.'

F-structure

(37)

$$\begin{bmatrix} \text{PRED} & \text{`}chi\text{ <SUBJ, `}doufu\text{ <OBJ>'>'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`Sanbai'} \\ \text{NUM} & \text{SG} \\ \text{PERS} & 3 \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`Yunniang'} \\ \text{NUM} & \text{SG} \\ \text{PERS} & 3 \end{bmatrix} \\ \text{ADJ} & \{ [\ \text{PRED} \quad \text{`}changchang\text{'}\ ] \} \end{bmatrix}$$

There are three major advantages with this analysis. First, it helps us recognize both objects' status as an argument. As demonstrated earlier, only one argument is realized in the previous analyses of Chinese VOCs within the LFG framework. Second, it captures the internal structure of VOCs when NP_2 is encoded as an argument of the verb as part of the complex predication. Third, NP_1 is assigned the right grammatical function, which provides a perfect explanation why it is NP_1 (not NP_2) that can be passivized.

5. CONCLUSION

In this paper, we have solved the puzzle on the nature of two objects within the discontinuous VOC. Some interesting phenomena are presented. Both objects are amply demonstrated as a real argument through a series of syntactic tests in our analysis. Some previous analyses have often run into problems mostly because they assign the wrong grammatical function to these objects in question. For example, Huang (1990) fails to recognize NP_2's status as an argument and mistakes NP_1 for an oblique object. Furthermore, We have provided a complex predicate analysis by adopting the ideas of Ahmed et al. (2012). In current analysis, NP_2 contributes one further argument, namely, NP_1. As part of the VOC, NP_2 is then encoded as an argument of the verb to form the complex predication.

Following Ahmed et al. (2012), we have provided a reasonable account for argument relation within Chinese VOCs. The remaining issue is semantics. The direction for future research may rely on a lexical semantic approach to represent the underlying representation and derive the idiomatic meaning.

REFERENCES

Ackerman, Farrell and Philip Lesourd. 1997. Toward a lexical representation of phrasal predicates. In Alsina, Alex, Joan Bresnan & Peter Sells (eds.), *Complex predicates*. Stanford, CA: CSLI Publications, pp. 67–106.

Ahmed, Tafseer, Miriam Butt, Annette Hautli and Sebastian Sulger. 2012. A reference dependency bank for analyzing complex predicates. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation.* Istanbul, Turkey.

Ahmed, Tafseer and Miriam Butt. 2011. Discovering semantic classes for Urdu NV complex predictates. In *Proceedings of the Ninth International Conference on Computational Semantics,* pp. 305-309.

Alsina, Alex. 1996. *The Role of Argument Structure in Grammar: Evidence from Romance.* CSLI Publications, Stanford University.

Alsina, Alex. 1993. *Predicate Composition: A Theory of Syntactic Function Alternations*. Ph.D. dissertation. Stanford University.

Alsina, Alex, Joan Bresnan & Peter Sells. 1997. *Complex Predicates*. Stanford, CA: CSLI Publications.

Arnold, Doug. 2015. Glue semantics for structurally regular MWEs. Poster presented at *the PARSEME 5th general meeting*, 23-24th September 2015, Iași, Romania.

Asudeh, Ash, Mary Dalrymple and Ida Toivonen. 2008. Constructions with lexical integrity: templates as the lexicon-syntax interface. In Miriam Butt and Tracy Holloway King, eds., *Online Proceedings of the LFG08 Conference*, pp. 68-88.

Bodomo, Adams, So-sum Yu and Dewei Che. 2017. Verb-Object Compounds and Idioms in Chinese. *Computational and Corpus-based Phraseology: EUROPHRAS 2017*. Mitkov, R. (Hrsg.). Springer International Publishing AG, pp. 383-396. (Lecture Notes in Computer Science (LNCS) book series; Band 10596)

Bodomo, Adams. 1998. Serial verbs as complex predicates, in Dagaare and Akan. In: Maddieson I., Hinnebusch, T. J. (eds.), *Language history and linguistics description in Africa: trends in African linguistics* 2, pp. 195–204. Africa World Press, Trenton, N.J.

Bodomo, Adams. 1997. *Paths and Pathfinders: Exploring the Syntax and Semantics of Complex Verbal Predicates in Dagaare and Other Languages.* Doctoral dissertation, Norwegian University of Science and Technology.

Bodomo, Adams. 1996. Complex verbal predicates: The case of serial verbs in Dagaare and Akan. In Miriam Butt and Tracy Holloway King, eds., *Online Proceedings of the LFG96 Conference*.

Bresnan, Joan. 1982. The passive in lexical theory. In: Bresnan, J. (ed.), *Mental representation of grammatical relations*, pp. 3–86. MIT Press, Cambridge, MA.

Bresnan, Joan and Sam Mchombo. 1995. The lexical integrity principle: evidence from Bantu. *Natural language and linguistic theory* 13: 181–254.

Butt, Miriam. 2014. Control vs. Complex Predication. Comment on Alice Davison. *Natural Language and Linguistic Theory*, Vol. 32 Issue 1, pp. 165-190.

Butt, Miriam. 2010. The Light Verb Jungle: Still Hacking Away, In M. Amberber, M. Harvey and B. Baker (eds.) *Complex Predicates in Cross-Linguistic Perspective*, pp. 48–78. Cambridge University Press.

Butt, Miriam. 2003. *The Morpheme That Wouldn't Go Away*. Ms., University of Konstanz.

Butt, Miriam. 1998. Constraining argument merger through aspect. In: Hinrichs E, Kathol A, Nakazawa T (eds) *Complex Predicates in Nonderivational Syntax*, Academic Press, pp. 73–113.

Butt, Miriam. 1995. *The Structure of Complex Predicates in Urdu.* CSLI Publications.

Butt, Miriam, Tracy Holloway King and John T. Maxwell. 2003. Productive encoding of Urdu complex predicates in the ParGram Project. In *Proceedings of the EACL03 Workshop on Computational Linguistics for South Asian Languages: Expanding Synergies with Europe*, pp. 9–13.

Cai, S. 2010. Semantic Properties and Syntactic Constructing Process of the Specific Dative Construction "V+X+de+O". *Chinese Teaching in the World* (3): 363–372.

Chao, Yuen Ren. 1968. *A grammar of spoken Chinese.* University of California Press, California.

Che, Dewei. 2014. *The Syntax of Particles in Mandarin Chinese*. Ph.D. dissertation, The University of Hong Kong.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Foris, Dordrecht.

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

Fillmore, Charles. 1988. The mechanisms of 'Construction Grammar'. In Shelley Axmaker, Annie Jaisser, and Helen Singmaster, eds., *Proceedings of Fourteenth Annual Meeting of the Berkeley Linguistics Society*, pp. 35–55. Berkeley: Berkeley Linguistics Society.

Findlay, Jamie. Y. 2017. Multiword expressions and Lexicalism. In Miriam Butt and Tracy Holloway King, eds., *Online Proceedings of the LFG17 Conference*, pp. 209-229.

Frank, Anette. 1996. A note on complex predicate formation: Evidence from auxiliary se- lection, reflexivization, and past participle agreement in French and Italian. In Miriam Butt and Tracy Holloway King, eds., *Online Proceedings of the LFG96 Conference*.

Her, One-Soon. 1999. Grammatical Representation of Idiom Chunks. In: *International Association of Chinese Linguistics 8th Annual Conference*, Melbourne, Australia.

Her, One-Soon. 1997. *Interaction and variation in the Chinese VO construction*. The Crane Publishing Co., Ltd, Taipei.

Huang, Chu-Ren. 1990. A Unification-Based LFG Analysis of Lexical Discontinuity. *Linguistics* 28: 263–307.

Huang, C. T. James. 1988. Wo Pao De Kuai and Chinese phrase structure. *Language* 64: 274–311.

Huang, C. T. James. 1984. Phrase structure, lexical ambiguity, and Chinese compounds. *Journal of the Chinese language teachers association* 19(2): 53–78.

Kaplan Ronald. M. and Annie Zaenen. 2003. West-Germanic verb clusters in LFG. In *Verb constructions in German and Dutch*, Pieter Seuren and Gerard Kempen (eds.), pp. 127-150. John Benjamins Publishing Company, Amsterdam, The Netherlands.

Kaplan, Ronald. M. and Jürgen Wedekind. 1993. Restriction and correspondence-based translation. In *Proceedings ofthe 6th Meeting ofthe EACL*. European Chapter of the Association for Computational Linguistics, University of Utrecht.

Kay, Paul and Charles Fillmore. 1999. Grammatical Constructions and Linguistic Generalizations: the What's X doing Y? Construction. *Language* 75: 1–33.

Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicag Grammatical Constructions and Linguistic Generalizations: the What's X doing Y? Construction o Press.

Li, Charles N. and Sandra A. Thompson. 1981. *Mandarin Chinese: A functional reference grammar*. Univ of Calif. Press, CA.

Li, G. 2009. On the Possessive Object Construction "VN de O". *Chinese Language Learning* (3): 63–69 (in Chinese).

Megerdoomian, Karine. 2012. The status of the nominal in Persian complex predicates. *Natural Language & Linguistic Theory*, Volume 30, Issue 1, pp. 179-216.

Mohanan, Tara. 1997. Multidimensionality of representation: NV complex predicates in Hindi. In Alsina, A., Bresnan, J. & Sells, P. (eds.), *Complex predicates*. Stanford, CA: CSLI Publications: 431–472.

Mohanan, Tara. 1995. Wordhood and lexicality: Noun incorporation in Hindi. *Natural Language and Linguistic Theory* 13 (I): 75-134.

Nunberg, Geoffrey, Ivan A. Sag and Thomas Wasow. 1994. Idioms. *Language*, Vol. 70, No. 3. pp. 491-538.

Ouhalla, Jamal. 1999. *Introducing Transformational Grammar: From Principles and Parameters to Minimalism*. Edward Arnold, London.

Tang, Sze-Wing. 2000. Word Order in Natural Languages and the Theory of Phrase Structure. *Contemporary Linguistics* 2: 138–154 (in Chinese).

Wasow, Thomas, Ivan A. Sag and Geoffrey Nunberg. 1983. Idioms: an interim report. In *Proceedings of the XIIIth International Congress of Linguistics*, S. Hattori and K. Inoue (eds.), pp. 102-115.

Zhuang, Huibin, Zhenqian Liu and Yuan Zhang. 2013. VO Verbal Compounds and the Realization of Their Objects. In: Ji D., Xiao G. (eds) *Chinese Lexical Semantics. CLSW 2012*, pp. 268-279. Lecture Notes in Computer Science, vol 7717. Springer, Berlin, Heidelberg.

The Malefactive Topic Role in Cantonese Indirect Passives

Pui Lun Chow
The University of Hong Kong

Proceedings of the LFG'18 Conference

University of Vienna

Miriam Butt, Tracy Holloway King (Editors)

2018

CSLI Publications

pages 191–207

http://csli-publications.stanford.edu/LFG/2018

Keywords: Cantonese indirect passives, information structure, topic, malefactive role, syntactic valency

Chow, Pui Lun. 2018. The Malefactive Topic Role in Cantonese Indirect Passives. In Butt, Miriam, & King, Tracy Holloway (Eds.), *Proceedings of the LFG'18 Conference, University of Vienna*, 191–207. Stanford, CA: CSLI Publications.

Abstract

In this paper, I aim to account for the formation of an under-studied subtype of Cantonese passives, namely the indirect passives. Having examined the indirect passive constructions in examples from a corpus, I establish that the indirect passive subject carries the information structure role *topic* and this particular syntactic structure is obligatorily associated with adversity. Instead of the patient argument as in canonical passives, a malefactive argument is borne by the sentence initial NP in indirect passives, leading to an extension of syntactic valency. In sum, the indirect passive with the topical part of the patient NP expressed as the subject and the non-topical part remaining as an object-in-situ is an outcome of interaction of information packaging and grammatical relations.

1 Introduction[1]

Unlike English, the agent in Cantonese passives[2] is obligatory. Mapping in canonical passives involves the agent being expressed as a non-core GF OBL_{θ}, making the patient NP the most prominent semantic role bearing the SUBJ function. The subject in this case is the default topic, as default topic is associated with subject in Cantonese (Fung, 2007).

This paper investigates an under-studied subtype of Cantonese passives which has been analyzed in Mandarin (Huang 1999; Kit 1998; Her 2009; Peltomaa 1996, among others), generally known as *Indirect Passives*.[3] In indirect passives, the subject corresponds to part of the patient argument, often the possessor of the object, rather than the active object/patient. I argue that a

[1] I am very grateful to the attendees and the audience of LFG18 for their attention and valuable comments, in particular to Prof. Mary Dalrymple, Prof. Miriam Butt, and Prof. Alex Alsina. I am also grateful to my supervisors Prof. Stephen Matthews and Dr. Olivia Lam for their contributive comments and support. Naturally all errors are my own.

[2] The basic structure of Cantonese passives is SUBJ - bei2 - agent - V - (OBJ). There are two main types of passives: canonical passives and indirect passives.

[3] Indirect passives are also found in Japanese and Vietnamese. However, there are differences between Cantonese indirect passives and those in Japanese and Vietnamese in terms of selectional restrictions of verbs and syntactic relations.

crucial distinction between the subject in canonical and indirect passives involves information packaging. This article is structured as below: in Section 2, I give a brief description of the syntactic information packaging in Cantonese; in Section 3, evidence of *topic*-bearing subjects in indirect passives from coprus data is provided; discussion of the findings is carried out in Section 4, followed by the corresponding structural representations.

2 The *Topic* Role in Cantonese (Passives)

Cantonese, like other Chinese languages, possesses little verbal morphology. The grammatical relations in Cantonese are specified structurally (Berman 1999). Despite its scant morphology, Cantonese allows pro-drop and flexible word order, facilitated by topic particles such as 'ne1 呢' and 'aa^{6} 呀' or prosodic signal, i.e. a pause (Matthews and Yip 2011) (see (1) and (2)).[4]

(1) Gwo^{3} hoi^{2} aa^{6}, dei^{6} tit^{3} zeoi^{3} faai^{3}
過 海 呀, 地 鐵 最 快
Cross sea Sfp underground most fast
'For crossing the harbor, the underground is fastest.'
(Matthews and Yip 2011:78)

(2) Luk^{6} sik^{1} sang^{1} wut^{6}, nei^{5} hoeng^{2} jing^{3} zo^{2} mei^{6}?
綠 色 生 活, 你 嚮 應 咗 未?
Green life-style 2nd respond Perf not-yet
'Green living – have you responded yet?
(Matthews and Yip 2011:77)

Information packaging in Cantonese follows a typical topic-focus arrangement. It is observed that the pre-verbal sentence-initial position, or [Spec S], is the

[4] Symbols and abbreviations used in this paper: * = Ungrammatical; 1st = First Person; 2nd = Second Person; 3rd = Third Person; Ag = Agent; Pt = Patient; CL. = Classifier; DEF = Definite Determiner; NEG = Negation Marker; OBJ = Object; Perf = Perfective Aspect; Pass = Passive marker; PRED = Predicate; SG = Singular; SUBJ = Subject; Th = Theme; Sfp=sentence final particle. The romanization scheme adopted in this paper is based on the one developed by The Linguistic Society of Hong Kong (2002). There are altogether six tones in this scheme: 1 = high level; 2 = high rising; 3 = mid level; 4 = low falling; 5 = low rising; 6 = low level. The tone is marked as superscript of each romanized character.

default position for *topics*, followed by focus (traditionally termed comment), see (3) and (4):

(3)

(3-i) Aa^{3} can^{2} zou^{6} me^{1} aa^{3}
阿 陳 做 咩 呀?
Ah Chan do what Sfp
What happened to Chan?

(3-ii) keoi^{5} bei^{2} jan^{4} caau^{2} zo^{2} jau^{4} jyu^{2} aa^{3}
佢 畀 人 炒 咗 魷 魚 呀
3rd sgPass people fired Sfp
TOPIC FOCUS/COMMENT
He/She was fired.

The information exchange of (1) can be represented as below:

a. pragmatic presupposition: Chan undergo X
b. pragmatic assertion: *X = being fired*
c. focus: *being fired*

(4)

(4-i) Aa^{3} can^{2} bou^{6} ce^{1} gaau^{2}-me^{1} aa^{3}
阿 陳 部 車 搞-咩 呀?
Ah Chan CL car what-happen Sfp
What happened to Chan's car?

(4-ii) bei^{2} jan^{4} zong^{6} zo^{2} aa3
畀 人 撞 咗 呀
Pass person crash Perf Sfp
FOCUS/COMMENT
(Chan's car) was crashed by someone.

The information exchange of (2) can be represented as below:

a. pragmatic presupposition: Chan's car undergo X
b. pragmatic assertion: *X = being crashed by someone*
c. focus: *being crashed by someone*

The topic 'aa^{3} can^{2} bou^{6} ce^{1} 阿陳部車'(Chan's car) in the response in (4-ii) is understood from previous context and is not expressed.

The passive sentences (3-ii) and (4-ii) are typical predicate-focus structures. The sentences express comments about the topic referents which are also the passive subjects, i.e. 'aa^{3} can^{2} 阿陳' (Chan) in (3-ii) and 'aa^{3} can^{2} bou^{6} ce^{1} 阿陳部車' (Chan's car) in (4-ii).

In canonical passives, the sentence-initial subject carries the informational topic *role*.[5] The resultant grammatical function SUBJ carried by the patient via passive mapping rules makes it a topic by default:[6]

Mapping in canonical passives:

(5)

Can^{4} saang^{1}	gaa^{3} ce^{1}	bei^{2}	tung^{4} si^{6}	zong^{6} laan^{6}	zo^{2}
[陳 生	架 車]	[畀	同事	撞 爛	咗]
Mr. Chan	CL car	Pass	colleague	crash-broken	Perf

Mr. Chan's car was crashed by his colleague.

(6)

	tung^{4} si^{6}	can^{4} saang^{1} gaa^{3} ce^{1}
	同事	陳 生 -架 車
	(colleague)	(Mr.Chan's car)
zong^{6} laan^{6}	< Ag	Pt >
撞 爛 passive	|	|
Crash-broken	OBL_{θ}	SUBJ *(default topic)*

In canonical passives, the rearrangement of grammatical function and semantic role mapping is triggered entirely by the coverb '畀 bei^{2}' and a straightforward mapping results according to the passive mapping rules. In the case of an indirect passive like (7), the mapping is more complicated. A motivation which targets only part of the constituent is needed. The claim of this paper is that

[5] I adopt Butt and King's (2000) definition of topic and focus:
TOPIC is old or known information that is relevant in the current context.
FOCUS is new and prominent information.

[6] Default information-structure roles are often associated with particular grammatical functions. In the majority of cases, the default topical GF is the subject. See detailed discussion in (Dalrymple and Nikolaeva (2011) Chapter 5).

indirect passive is the outcome of interaction between the informational roles (i-structure) and functional relations (f-structure).

(7)

Can[4] saang[1]	bei[2]	tung[4] si[6]	zong[6] laan[6] zo[2]	gaa[3] ce[1]
陳 生	[畀	同 事	撞 爛 咗]	架 車
Mr. Chan	Pass	colleague	crash-broken Perf	CL car
Possessor of patient		Agent		head of patient

In the following section, I demonstrate with corpus data that the indirect passive subject carries the discourse function of topic.

3 Corpus Data

Indirect passives cover a small proportion of passive sentences found in the corpus. Out of 61 passive '畀 bei[2]' sentences, only 4 indirect sentences are found.[7] The limited use of indirect passives is hypothesized to be related to its special pragmatic connotations. In (8), I provide an extract of a dialogue from HKCanCorp (Luke and Wong 2015). The passive sentences are underlined for easier reference.

(8)

(S1)

Waak[6] ze[2] di[1] ngai[6] jan[4] gam[2] joeng[2] lam[2] zyu[6] wan[2] go[3] san[1] sai[3] gaai[3] gam[2] joeng[2]

或 者啲 藝 人 嗷 樣 諗 住 搵 個 新 世 界 嗷 樣。

Or Det $\frac{actors}{actresses}$ that way think find CL new world Part Part

Or, the actors/actresses were thinking of living a new life

[7] The Hong Kong Cantonese Corpus (HKCanCorp) (Luke and Wong 2015) consists of 93 recording and approx. 230,000 Chinese words. 471 sentences are found to involve the morpheme 'bei2 畀'. The other uses of 'bei2 畀' discovered in the corpus and their corresponding proportion are listed as follows: (i) as the lexical verb 'give' (29 %); (ii) as the lexical verb 'let' (34 %); (iii) as a preposition marking benefactory role (23 %) (iv) as a particle meaning 'if it were' (1 %).

(S2)

Dim^{2} zi^{2} fat^{1} jin^{4} gaan^{1} jau^{6} m^{4} dak^{1} laak^{3}

點 知 忽 然 間 又 唔 得 嘞。

Unexpectedly suddenly Part NEG work Part

Unexpectedly, it turned out that it did not work.

(S3)

Zik^{1}hai^{6} teng^{1} gong^{2} **gwo^{2} di^{1} jau^{6} waa^{6} ji^{1} gaa^{3} bei^{2} jan^{4} fong^{1} uk^{1} aa^{3}**…

即 係 聽 講 **嗰 啲 又 話 而 家 畀 人 封 屋 呀**…

That is rumor those Part say now PASS people seal house Sfp

Rumor has it that those people had their houses sealed up.

(8)-(S3) is an example of indirect passive: the subject of the VP 'bei^{2} jan^{4} fung^{1} uk^{1} 畀人封屋' (having (their) houses sealed up) is expressed by the demonstrative pronoun 'gwo^{2} di^{1}' (those) which refers to 'di^{1} ngai^{6} jan^{4} 啲藝人' (the actors/actresses) in the previous discourse (S1). This is consistent with the assumption that topics must be referential and may or may not be overtly represented by noun phrases, while foci must be overtly expressed (Dalrymple and Nikolaeva 2011:50). The topichood of the DP 'gwo^{2} di^{1}' (those) (= 'di^{1} ngai^{6} jan^{4} 啲藝人' the actors/actresses) is further supported by the 'what-about' test for topichood (Dalrymple and Nikolaeva 2011).

(9)

(9-i) Gwo^{2}di^{1} ngai^{6} jan^{4} dim^{2} aa^{3}?

嗰 啲 藝 人 點 呀

those $\frac{actors}{actresses}$ how Sfp

What about the actors/actresses?

(9-ii) Zik^{1}hai^{6} teng^{1} gong^{2} gwo^{2}di^{1} jau^{6} waa^{6} ji^{4} gaa^{3} bei^{2}jan^{4} fong^{1} uk^{1} aa^{3}

即 係 聽 講 嗰 啲 又 話 而 家 畀 人 封 屋 呀

That is rumor those Part say now Pass people seal house Sfp

Rumor has it that those people had their houses sealed up.

Following Dalrymple and Nikolaeva (2011), the information exchange can be represented as below:

a. pragmatic presupposition: the actors/actresses undergo X

b. pragmatic assertion: *X = their houses being sealed up*

c. focus: *their houses being sealed up*

Building on the previous context, 'di^{1} ngai^{6} jan^{4} 啲藝人' (the actors/actresses) which is the antecedent of the demonstrative pronoun in the later context is the most salient referent in the utterance. The theme of discussion continues with an indirect passive which has a possessor NP as the subject. A response using the canonical passive with the patient 'ngai^{6} jan^{4}ge^{3}uk^{1} 藝人嘅屋' as the subject (as in (10-ii)) would be odd.

(10)

(10-i) Gwo^{2}di^{1} ngai^{6} jan^{4} dim^{2} aa^{3}?

嗰 啲 藝 人 點 呀

Those $\frac{actors}{actresses}$ how Sfp

How are the actors/actresses?

(10-ii)# di^{1} ngai^{6} jan^{4} ge^{3} uk^{1} ji^{4} gaa^{1} bei^{2} jan^{4} fong^{1} zo^{2} aa^{3}

啲 藝 人 嘅 屋 而 家 畀 人 封 咗 呀…

Det $\frac{actors}{actresses}$ Poss house now Pass people seal Perf Sfp

The actors/actresses' houses have now been sealed up.

The discourse established in (10-i) calls for a response with 'di^{1} ngai^{6} jan^{4} 啲藝人' (the actors/actresses)' as the subject, i.e. *They had their houses sealed*. A direct passive construction (10-ii) causes an inevitable mismatch between the theme of discussion in the context and the subject/topic in the sentence. In response to a 'what-about' question concerning the actors/actresses which are the TOPIC, an indirect passive with only the possessor but not the entire patient NP is preferred. In other words, the possessor NP in the indirect passive construction carries the discourse function TOPIC.

4 Discussion

4.1 The Semantic Restriction

Apart from the essential informational topic feature of the indirect subject, an additional malefactive restriction is imposed on the structure of indirect sentence. Such restriction is also shown in corpus example above. It has been observed in early studies of passives that adversative meaning is associated with indirect passives (Shibatani 1985, Lapolla 1988, Huang 1999 and among others). It is noted in Shibatani (1985:841) that,

> *The affected nature of the passive subject, when strongly felt, may lead to the use of passive morphology/syntax in a situation where the subject is in directly affected by an event. Thus in Korean, Vietnamese, Chinese, and Japanese, the possessor of a body part or an article that is directly affected can stand in subject position in a passive…*

The semantic constraint of indirect passive is applied quite strictly in Cantonese. While canonical passives allow both adversative and non-adversative meanings such as (11) and (12), indirect passive is restricted to adversative events, see (13) and (14).

(11) keoi^{5} fuk^{1} waa^{2} bei^{2} lou^{5} si^{1} tip^{3} tong^{4}
佢 幅 畫 畀 老師 貼 堂
3rd sg CL picture Pass teacher display
His/her picture was displayed by the teacher.
keoi^{5} hou^{2} zung^{1} ji^{3} bei^{2} jan^{4} zaan^{3}
(Matthews & Yip 1994, p.170)

(12) keoi^{5} hou^{2} zung^{1} ji^{3} bei^{2} jan^{4} zaan^{3}
佢 好 鍾 意 畀 人 讚
3rd sg very like Pass people praise
S/he likes being praised so much.

(13)

(13-i) #keoi^{5} bei^{2} lou^{5} si^{1} tip^{3} tong^{4} fuk^{1} waa^{2} (positive event)

佢 畀 老師 貼 堂 幅 畫

3rd sg Pass teacher display CL picture

(The sentence can only be understood as a relative clause, meaning "The painting that is/was displayed …")

(13-ii) keoi^{5} bei^{2} lou^{5} si^{1} pai^{1} ping^{4} fuk^{1} waa^{2}

佢 畀 老 師 批 評 幅 畫

3rd sg Pass teacher criticize CL picture

S/he had his/her picture being criticized by the teacher.

(14) *keoi^{5} zung^{1} ji^{3} bei^{2} jan^{4} zaan^{3} bun^{2} syu^{1}

佢 鍾 意 畀 人 讚 本 書

3rd sg like Pass people praise CL book

(Intended meaning: He/She likes his/her book being praised.)

The emphasis on the affected possessor/sufferer suggests that the marked indirect passive construction is motivated by pragmatic reasons. To sum up, the subjects of indirect passives are believed to carry the discourse role *topic* and are necessarily associated with a malefactive semantic role.

Having established that a negatively affected *topical* subject is a crucial feature of indirect passives, a question that follows is: what are the syntactic relations of the constituents in indirect passives, and what are the consequences of such an analysis?

4.2 Syntactic relation of pre-畀 bei^{2} and post-畀 bei^{2}NP

The passivization rule states that the agent role is expressed as an adjunct or an OBL_{θ} and the patient NP is then expressed as the subject by LMT (Bresnan and Kanerva 1989). A possible relation between the topical pre-畀bei^{2} and post-畀bei^{2} NP (the highlighted NP in (15)) is that they are a discontinuous subject.

(15) (repeated in (8)-(S3))

Zik^{1}hai^{6} teng^{1} gong^{2} **gwo^{2} di^{1}** jau^{6} waa^{6} ji^{1} gaa^{3} bei^{2} jan^{4} fong^{1} **uk^{1} aa^{3}**…

即 係 聽 講 **嗰 啲** 又 話 而 家 畀 人 封 **屋** **呀**…

That is rumor those Part say now PASS people seal house Part

Rumor has it that those people had their houses sealed up.

The hypothesis can be tested by restoring the 'original' structure, i.e. putting the two NPs together. Consider a construction with discontinuous NP in Cantonese:

(16) **taai^{4} zi^{2}** ngo^{5} sik^{6} zo^{2} **saam^{1} lap^{1}** (discontinuous NP)

提 子 我 食 咗 三 粒

Grapes 1stsg eat Perf three CL

For grapes, I have eaten three.

(17) ngo^{5} sik^{6} zo^{2} **saam^{1}lap^{1}taai^{4} zi^{2}** (restored construction)

我 食 咗 **三 粒 提 子**

1stsg eat Perf three CL grapes

I have eaten three grapes.

Going back to (15), the demonstrative pronoun '嗰啲 gwo^{2} di^{1}' (those) refers to the actors or actresses ('di^{1} ngai^{6} jan^{4} 啲藝人') in the context. It is discovered that the possessor-possessee NP formed by the two NPs is not a grammatical one, see (18):

(18) *gwo^{2} di^{1} ngai^{6} jan^{4} uk^{1} jau^{6} waa^{6} ji^{1} gaa^{3} bei^{2} jan^{4} fong^{1}

嗰 啲 藝 人 屋 又 話 而 家 畀 人 封

those actors/actresses house Part say now PASS people seal

(Intended meaning: the actors'/actresses' houses are said to be sealed.)

Failing to form a grammatical NP in the 'restoration' test suggests that the pre-畀 bei^{2} and post-畀 bei^{2} NP are not a discontinuous subject.

Diagnosis of grammatical status of constituents in Cantonese is rather difficult as Cantonese has very little morphological marking. Constructions of the same type in other languages serve as a good pointer for this kind of unclarity. Indirect passives in Japansese and Korean are two good pointers in this matter. Consider the indirect passives in Japanese and Korean below:

(19) Ken$_i$-wa tomodachi-ni zitensya$_i$-o kowas-are-ta
Ken-TOP friend-DAT bike-ACC break-PASS-PAST
Ken had (his) bike broken by his friends.
(Ishizuka 2010)

(20) Ken$_i$-ga Naomi-ni kao$_i$-o tatak-are-ta
Ken-NOM Naomi-DAT face-ACC hit-PASS-PAST
Ken was hit in the face by Naomi.
(Ishizuka 2010)

(21) haksayng-i sensayngnim-eykey son-ul cap-hi-ess-ta
student-nom teacher-dat hand-acc catch-pass-past-decl.
The student had his hand caught by the teacher.
(The student was caught by the hand by the teacher.)
(Huang 1999, p.52)

In the passive constructions in (19)-(20), the possessors are topicalized and marked by either the topic marker 'wa' as in (19), i.e. Ken-***wa***, or the nominative marker 'ga' as in (20), i.e. Ken-***ga***. The heads of the patients are marked by the accusative marker, i.e. zitensya-***o*** (bike) in (19) and kao-***o*** (face) in (20). The phenomenon is also found in Korean indirect passives. In (21), the possessor of the patient is marked by the nominative marker '***i***', i.e. haksayng-***i*** (the student) whereas the head is marked by the accusative marker '***ul***', i.e. son-***ul*** (hand). It is shown clearly by the case markers on the head of the patient NP in Japanese and Korean indirect passive constructions that the non-topical part of the patient retains its object status.

So far, we have established the grammatical relations and semantic roles of an indirect passive construction as below:

(22)

	NP_1	bei[2]	NP_2	V	NP_3
grammatical functions :	SUBJ		OBL_θ		OBJ
semantic roles :	malefactive		agent		patient

4.3 Structural Representation

As established in the previous section, the indirect passive subject which is usually the possessor of the post-畀 bei^{2} NP is associated with a malefactive and the post-畀 bei^{2} NP is an object. An important point about these features of indirect passives in Cantonese is that they are associated with this particular 'NP_1-畀 bei^{2}-NP_2-V-NP_3' structure. In other words, there is a change in syntactic valency associated with this structure. There is an extension of valency of the PRED from two to three because of the additional malefactive topic role.[8] The theta-role assignment and f-structure of an indirect passive construction is proposed as below:

(23) $< \frac{MALEFACTIVE \quad AGENT \quad PATIENT}{SUBJ \qquad \text{OBL}\theta \qquad \text{OBJ}} >$

(24)

(↑ PRED) = verb < SUBJ OBL$_\theta$ OBJ >

(↑ VOICE) = PASSIVE

(↑FORM) = bei^{2}

(↑ TOP) = SUBJ

Take (7) (repeated in (25)) in Section 1 as an example. The possessor 'can^{4} saang^{1} 陳生' (Mr. Chan) of the patient 'can^{4} saang^{1}gaa^{1} ce^{1} 陳生架車' (Mr. Chan's car) is associated with a topic role in the i-structure. The corresponding lexical entry specification of 'Mr. Chan' is represented as (26):

(25) Can^{4} saang^{1} bei^{2} tung^{4} si^{6} zong^{6}-laan^{6} zo^{2} gaa^{3} ce^{1}

陳 生 畀 同 事 撞 -爛 咗 架 車

Mr. Chan Pass colleague crash-broken Perf CL car

Mr. Chan had his car crashed by his colleague.

[8] We thank an anonymous reviewer for this suggestion.

(26) Lexical specification of 'can^4 saang^1 陳生':

can^4 saang^1

陳生　　N　　(↑PRED) = '陳生'

chan ∈ (↑$_{\sigma 1}$ (↑$_\sigma$ DF))

The functional specification in the lexical entry, i.e. [***chan*** ∈ (↑$_{\sigma 1}$ (↑$_\sigma$ DF))], requires the NP 'can^4 saang^1 陳生' (Mr. Chan) to be a member of discourse functions, i.e. TOPIC in this case. The Extended Coherence Condition (Bresnan and Mchombo 1987) states that:

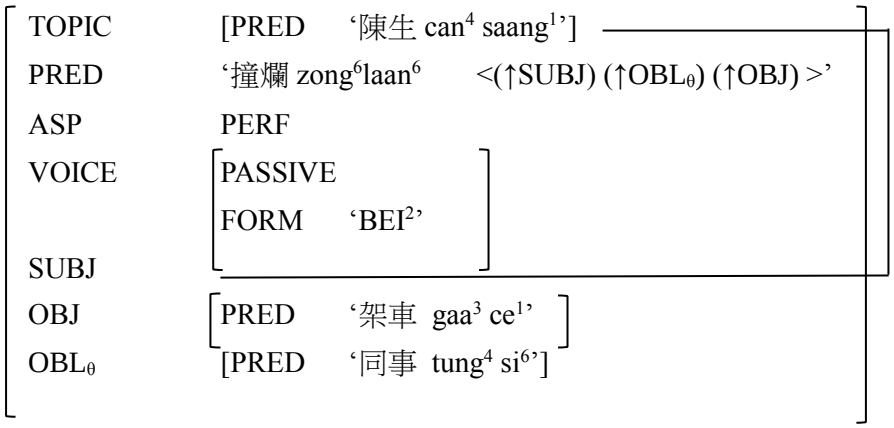
> *FOCUS and TOPIC must be linked to the semantic predicate argument structure of the sentence in which they occur, either by functionally or by anaphorically binding an argument.*

'Can^4 saang^1 陳生' (Mr. Chan) expressed as '***chan***' in the i-structure, is the topic while the rest of the proposition (i.e. what happens to '***chan***') is the focus. The topical NP 'can^4 saang^1 陳生' (Mr. Chan) is linked to SUBJ in the f-structure. According to the *Uniqueness* Condition, each attribute in the f-structure must have a unique value only. With the subject position now being occupied by the possessor 'can^4 saang^1 陳生' (Mr. Chan), the possessee NP 'gaa^3 ce^1 架車' (the car) cannot be mapped to the SUBJ.

Combining the i-structure and the f-structure in (26), the resultant representation of the indirect passive sentence in (25) is shown in (27):

(27)

$$\begin{bmatrix} \textbf{TOPIC \{ } \textbf{\textit{chan}} \textbf{ \}} \\ \\ \textbf{FOCUS \{ } \textbf{\textit{got crashed by colleague - car}} \textbf{ \}} \end{bmatrix}$$



5 Conclusion

In sum, Cantonese indirect passives involve an extension of syntactic valency due to an additional topical malefactive argument. I argue that a crucial distinction between the subject in canonical passives and that in indirect passives lies in the 'contributing factor' of the *topic* role. In canonical passives, the resultant grammatical function SUBJ carried by the patient NP via passive mapping rules is a topic by default and is expressed in sentence initial position, as illustrated in (6). On the other hand, in indirect passives, a NP carrying the malefactive role (not the entire patient) is given a topic role through the context. This particular syntactic structure is the outcome of an interaction between information packaging and grammatical relations.

References

Berman, Judith. 1999. Does German satisfy the subject condition? In Butt and Tracy Holloway King (eds.), *Proceedings of LFG conference99*. Stanford University: CSLI Publication.

Butt, Miriam and Jabeen, Farhat. 2018, July. *Word Order Variation In Urdu/Hindi Wh-Constituent Questions*. Presented at The 23rd International Lexical-Functional Grammar Conference, Vienna, Austria.

Dalrymple, Mary and Nikolaeva, Irina. 2011. *Objects and information structure*. Cambridge: Cambridge University Press.

Fung, Suet Man. 2007. *TOPIC and FOCUS in Cantonese: An OT-LFG Account* MPhil Thesis, The University of Hong Kong.

Her, One-Soon. 2009. Unifying the Long Passive and the Short Passive: On the Bei Construction in Taiwan Mandarin. *Language and Linguistics, 10(3)*, 421-470.

Huang, James C.-T. 1999. Chinese passives in comparative perspective. *Tsing Hua Journal of Chinese Studies* ***29*** **(4)**: 423-509.

Ishizuka, Tomoko. 2010. *Towards a Unified Analysis of Passives in Japanese: A Cartographic Minimalist Approach*. PhD Thesis. University of California.

LaPolla, Randy J. (1988). 'Subject' and referent tracking: Arguments for a discourse-based grammar of Chinese. In by Joseph Emonds, P.J. Mistry, Vida Samiian & Linda Thornburg (eds), *Proceedings of the West Coast Conference on Linguistics*, Volume I, 160-173. Dept. of Linguistics, California State University, Fresno.

Kit, Chunyu. (1998). Ba and bei as multi-valence prepositions in Chinese. In Benjamin K. T'sou (ed.), *Studia Linguistica Sinica*, 497-522. Language Information Sciences Research Centre, City University of Hong Kong.

Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge University Press.

Luke, Kang Kwong and Wong, Lai Yin May. 2015. The Hong Kong Cantonese Corpus, Hong Kong. http://compling.hss.ntu.edu.sg/hkcancor/

Matthews, Stephen and Yip, Virginia. 1994. *Cantonese: A Comprehensive Grammar*. Routledge, New York.

Matthews, Stephen and Yip, Virginia. 2011. *Cantonese: a Comprehensive Grammar*. London: Routledge. 2nd edition

Peltomaa, Marja. 1996. Pragmatic nature of Mandarin passive-like constructions. In Werner A. and Larisa L. (eds.), *Passivization and*

Typology, 83-114, USA: John Benjamin B.V.

Rosen, Victoria. 1998. *Topics and empty pronouns in Vietnamese*. Doctoral Dissertation, University of Bergen.

Shibatani, Masayoshi. 1985. Passives and related constructions: A prototype analysis. *Language, 61,* 821–848.

Glue semantics for Universal Dependencies

Matthew Gotham
University of Oxford

Dag Trygve Truslew Haug
University of Oslo

Proceedings of the LFG'18 Conference

University of Vienna

Miriam Butt, Tracy Holloway King (Editors)

2018

CSLI Publications

pages 208–226

http://csli-publications.stanford.edu/LFG/2018

Keywords: universal dependencies, glue semantics

Gotham, Matthew, & Haug, Dag Trygve Truslew. 2018. Glue semantics for Universal Dependencies. In Butt, Miriam, & King, Tracy Holloway (Eds.), *Proceedings of the LFG'18 Conference, University of Vienna*, 208–226. Stanford, CA: CSLI Publications.

Abstract

Universal Dependencies (UD) is a very widely-used standard for cross-linguistic annotation of syntactic structure. There is, therefore, interest in deriving semantic representations from UD structures, ideally in a language-independent way. In this paper we report on an approach to deriving semantic representations from UD structures that relies on adapting and exploiting techniques from Glue semantics for LFG.

1 Introduction

In recent years, the Universal Dependencies initiative (de Marneffe et al., 2014) has established itself as something of a *de facto* annotation standard for cross-linguistic annotation of syntactic structure (treebank development) and subsequent statistical parsing with models trained on those treebanks. However, many downstream tasks require not dependency parses but rather more elaborate semantic structures that must be derived from those parses. The challenge in any attempt to derive such structures is is to do so while retaining the principal advantages of UD, which means relying as little as possible on language-specific, typically lexical, resources that are not available for many of the 60 languages for which there are UD treebanks.

In this paper we outline an approach to this problem that builds on techniques developed for LFG + Glue. There are several motivations for this. First, LFG's f-structures track the same aspect of syntactic structure as UD dependency trees. Second, the particular version of dependency grammar that UD embodies has inherited much from LFG via the Stanford Dependencies and the PARC dependencies. Third, unlike many other approaches, LFG + Glue does not assume a one-to-one mapping from syntactic to semantic structures, but instead develops a syntax-semantics interface that can map a single syntactic structure to several meaning representations (i.e. the syntax underspecifies the semantics). The latter point becomes especially important because UD—for all its LFG inheritance—is a compromise between theoretical concerns such as language-specific and typological adequacy on the one hand, and computational concerns such as efficient annotation and reliable statistical parsing on the other hand. A typical UD tree therefore contains much less information than the ideal syntactic representations assumed in theoretical formal semantics.

The paper is organized as follows. In Section 2 we describe some relevant properties of UD syntactic structures that illustrate the task we set ourselves. In Section 3 we describe the meaning representation language (version of typed lambda calculus) and meaning composition language (fragment of linear logic) that make up

[†]This work was conducted during the authors' fellowship at the Oslo Center for Advanced Study at the Norwegian Academy of Science and Letters. We gratefully acknowledge their support. We also thank the members of the CAS research group for valuable feedback, and Johan Bos for help with the Boxer software.

the Glue semantics aspect of our proposal. In Section 4 we describe specifically how we connect that Glue theory to UD trees of the form described in Section 2. In Section 5 we discuss the strong points and limitations of our proposal, and point the way to future work.

2 The challenge of UD syntax

In common with other dependency grammar formalisms, UD structures always form a rooted tree over the tokens of the sentence. This means that every node corresponds to an overt token of the sentence (the **overt token constraint**) and has exactly one mother (the **single head constraint**), unless it is the root, in which case it has no mother.[1]

Both the single-head constraint and the overt token constraint limit the expressivity of the syntactic formalism in a way that impairs meaning construction. We illustrate this point on the basis of three example sentences that we will keep returning to, (1)–(3). These are chosen to illustrate different challenges faced by our endeavour and exemplify, respectively, control, VP coordination and a bare relative clause.

(1) Abrams persuaded the dog to bark.

(2) He laughed and smiled.

(3) The dog they thought we admired barks.

Figure 1 shows the UD annotation of (1).[2] The single head constraint makes it impossible to express that *the dog* is simultaneously the object of *persuaded* and the subject of *to bark*, and the overt token constraint makes it impossible to insert any other subject for *bark*, e.g. a PRO that could be coindexed with *the dog*. Compare this with the richer f-structure formalism of LFG, where the relevant information could be captured either through structure sharing (functional control) or a coindexed PRO (obligatory anaphoric control).

For similar reasons, there is no way to indicate in the UD annotation of (2), shown in Figure 2, that *he* is the subject of *smiled*. Nor is there any way to indicate the position of the gap in the UD annotation of a relative clause structure without a relativizer (relative pronoun or complementizer), such as Figure 3, which is the annotation of (3).

[1] These constraints do not apply to *enhanced* UD (http://universaldependencies.org/u/overview/enhanced-syntax.html), which is also part of the Universal Dependencies initiative. However, we restrict ourselves to basic (non-enhanced) UD because (i) enhanced dependency annotations are only available for a very small proportion of the UD treebanks, and (ii) state of the art parsing speed and accuracy is significantly worse for enhanced UD than for basic UD. As such, as things currently stand enhanced UD lacks two of the major attractions of UD.

[2] In these annotations, the first line of text shows the tokens, the second line the lemmas, and the third line the parts of speech. Other information (for example, features) is included in full UD annotations.

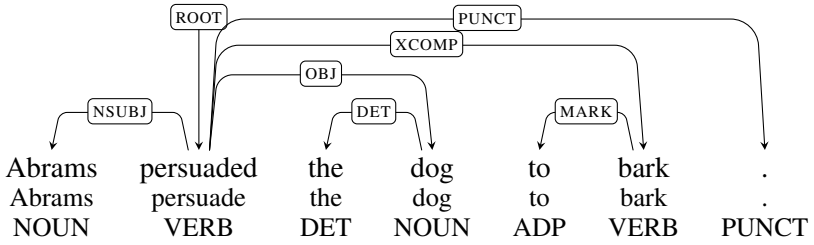


Figure 1: The UD annotation of (1)

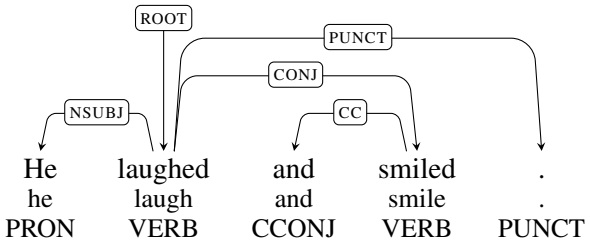


Figure 2: The UD annotation of (2)

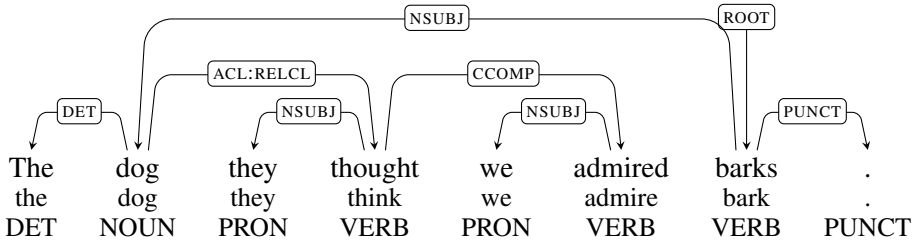


Figure 3: The UD annotation of (3)

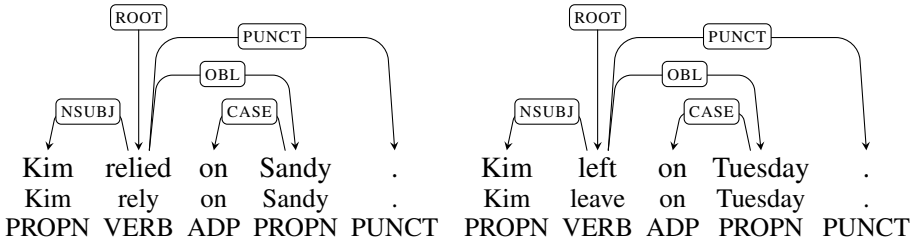


Figure 4: The UD annotations of (4)–(5)

We should also mention at this point that some UD design choices pose challenges from the perspective of semantics that are not shared with other dependency formalisms. For example, the UD annotation guidelines leave no room for an argument/adjunct distinction: both kinds of dependency can be annotated with the relation OBL(ique). An example is shown in Figure 4, which gives the UD annotations of (4)–(5). It can be seen that the two annotations have exactly the same edges and parts of speech, despite the fact that one '*on* phrase' is an argument of the main verb, while the other is an adjunct.

(4) Kim relied on Sandy.

(5) Kim left on Tuesday.

We will return to the discussion of (4)–(5) in Section 5.

3 Semantics

3.1 Meaning representation

Our target meaning representations are Discourse Representation Structures, the format of which is inspired by Boxer (Bos, 2008). The most obvious difference from the Boxer format is that we do not have separate DRSs for presupposed and asserted content. Instead, presupposed conditions are marked with the connective ∂, which is the propositional operator that maps TRUE to TRUE and maps FALSE or # (undefined) to # (we are working in a trivalent semantics). Presupposed discourse referents are given as arguments to the predicate ant; basically, this requires the discourse referent to have an antecedent. The predicates pron.he/they/we are sugaring for ant combined with the appropriate gender/number presuppositions.

With these considerations in mind, we can give the target DRSs for (1)–(3) in Figure 5. As in Boxer, we have three sorts for discourse referents: entities (x_n), eventualities (e_n) and propositions (p_n).

How are these DRSs put together compositionally? Conceptually, we are assuming an updated version of Partial Compositional Discourse Representation Theory (PCDRT, Haug 2014) in which, for example, the representation of (2) given in Figure 5 is an abbreviation of (6).

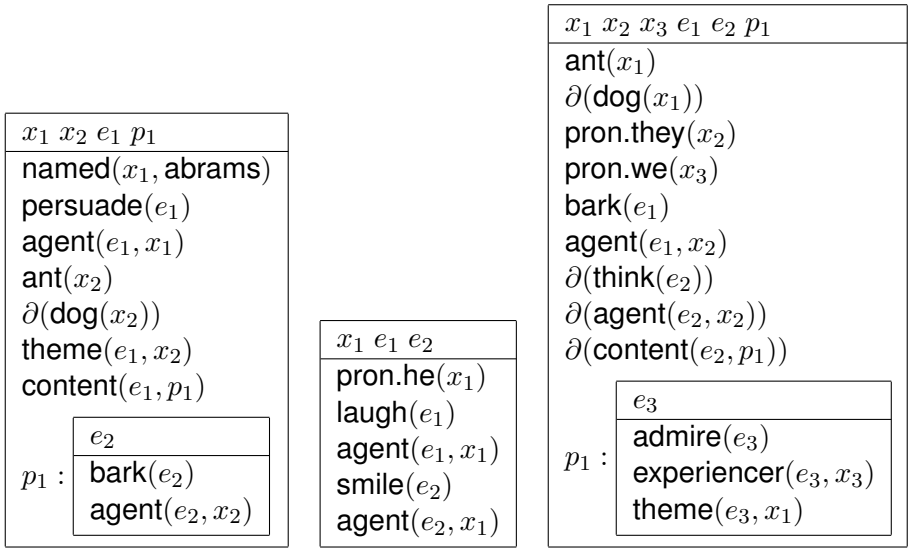


Figure 5: Target meaning representations for (1)–(3)

(6) $\lambda i.\lambda o.\partial(i[x_{i1}\ e_{i1}\ e_{i2}]o) \wedge \mathsf{ant}(o)(x_{i1}) \wedge \partial(\mathsf{male}(\nu(o)(x_{i1})))$
$\wedge\ \mathsf{laugh}(\nu(o)(e_{i1})) \wedge \mathsf{agent}(e_{i1}, x_{i1}) \wedge \mathsf{smile}(e_{i2}) \wedge \mathsf{agent}(e_{i2}, x_{i1})$

That is to say, (6) represents a relation between states i and o such that o extends i by a male individual (identical to one already defined in i) and two events, one of that individual laughing and one of him smiling. Unlike in other approaches, then, the lexical semantics of the word *he* introduces a new discourse referent, albeit one that must be identified with a contextually available discourse referent. This is shown in (7), which is an abbreviation of (8).

(7) $\lambda P.\ \boxed{\begin{array}{|c|}\hline x_1 \\ \hline \mathsf{pron.he}(x_1) \\ \hline \end{array}}\ ; P(x_1)$

(8) $\lambda P.\lambda i.\lambda o.\exists j.\partial(i[x_{i1}]j) \wedge \mathsf{ant}(j)(x_{i1}) \wedge \partial(\mathsf{male}(\nu(j)(x_{i1}))) \wedge P(j)(o)$

We assume PCDRT because:

1. It is defined in typed lambda calculus, and hence is straightforwardly compatible with Glue.

2. It has a treatment of unresolved anaphora, which is essential for an adequate meaning representation for many naturally-occurring examples such as are collected in treebanks.

3. It is representationally similar to standard DRT, allowing for comparison with computational linguistic resources prepared on the basis of DRT.

The assumption of PCDRT is certainly not crucial, however, any theory that meets conditions 1–3 would serve equally well. In our current implementation we use the beta reduction software for λ-DRT described by Blackburn & Bos (2006) and implemented in Boxer (Bos, 2008).

3.2 Meaning composition

On the meaning composition side, we assume a fragment of propositional linear logic that has $\multimap$ as the only connective and three undefined propositional function symbols: e, v and t—mnemonic for entities, eventualities (events and states) and truth values respectively. Following Andrews (2010), we express the fact that certain expressions can take scope at multiple locations by means of an inside-out functional uncertainty (over UD structures), and not in the linear logic fragment itself, which has no quantification. For example, the linear logic type of a quantifier is usually given in higher order glue as (9), where $\uparrow_\sigma$ is the semantic structure of the argument position in which the quantifier occurs, and H is any semantic structure, representing the fact that the quantifier can scope higher than the predicate of which it is an argument. By contrast, in a propositional glue setting, the quantification is replaced by a standard functional uncertainty as in (10).

(9) $\forall H.(\uparrow_\sigma \multimap H) \multimap H$

(10) $(\uparrow_e \multimap (\%H_t)) \multimap \%H_t, \text{GF*} \uparrow = \%H$

Using propositional glue makes it easier to exploit existing tools for linear logic.

In our lexicon, we will assign interpretations (and accompanying linear logic formulae) both to nodes and to edges of UD structures. The up and down arrows then should be read as shown in (11).[3]

(11)

	$\downarrow$	$\uparrow$
node	this node	this node's mother
edge	this edge's target	this edge's source

In our descriptions of linear logic formulae used in lexical entries we also make use of the Kleene star * and local names. One such local name, %R, is special in that it always picks out the root node in a dependency structure—this is how we replicate the treatment of proper names given by Kamp & Reyle (1993) insofar as they always take widest scope and hence remain as accessible antecedents for pronouns.

3.3 The form of lexical knowledge assumed

As alluded to above, we are attempting to give our semantics for UD in such a way as to postpone as much as possible the need for language-specific lexical knowl-

[3] In the actual UD encoding, edges are treated as features of their target nodes, and so the bifurcation given in (11) is strictly speaking unnecessary.

edge. This means that our lexical entries are underspecified in various respects. The principle can best be illustrated by way of an example such as (12),[4] which is what 'admired' retrieves from the lexicon when instantiated as in Figure 3 (there will be more on how this works in the Section 4).

(12) $\lambda x.\lambda F.\left[\begin{array}{|l|}\hline e_1 \\ \hline \mathsf{admire}(e_1) \\ \mathsf{nsubj}(e_1, x) \\ \hline\end{array}\right]; F(e_1) : e_{\downarrow \text{NSUBJ}} \multimap (v_\downarrow \multimap t_\downarrow) \multimap t_\downarrow$

The F argument leaves the event variable open for further modification; we will explain how existential closure happens in Section 4. There are two main points to note about how (12) is underspecified. Firstly, we are not assuming that we have subcategorization information available for individual verbs. Consequently, (12) involves abstraction over one argument of type e just because the node in Figure 3 has one dependent; we don't have the information that there's a 'missing' object in that structure. Secondly, we are not assuming that we have thematic information available either. The thematic relation name nsubj shown in the DRS is lifted from the label of the arc going from *admired* to *we*. In the same way, the name of the event predicate admire is taken from the lemma of the linguistic token. For many purposes these underspecified representations will have to be more fully fleshed out; we will discuss how we anticipate this working in Section 5.

On the other hand, for our approach to produce anything usable we are going to have to assume some lexical knowledge, specifically that associated with 'logic words' (determiners and conjunctions). For example, the treatment of conjunction as exemplified in the case of (2) depends on each of the meaning constructors shown in (13)–(15) below. The meaning constructor in (13) is triggered by a CONJ edge with a verb target node, while (14) is triggered by a CC edge with a source node the mother of which is a verb—hence, in principle, these are independent of specific lexical knowledge. However, the whole analysis still depends on the meaning constructor given in (15), which is triggered by the lemma *and*.[5] The precise nature of how this triggering works will be described in Section 4.

(13) $\lambda P.\lambda S.\lambda C.\lambda E.C(P(E))(S(C)(E)) : ((v_\downarrow \multimap t_\downarrow) \multimap t_\downarrow) \multimap$
$((t_\uparrow \multimap t_\uparrow \multimap t_\uparrow) \multimap ((v_\uparrow \multimap t_\uparrow) \multimap t_\uparrow)) \multimap$
$(t_\uparrow \multimap t_\uparrow \multimap t_\uparrow) \multimap (v_\uparrow \multimap t_\uparrow) \multimap t_\uparrow$

(14) $\lambda P.\lambda _.P :$
$((v_{\uparrow\uparrow} \multimap t_{\uparrow\uparrow}) \multimap t_{\uparrow\uparrow}) \multimap (t_{\uparrow\uparrow} \multimap t_{\uparrow\uparrow} \multimap t_{\uparrow\uparrow}) \multimap (v_{\uparrow\uparrow} \multimap t_{\uparrow\uparrow}) \multimap t_{\uparrow\uparrow}$

(15) $\lambda p.\lambda q.p; q : t_{\uparrow\uparrow} \multimap t_{\uparrow\uparrow} \multimap t_{\uparrow\uparrow}$

[4]To save space and improve readability, we write arguments to the propositional functions as subscripts rather than in brackets, e.g. we write '$t_\downarrow$' rather than '$t(\downarrow)$'.

[5]We are directly applying the analysis of coordination given by Asudeh & Crouch (2002), with CC fulfilling the role of the 'seed' conjunct and CONJ fulfilling the role of the non-'seed' conjunct(s).

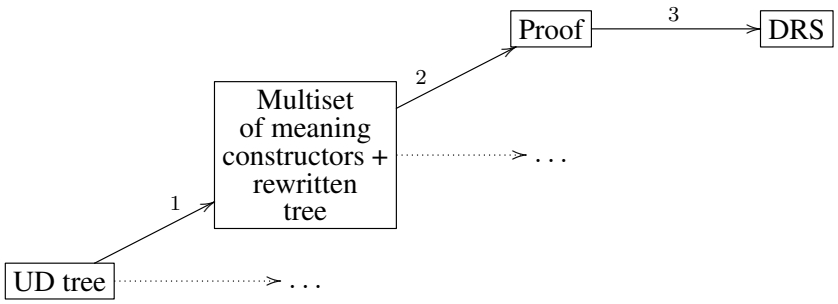


Figure 6: The pipeline

4 Our pipeline

4.1 Overview

The overall architecture of our system is shown in Figure 6. As can be seen, we proceed in three steps. In the first step, the UD tree is simultaneously being (possibly) enriched and rewritten as a multiset of glue-type meaning constructors in a non-deterministic manner. This yields a set of pairs $\langle T, M \rangle$ where T is a (possibly) enriched tree. One should not read too much into this tree enrichment, however: As we will see, the tree delivers the correct types for the meaning constructors and will in many cases show perspicuously which reading the meaning constructors capture, but it does not otherwise play a role in the further processing. The semantic derivation proceeds as in standard glue semantics, by combining the meaning constructors in one or more linear logic proofs (step 2) and then getting a meaning term (in our case, a DRS) via the Curry-Howard isomorphism (step 3). As in standard glue, step 2 is relational (i.e. there can be several different proofs from a single set of premises) but step 3 is functional (i.e. each proof corresponds to a single meaning). Steps 2 and 3 are implemented via Miltiadis Kokkonidis' Instant Glue linear logic prover[6] and Johan Bos's Boxer (Bos, 2008).

The basic idea behind our approach is to traverse the UD tree depth-first and create meaning constructors for each node. As meaning constructors are created at each visited node, the UD tree may be extended non-deterministically; and each of the extended trees are fed to the algorithm. That is, the function f that creates the meaning constructors is of type

(16) $\quad f :: \langle M, T* \rangle \rightarrow [\langle M, T* \rangle]$

where M is a multiset of meaning constructors and $T*$ is a UD tree with a pointer to the current node. The function is recursively applied using the bind operator of the List monad (Haskell's $>>=$).

The output of f is governed by a set of hand-written rules for creating meaning constructors. Each rule has two parts, a criterion, i.e. a tree description that must

[6] http://users.ox.ac.uk/~cpgl0036/prover/glue_prover.pl

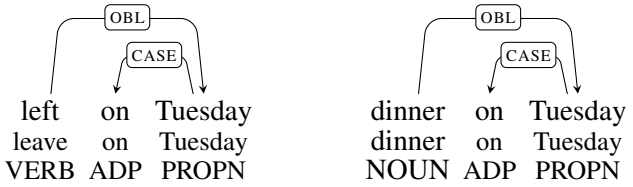


Figure 7: PP attachment to nouns and verbs

evaluate to true at the current node for the rule to apply, and a meaning constructor that will be created if the rule applies. Several meaning constructors can be created from a single node if it maches several criteria. There is a simple control structure in the rules file: rules are matched in the order that they are listed, and stop rules with an empty meaning constructor part will stop the algorithm from searching for more matching rules.

This control structure can be used to encode defaults. For example, in the UD annotation, prepositions are CASE dependents of what would be their complements in a phrase structure analysis. For example, as shown in Figure 7, *on* is a CASE dependent of *Tuesday*, rather than the head of a PP with *Tuesday* as its complement. Semantically, they denote a relation between their mother and their grandmother nodes. While the mother is always of type e, the grandmother can be of either type v or type e, so the type of the preposition is either $\langle e, \langle v, t\rangle\rangle$ or $\langle e, \langle e, t\rangle\rangle$. The following three rules assign the first type if the grandmother is a verb, otherwise the second type.

(17)
```
relation = case; ↑↑ pos = VERB ->
```
$\lambda y.\lambda x.$ [| :LEMMA:(y, x)] : $e_{\uparrow} \multimap v_{\uparrow\uparrow} \multimap t_{\downarrow}$
```
relation = case; ↑↑ pos = VERB ->
relation = case ->
```
$\lambda y.\lambda y.$ [| :LEMMA:(y, x)] : $e_{\uparrow} \multimap e_{\uparrow\uparrow} \multimap t_{\downarrow}$

The first rule matches any node whose relation is CASE and whose grandmother is a verb and assigns the appropriate semantics and type ($\langle e, \langle v, t\rangle\rangle$). The second rule stops further generation of meaning constructors from such nodes, and the final rule then assigns the default type $\langle e, \langle e, t\rangle\rangle$ to any (other) nodes bearing the CASE relation.[7]

Another use of the stop rules is to avoid giving any semantics at all for certain items. For example, there is a stop rule for elements bearing the `PType=Rel`

[7]This was a more or less random choice for our implementation. In a production system, the choice of default could of course affect performance.

feature, which applies *before* the rule that assigns semantics to pronouns, thereby ensuring the relative pronouns are treated as gaps with no semantics.

In the next step, the meaning constructors that come from the rules are *instantiated*. This means that we substitute actual node indices for the $\uparrow$ and $\downarrow$ metavariables. Furthermore, the function :LEMMA: extracts the lemma from the current node. For example, if we assume that the nodes in the left example in Figure 7 have indices 1, 2, 3, the meaning constructor produced by the first rule in (17) will be instantiated as in (18).

(18) $\lambda y.\lambda x.\left[\begin{array}{|c|}\hline \\ \hline \mathsf{on}(y,x) \\ \hline\end{array}\right] : e_2 \multimap v_1 \multimap t_3$

Instantiation can be more complex when we are dealing with verbal nodes. The verb rule looks like (19).

(19) `pos = VERB ->`
$\lambda F.\left[\begin{array}{|c|}\hline e \\ \hline \mathsf{:LEMMA:}(e) \\ \hline\end{array}\right]$; `:DEP:`(e); $F(e) : (v_\downarrow \multimap t_\downarrow) \multimap t_\downarrow$

As mentioned in section 3.3, we do not assume that we have an external valency lexicon available. Instead, we construct the valency from the syntactic tree. The function :DEP: extracts the appropriate semantics from the dependents of the verb according to a separate rule file. For example, if there is an NSUBJ and an OBJ dependent, the rule in (19) will instantiate as in (20), assuming that the indices of the subject, verb and object are 1, 2 and 3 respectively and the verb has the lemma *kiss*.

(20) $\lambda x.\lambda y.\lambda F.\left[\begin{array}{|c|}\hline e \\ \hline \mathsf{kiss}(e) \\ \hline\end{array}\right]$; $\left[\begin{array}{|c|}\hline \\ \hline \mathsf{nsubj}(e,x) \\ \mathsf{obj}(e,y) \\ \hline\end{array}\right]$; $F(e) :$
$e_1 \multimap e_3 \multimap (v_2 \multimap t_2) \multimap t_2$

The F argument here serves as a "handle" for further modification of the event, without making it possible for such modifiers to scope under the event variable. We refer to Champollion (2015) for more details. The semantic composition then ends by saturating F with a property of all events, rather than existential closure of the event variable itself. The relevant meaning constructor is triggered by the ROOT relation and is as in (21), where $\downarrow$ will be instantiated to the index of the root verb, in our case 2.

(21) $\lambda_.\left[\begin{array}{|c|}\hline \\ \hline \\ \hline\end{array}\right] : v_\downarrow \multimap t_\downarrow$

After the meaning constructors have been instantiated, composition can proceed as in ordinary Glue Semantics. We show this in more detail in the next section.

4.2 Worked examples

Let us now see in more detail how we derive the meaning for (1), which has the UD tree in Figure 1.

The first, and most interesting, step is the creation of a meaning for the root node *persuaded*. The relevant rule was given in (19). Instantiation of :DEP: gives (22).

(22) $\lambda P.\lambda y.\lambda x.\lambda F.$
$$\begin{array}{|l|}\hline e_1\ x_1\ p_1 \\ \hline \mathsf{persuade}(e_1) \\ \mathsf{controldep}(e_1, x_2) \\ \mathsf{xcomp}(e_1, p_1) \\ \mathsf{obj}(e_1, y) \\ \mathsf{nsubj}(e_1, x) \\ p_1 : P(x_1)(\lambda_.[\ |\ ]) \\ \hline \end{array} ; F(e_1) :$$
$(e_{\downarrow \text{XCOMP NSUBJ}} \multimap (v_{\downarrow \text{XCOMP}} \multimap t_{\downarrow \text{XCOMP}}) \multimap t_{\downarrow \text{XCOMP}}) \multimap (e_{\downarrow \text{NSUBJ}}) \multimap (e_{\downarrow \text{OBJ}}) \multimap (v_{\downarrow} \multimap t_{\downarrow}) \multimap t_{\downarrow}$

Several noteworthy things happen in (22). First, although we assume no lexical knowledge at this stage in the derivation, we know that *persuade* is a control verb, since it has an XCOMP dependent. But we do not know whether it is a subject or object control verb. Instead we introduce an individual x_1 which bears the relation CONTROLDEP to the matrix event, and is also fed to the embedded controlled predicate as its subject. Notice that CONTROLDEP is a purely semantic relation which does not correspond to anything in the syntactic tree. Furthermore, we introduce a propositional discourse referent p_1 for the proposition we get from feeding the downstairs verb with that subject and closing off the composition (in the way the ROOT relation would do for the matrix verb). This discourse referent bears the XCOMP relation to the matrix event.

Next, the meaning constructor in (22) must be instantiated. If we assume that the nodes in Figure 1 are indexed consecutively from 1, then $\downarrow$ instantiates to 2, $\downarrow$ NSUBJ to 1, $\downarrow$ OBJ to 4 and $\downarrow$ XCOMP to 6. But the linear logic type side of (22) also references the node $\downarrow$ XCOMP NSUBJ, which does not exist. When this happens, the tree is enriched with such a node, yielding the tree in Figure 8. Given this tree, we can instantiate (22) as (23), where 8 is the index of the newly created node.

(23) **persuade:** $\lambda P.\lambda y.\lambda x.\lambda F.$
$$\begin{array}{|l|}\hline e_1\ x_1\ p_1 \\ \hline \mathsf{persuade}(e_1) \\ \mathsf{controldep}(e_1, x_2) \\ \mathsf{xcomp}(e_1, p_1) \\ \mathsf{obj}(e_1, y) \\ \mathsf{nsubj}(e_1, x) \\ p_1 : P(x_1)(\lambda_.[\ |\ ]) \\ \hline \end{array} ; F(e_1) :$$
$(e_8 \multimap (v_6 \multimap t_6) \multimap t_6) \multimap e_4 \multimap e_1 \multimap (v_2 \multimap t_2) \multimap t_2$

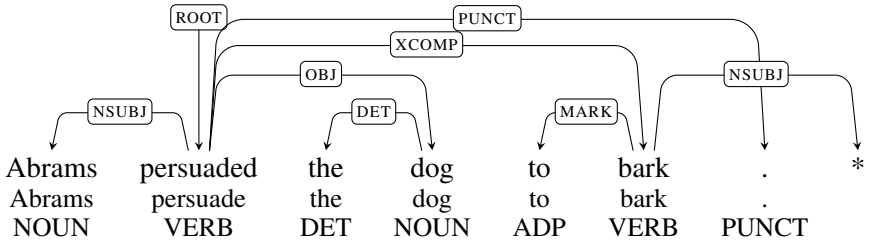


Figure 8: The enriched UD annotation of (1)

The creation and instantiation of meaning constructors for *Abrams*, *the* and *dog* is relatively trivial, keeping in mind that we do allow for use of lexical information about "logic words" such as *the*. The result is shown in (24).

(24) a. **Abrams:** $\lambda P.\boxed{\begin{array}{|l}\hline x_1 \\ \hline \mathsf{named}(x_1, \mathsf{abrams}) \\ \hline\end{array}}; P(x_1) : (e_1 \multimap t_2) \multimap t_2$

b. **the:** $\lambda P.\lambda Q.\boxed{\begin{array}{|l}\hline x_1\ p_1 \\ \hline \mathsf{ant}(x_1) \\ \partial(p_1) \\ p_1 : P(x_1) \\ \hline\end{array}}; Q(x_1) : (e_4 \multimap t_4) \multimap (e_4 \multimap t_2) \multimap t_2$

c. **dog:** $\lambda x.\boxed{\begin{array}{|l}\hline \\ \hline \mathsf{dog}(x_1) \\ \hline\end{array}} : e_4 \multimap t_4$

The meaning for the definite article requires some comment. It introduces a dref x_1. P is the restrictor argument, which in the case of a definite description is presupposed. To capture this, we introduce a propositional dref p_1 for the proposition $P(x_1)$ and put it in the scope of ∂.

The interesting part comes when we reach *bark*. The uninstantiated meaning constructor will be as in (19). And the NSUBJ dependent that was created during the processing of *persuaded* will now make sure that :DEP: triggers a dependency on the subject so that we get (25).

(25) **bark:** $\lambda x.\lambda F.\boxed{\begin{array}{|l}\hline e \\ \hline \mathsf{bark}(e) \\ \hline\end{array}}; \boxed{\begin{array}{|l}\hline \\ \hline \mathsf{nsubj}(e, x) \\ \hline\end{array}} : e_8 \multimap (v_6 \multimap t_6) \multimap t_6)$

Once we have the meaning constructors in (23), (24) and (25), we can assemble them in an ordinary glue proof, as shown in Figure 9. The lambda term corresponding to t_2 in that proof beta reduces to (26).

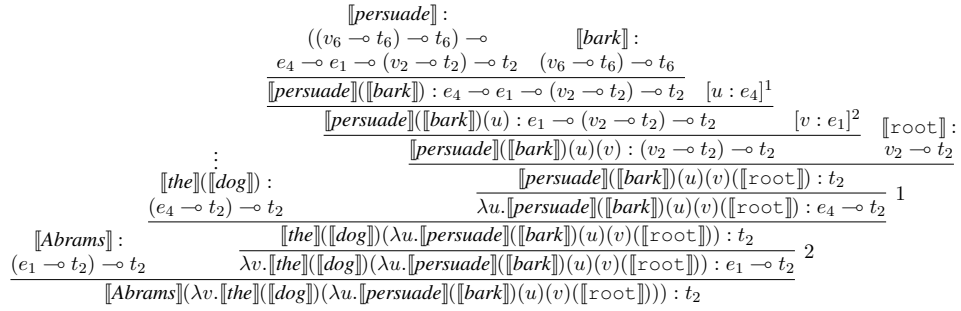


Figure 9: Linear logic proof for (1)

(26)

$x_1\ x_2\ x_3\ e_1\ p_1$
named(x_1, abrams), ant(x_2)
∂(dog(x_2)), persuade(e_1)
nsubj(e_1, x_1), obj(e_1, x_2)
controldep(e_1, x_3), xcomp(e_1, p_1)
p_1 : e_2 | bark(e_2), nsubj(e_2, x_3)

This, then, is the output we derive for Figure 9. In section 5 we will compare it to the target representation that we showed in Figure 5.

To illustrate another feature of our algorithm, we will now see how we derive the correct meaning constructor for bare relative clauses, as in (3), with the UD tree in Figure 3. The challenge in this case is that there is no indication of where the gap in the relative clause is. On the other hand, we do know (by virtue of the ACL:RELCL relation) that there is a gap somewhere. Drawing on the LFG apparatus, we model this using a functional uncertainty. The uninstantiated meaning constructor triggered by ACL:RELCL is as in (27).

(27) $\lambda P.\lambda V.\lambda x.P(x); V(x)(\lambda_.[\,|\,]) :$
$(e_{\uparrow} \multimap t_{\uparrow}) \multimap (e_{\downarrow \text{DEP}^* \text{ DEP}\{PType=Rel\}} \multimap (v_{\downarrow} \multimap t_{\downarrow}) \multimap t_{\downarrow}) \multimap e_{\uparrow} \multimap t_{\uparrow}$

We see that this takes a predicate and a clause with a gap and turns it into a new predicate – in other words, it turns the clause with the gap into a modifier of predicates, as we expect for a relative clause. The type of the clause with the gap is $(e_{\downarrow \text{DEP}^* \text{ DEP}\{PType=Rel\}} \multimap (v_{\downarrow} \multimap t_{\downarrow}) \multimap t_{\downarrow})$, where $(v_{\downarrow} \multimap t_{\downarrow}) \multimap t_{\downarrow})$ is the by now familiar type of verbs. The gap itself is $e_{\downarrow \text{DEP}^* \text{ DEP}\{PType=Rel\}}$, that is, a type e resource associated with some downstairs node (at any level of embedding, hence the Kleene star) bearing the feature PType=Rel. If there is none – as in our case – that node will be created. Let us assume the new node bears the index 9. Then the instantiation of (27) will be as in (28).

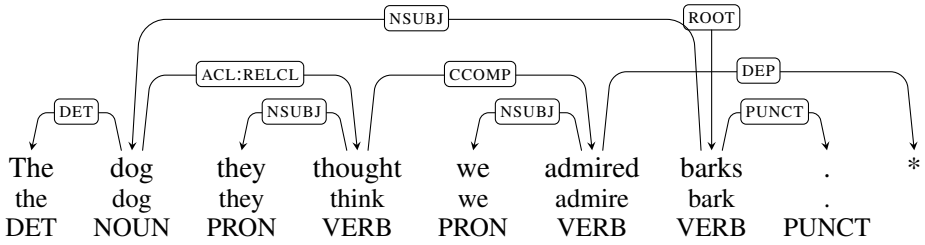


Figure 10: One enrichment of the UD annotation of (3)

(28) $(e_2 \multimap t_2) \multimap (e_9 \multimap (v_4 \multimap t_4) \multimap t_4) \multimap e_2 \multimap t_2$

However, there are many possibilities for situating the new node 9 in the tree, since all we know is that a path of DEP relations leads down to it from the verb of the relative clause. In Figure 10 we show the correct enrichment, with the new node attached under *admired*, but our system generates all four possible attachments (i.e. under *they, thought* and *we* as well as *admired*). Of these, the two readings that attach the gap to a pronoun will fail to produce a correct proof, since there is no interpretation of a gap under a pronoun. But both the attachment under *thought* and *admired* will produce possible meanings as shown in (29).

(29)

$x_1\ x_2\ x_3\ e_1\ e_2\ p_1$	$x_1\ x_2\ x_3\ e_1\ e_2\ p_1$
$\mathsf{ant}(x_1)$	$\mathsf{ant}(x_1)$
$\partial(\mathsf{dog}(x_1))$	$\partial(\mathsf{dog}(x_1))$
$\mathsf{pron.they}(x_2)$	$\mathsf{pron.they}(x_2)$
$\mathsf{pron.we}(x_3)$	$\mathsf{pron.we}(x_3)$
$\mathsf{bark}(e_1)$	$\mathsf{bark}(e_1)$
$\mathsf{nsubj}(e_1, x_2)$	$\mathsf{nsubj}(e_1, x_2)$
$\partial(\mathsf{think}(e_2))$	$\partial(\mathsf{think}(e_2))$
$\partial(\mathsf{nsubj}(e_2, x_2))$	$\partial(\mathsf{nsubj}(e_2, x_2))$
$\partial(\mathsf{ccomp}(e_2, p_1))$	$\partial(\mathsf{ccomp}(e_2, p_1))$
p_1 : e_3	$\partial(\mathsf{dep}(e_2, x_1))$
$\mathsf{admire}(e_3)$	p_1 : e_3
$\mathsf{nsubj}(e_3, x_3)$	$\mathsf{admire}(e_3)$
$\mathsf{dep}(e_3, x_1)$	$\mathsf{nsubj}(e_3, x_3)$

The choice between these two readings can in fact only be made once we use valency information to discard the reading where *thought* takes a subject, a complement clause and a third nominal argument. We return to this point in the next section.

Finally, let us have a look at the coordination example (2), with the UD annotation in Figure (2). The uninstantiated meaning constructors were shown in (13) for the CONJ relation, so ↓= 4 and ↑= 2, (14) for the CC relation (↑= 4, ↓= 3) and

(15) for the lemma *and* (↑= 4). Using these instantiations, we get the meaning constructors in (30)–(32).

(30) **conj:** $\lambda P.\lambda S.\lambda C.\lambda E.C(P(E))(S(C)(E)) : ((v_4 \multimap t_4) \multimap t_4) \multimap$
$((t_2 \multimap t_2 \multimap t_2) \multimap ((v_2 \multimap t_2) \multimap t_2)) \multimap$
$(t_2 \multimap t_2 \multimap t_2) \multimap (v_2 \multimap t_2) \multimap t_2$

(31) **cc:** $\lambda P.\lambda_.P :$
$((v_2 \multimap t_2) \multimap t_2) \multimap (t_2 \multimap t_2 \multimap t_2) \multimap (v_2 \multimap t_2) \multimap t_2$

(32) **and:** $\lambda p.\lambda q.p\,;q : t_2 \multimap t_2 \multimap t_2$

In addition we will have the meaning constructors for **he-laughed** and **smiled** in (33) and (34).

(33) **he-laughed:** $\lambda F.$ $\begin{array}{|l} \hline e_1\ x_1 \\ \hline \mathsf{laugh}(e_1) \\ \mathsf{nsubj}(e_1, x_1) \\ \hline \end{array}$ $;F(e_1) : (v_2 \multimap t_2) \multimap t_2$

(34) **smiled:** $\lambda F.$ $\begin{array}{|l} \hline e_1 \\ \hline \mathsf{smile}(e_1) \\ \hline \end{array}$ $; F(e_1) : (v_4 \multimap t_4) \multimap t_4$

We see that **cc** can take **he-laughed** as its first argument, and **conj** can take **smiled**. Next, **cc**(**he-laughed**) fits as the argument of **conj**(**smiled**). Finally, we can apply the result to **and**, which results in the merger of the two DRSs in (33) and (34), which gives the end result in (35).

(35)
$\begin{array}{|l} \hline x_1\ e_1\ e_2 \\ \hline \mathsf{pron.he}(x_1) \\ \mathsf{laugh}(e_1) \\ \mathsf{nsubj}(e_1, x_1) \\ \mathsf{smile}(e_2) \\ \hline \end{array}$

In effect, what has happened is that, since the UD annotation does not distinguish between VP and sentence coordination, we are forced to treat everything as sentence coordination. Our representation does not therefore capture the fact that *he* is the subject of both verbs.

5 Discussion

Let us take stock at this stage and compare the target meaning representations shown in Figure 5 with what our system gets us so far, shown in (26), (29) and (35). While we anticipate that these underspecified representations will be adequate for many purposes, they are of course lacking plenty of information present in Figure 5. We expect that much of this information can be recovered with some addition of language-specific lexical information at this late stage.

First of all, let us take the θ-role names. As alluded to above, these have simply been lifted from the respective UD edge labels and as such are uninformative. With the aid of meaning postulates such as those shown in (36), however, more informative thematic relations can be inferred.

(36)
$$\forall e \forall x((\textsf{persuade}(e) \wedge \textsf{nsubj}(e,x)) \rightarrow \textsf{agent}(e,x))$$
$$\forall e \forall x((\textsf{persuade}(e) \wedge \textsf{obj}(e,x)) \rightarrow \textsf{theme}(e,x))$$
$$\forall e \forall p((\textsf{persuade}(e) \wedge \textsf{xcomp}(e,p)) \rightarrow \textsf{content}(e,p))$$
$$\forall e \forall x((\textsf{bark}(e) \wedge \textsf{nsubj}(e,x)) \rightarrow \textsf{agent}(e,x))$$
$$\forall e \forall x((\textsf{laugh}(e) \wedge \textsf{nsubj}(e,x)) \rightarrow \textsf{agent}(e,x))$$
$$\forall e \forall x((\textsf{think}(e) \wedge \textsf{nsubj}(e,x)) \rightarrow \textsf{agent}(e,x))$$
$$\forall e \forall p((\textsf{think}(e) \wedge \textsf{ccomp}(e,p)) \rightarrow \textsf{content}(e,p))$$
$$\forall e \forall x((\textsf{admire}(e) \wedge \textsf{nsubj}(e,x)) \rightarrow \textsf{experiencer}(e,x))$$

Next, let us look at the control example. As mentioned above, what has happened in (26) is that the meaning constructor triggered by the token of *persuade* accompanied by an XCOMP dependent has introduced an xcomp relation between the persuading event e_1 and the proposition p_1 that there is a barking event e_2, and introduced an individual x_3 as the nsubj of e_2 and the controldep of e_1. To go further, we need lexical knowledge, specifically the knowledge that *persuade* is an object control verb. That knowledge can be encoded in the meaning postulate shown in (37).

(37)
$$\forall e \forall x((\textsf{persuade}(e) \wedge \textsf{controldep}(e,x)) \rightarrow \textsf{obj}(e,x))$$

The DRS shown in (38) then follow logically from (26) and the meaning postulates given in (36)–(37). If we further assume thematic uniqueness, then we can infer that $x_2 = x_3$ in this case and hence derive a representation equivalent to (1) in Figure 5.

(38)

$x_1\ x_2\ x_3\ e_1\ p_1$
$\textsf{named}(x_1, \textsf{abrams})$, $\textsf{ant}(x_2)$
$\partial(\textsf{dog}(x_2))$, $\textsf{persuade}(e_1)$
$\textsf{agent}(e_1, x_1)$, $\textsf{theme}(e_1, x_2)$
$\textsf{theme}(e_1, x_3)$, $\textsf{content}(e_1, p_1)$
p_1 : $\begin{array}{\|l\|}\hline e_2 \\ \hline \textsf{bark}(e_2) \\ \textsf{agent}(e_2, x_3) \\ \hline \end{array}$

As for the bare relative clause example, if lexical information is to help in selecting the right interpretation of the two shown in (29), and then enriching dep to theme (possibly via obj), the way in which it does so will have to be somewhat less direct than simple entailments on the basis of meaning postulates. We can write one to the effect that every event of admiring has a theme (for example), but that won't in and of itself guarantee that x_1 is that theme, even if we know that it is some

dependent. A different kind of lexical information and/or reasoning process will be needed.

The situation with (2) is similar. We can write a meaning postulate to the effect that every smiling event has an agent, but to get from there to the inference that x_1 is that agent requires a bit more work. The different meanings of *on* are also a difficult case: while *rely* in (4) clearly subcategorizes for *on*, which does not contribute any meaning, it can be hard to reliably guess exactly what meaning of *on* is present in non-subcategorized examples, although the presence of the complement *Tuesday* is a robust cue for a temporal meaning. In all these cases, it might turn out useful to use default reasoning captured in a non-monotonic logic.

Ours is not the first work on semantics for UD; in particular, Reddy (2017) presents a much more developed system. The choice of Glue semantics has certain advantages and disadvantages in comparison with that system: on the plus side, with Glue there is no need for the UD trees to be binarized to get composition off the ground, and we get a flexible approach to scope taking yielding different readings that aren't derivable in a more rigid approach. However, that flexibility comes at a cost: practically it quicky becomes costly to compute lots of uninteresting scope differences, and in terms of design it can be hard to exclude non-existant readings.

In summary, the work described in this paper constitutes a proof of concept tested on carefully crafted examples, where we have applied LFG techniques (functional uncertainties) to enrich underspecified UD syntax, and applied Glue semantics to dependency structures. We have achieved some encouraging results, however we are very far from something practically useful: while we have basic coverage of the UD relations (though not yet VOCATIVE, DISLOCATED, CLF, LIST, PARATAXIS, ORPHAN), there has not yet been much work on interactions, special constructions or real data noise. These, and the limitations we have identified, provide plenty of opportunity for further work in this framework.

References

Andrews, Avery D. 2010. Propositional glue and the projection architecture of LFG. *Linguistics and Philosophy* 33(3). 141–170.

Asudeh, Ash & Richard Crouch. 2002. Coordination and parallelism in Glue semantics. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG02 conference*, 19–39. Stanford, CA: CSLI Publications.

Blackburn, Patrick & Johan Bos. 2006. Working with Discourse Representation Theory. Unpublished book draft. http://www.let.rug.nl/bos/comsem/book2.html.

Bos, Johan. 2008. Wide-coverage semantic analysis with Boxer. In *Proceedings of the 2008 conference on semantics in text processing* STEP '08, 277–286.

Stroudsburg, PA, USA: Association for Computational Linguistics. http://dl.acm.org/citation.cfm?id=1626481.1626503.

Champollion, Lucas. 2015. The interaction of compositional semantics and event semantics. *Linguistics and Philosophy* 38(1). 31–66. doi:10.1007/s10988-014-9162-8. https://doi.org/10.1007/s10988-014-9162-8.

Haug, Dag Trygve Truslew. 2014. Partial dynamic semantics for anaphora. *Journal of Semantics* 31. 457–511.

Kamp, Hans & Uwe Reyle. 1993. *From discourse to logic* (Studies in Linguistics and Philosophy 42). Dordrecht: Kluwer.

de Marneffe, Marie-Catherine, Timothy Dozat, Natalia Silveira, Katri Haverinen, Filip Ginter, Joakim Nivre & Christopher D. Manning. 2014. Universal Stanford dependencies: A cross-linguistic typology. In *LREC*, Reykjavik, Iceland: ELRA.

Reddy, Siva. 2017. *Syntax-mediated semantic parsing*: University of Edinburgh dissertation.

Modelling Possession and Agreement in Hungarian DPs: A paradigmatic approach

Tibor Laczkó
University of Debrecen

Proceedings of the LFG'18 Conference

University of Vienna

Miriam Butt, Tracy Holloway King (Editors)

2018

CSLI Publications

pages 227–247

http://csli-publications.stanford.edu/LFG/2018

Keywords: Hungarian, inflectional morphology, possessive noun phrases, word and paradigm

Laczkó, Tibor. 2018. Modelling Possession and Agreement in Hungarian DPs: A paradigmatic approach. In Butt, Miriam, & King, Tracy Holloway (Eds.), *Proceedings of the LFG'18 Conference, University of Vienna*, 227–247. Stanford, CA: CSLI Publications.

Abstract

In this paper I will present a formal LFG account of the basic morphosyntactic properties of Hungarian possessive DPs. I will argue for a Word and Paradigm mode of analysis (as opposed to the Item and Arrangement and the Item and Process alternatives, frequently used in other theoretical frameworks). In addition, I will discuss the relevant implementational issues.

1 Introduction

Earlier LFG analyses of Hungarian possessive DPs have, for the most part, dealt with some basic (morpho)syntactic issues, for instance, c-structure representation, the grammatical/discourse functions of nominative and dative possessors, the treatment of pro-drop, and the encoding of definiteness in possessive DPs with various types of possessors; see Laczkó (1995), Chisarik & Payne (2001), Charters (2014), and Laczkó (2017), among others. In this paper, I set out to develop a formal LFG analysis of the fundamental morphosyntactic aspects of the behaviour of Hungarian possessive DPs in comparison with some important previous accounts in alternative frameworks. My approach will be along the general Word and Paradigm lines (as opposed to the Item and Arrangement and Item and Process lines) of morphological analysis. I will also discuss implementational issues.

The structure of the paper is as follows. I present the relevant data in section 2. I discuss three previous accounts that represent the three major lines of morphological investigation in section 3. I develop my Word and Paradigm analysis in section 4. I give a brief summary in section 5.

2 The basic facts

In this section, I briefly present the phenomena under investigation.

(A) Hungarian possessive DPs host nominative or dative possessors, see (1a) and (1b).[1]

(1) a. *Kati* *toll-a*
Kate.NOM pen- POSS.3SG
'Kate's pen'

b. *Kati-nak* *a* *toll-a*
Kate-DAT the pen- POSS.3SG
'Kate's pen'

[1] The general pattern is this: NP_{DAT} D NP_{NOM}, i.e. if the definite article is present in the DP, it obligatorily follows the dative possessor, and it obligatorily precedes the nominative possessor; and only one of the two possessors can occur in any single possessive DP (see Laczkó 1995 and Bartos 2000, for instance).

(B) The possessed noun agrees with the possessor, see (1) and (2), and possessor pro-drop is possible (typical), see (2).

(2) *a* *(mi)* *toll-unk*
the we.NOM pen-POSS.1PL
'our pen'

(C) The possessum exhibits rich inflectional behaviour: it is morphologically marked for (i) possession, (ii) number, (iii) agreement with the possessor. In the most transparent (i.e. truly agglutinative) cases, three different (strictly ordered) morphs[2] encode these morphosyntactic features, see (3a). However, descriptively speaking, there are several feature value combinations in the case of which we can only find two overt inflectional elements or one attached to the noun stem, see (3b) and (3c,d), respectively. Note that *-i* is the plural marker of the possessum, see (3a,b,d), and *-k* is the plural marker of ordinary nouns, e.g. *a hajó-k* 'the ships'.[3]

(3) a. *a* *toll-a-i-nk*
the pen-POSS-PL-1PL
'our pens'

b. *a* *toll-a-i*
the pen-POSS-PL.3SG
'her pens'

c. *a* *toll-a*
the pen-POSS.3SG
'her pen'

d. *a* *hajó-i*
the ship-POSS.PL.3SG
'her ships'

In any theoretical framework, a formal analysis of the phenomena presented above is bound to face the following challenge at least: the treatment of morphological forms that are not (fully) agglutinative, see (3b-d) vs. (3a).

[2] Several morphological approaches (see Kiefer 2000 and Bartos 2000 below, for instance) use the classical (abstract) notion of a morpheme, which can be realized by several different (allo)morphs. In my analysis in this paper, I do not assume the existence of morphemes, and I only employ morphs, i.e. overt morphological elements.

[3] In what follows, I will omit POSS from the glosses when it is not relevant for the discussion of the given examples.

3 On some previous analyses

When the morphological composition of a word is not (fully) agglutinative, as in (3b-d), basically there are three strategies that can be followed: *Item and Arrangement* (IA), *Item and Process* (IP), and *Word and Paradigm* (WP); see Spencer (1991, 2004).

A) IA is templatic in nature: it assumes strictly ordered morpheme positions, and, consequently, it needs to admit zero (allo)morphs when there is no full (overt) agglutination.
B) IP, instead, fuses two or more ("underlying") morphemes into a single morph in such cases.
C) WP, by contrast, employs paradigmatic slots, the feature value combinations of which are realized by particular word forms of varied morphological compositions (whether fully agglutinative or not).

Interestingly, there have been analyses of the basic Hungarian possessive morphological phenomena along all the three strategic lines listed above.

(A) Kiefer (2000: 592-593) analyzes the examples in (4) as shown in (5).

(4) a. *ház-a-i-m*
house-POSS-PL-1SG
'my houses'

b. *kalap-om*
hat-POSS.SG.3SG
'my hat'

(5) a.

STEM	POSS	PL	AGR
ház	a	i	m

b.

STEM	POSS	AGR
kalap	∅	om

As (5) demonstrates, this is an IA style templatic approach that assumes the notion of morpheme, and uses zero allomorphs where necessary.[4]

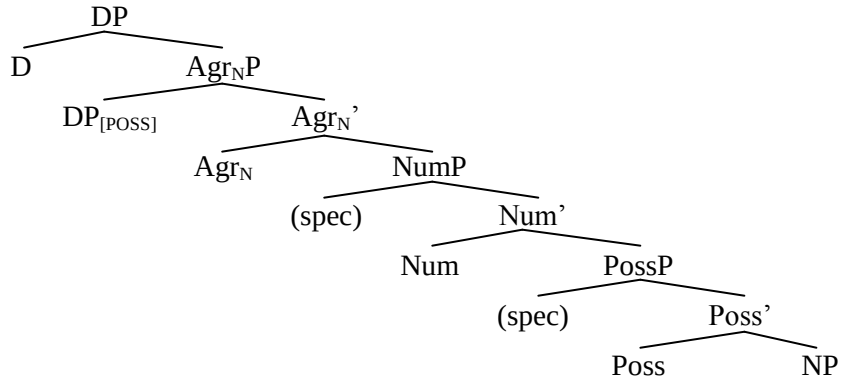
It should be obvious that an LFG analysis cannot adopt (or adapt) an approach along these templatic and zero allomorphic lines as the

[4] It is just a minor point that in Kiefer's system there is no SG morpheme (always to be realized by a zero (allo)morph), see his representation in (5b). Thus, there is no general (SG/PL) templatic slot here: the relevant slot is always PL.

overwhelming majority of approaches in this framework strictly reject empty elements either in syntax or in morphology.[5]

(B) Bartos (2000), in his MP and Distributional Morphology[6] framework, follows the IP strategy. Subscribing to Baker's (1985, 1988) Mirror Principle, he proposes the following internal structure for Hungarian DPs.

(6)



In his system, Poss, Num and Agr_N are suffixal (morphemic) heads.[7] He employs two major operations: morphosyntactic merger and morphological fusion. In the derivation, first Poss and NP are morphosyntactically merged (and, because of the bound morphemic nature of Poss, Poss attaches to the noun head). Next, Num merges with the Poss+NP complex. Finally, AgrN merges with the Num+Poss+NP unit.[8] The appropriate "spell-out" of these abstract morphosyntactic feature value combinations quite often requires morphological fusion (when there is no (full) transparency, i.e. agglutination, and a single morph encodes the feature values of more than one morpheme). The most important aspects of the analysis of (4a) and (4b) in Bartos's system are as follows.

In the case of the transparent (i.e. fully agglutinative) (4a), there are three instances of morphosyntactic merger, see (7a-c), and there is no need for morphological fusion, because each relevant (bound) morpheme has its respective morph realization.

---

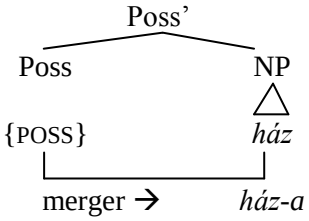
[5] For a salient exception in the domain of syntax; see Bresnan's (2001) treatment of certain long-distance dependency relations by dint of an empty category in c-structure. However, I am not aware of such exceptional treatments in the domain of morphology.

[6] See Halle and Marantz (1993, 1994) and Marantz (1997), for instance.

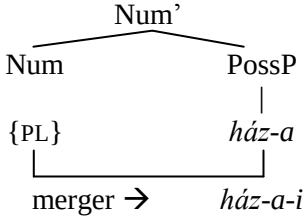
[7] Hungarian morpheme order is appropriately mirrored by the syntactic functional category arrangement in (6).

[8] The syntactic hierarchy of the three relevant functional projections and the merge operation prescribe the order of the morphemes in question attaching to the noun stem, thereby obeying the Mirror Principle (head+Poss+Num+Agr_N).

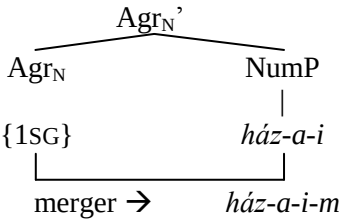
(7) a.



b.



c.



Bartos's analysis of (4b) is as follows (2000: 677).

(8) [$_{word}$ {kalap} + {POSS} + {1SG}] **fusion** →

[$_{word}$ {kalap} + {POSS, 1SG}]

insertion:	*kalap*	*-m*
phonological adjustment:	*kalapo*	*-m*

It should be obvious (again) that an LFG analysis cannot adopt (or adapt) an approach along these (either syntactic or lexical) fusional (IP) lines, given its WYSIWYG principle.

(C) Spencer and Stump (2013) outline an inferential-realizational approach to the inflectional properties of Hungarian possessive noun phrases in Stump's (2001) Paradigm Function Morphology (PFM) framework.[9] It is

[9] The main goal of Spencer and Stump (2013) is to develop a principled morphological analysis of Hungarian inflected (i.e. case-marked) personal pronouns. They exhibit a rather marked behaviour: in terms of their role in the syntax, they are case-marked pronouns; however, morphologically the case marker functions as the stem and it is inflected for person and number just like a possessive noun. Capitalizing on Stump's (2002) extended PFM model (PFM2), this approach

inferential in that its rules deduce complex morphological forms (like *sings*) from more basic forms (like *sing*). It is realizational in that it assumes that the association of an inflected word with its set of morphosyntactic feature values serves as a precondition for (and not a consequence of) its morphological formation.[10] They describe the most crucial aspects of PFM's formalism as follows.

> Each cell in the paradigm of a lexeme corresponds to a particular morphosyntactic property set for which that lexeme inflects; accordingly, each cell in the paradigm of lexeme L is […] formalized as a pairing <R,σ> of L's root R with the morphosyntactic property set σ to which that cell corresponds. A language's paradigm function is therefore a function PF whose domain is the set of such pairings and whose range is the set of realizations for those pairings. We assume that each realization is itself the pairing of a word form w with the morphosyntactic property set that it realizes. That is, where (i) σ is a complete and coherent property set for which lexeme L inflects, (ii) R is a L's root, and (iii) w is the inflected word form realizing lexeme L and the property set σ, PF(<R,σ>) = <w,σ>" (2013: 1222).

For instance, their representation of (9) is as in (10).[11]

---

employs two paradigmatic dimensions: content paradigm (specifying the syntactic properties of the word) and form paradigm (specifying the morphological properties of the word). Normally, there is a default linkage between the two paradigms; however, there can also be a "mismatch" between them, and that is how Hungarian case-marked pronouns can be treated. Given that Spencer and Stump (2013) concentrate on the PFM2 analysis of these miscreant Hungarian pronouns, they do not demonstrate in a detailed fashion how ordinary inflected possessive nouns can be treated in the original PFM model. However, the essence of their presentation and exemplification of the model is fully satisfactory for the purposes of this paper, see below.

[10] "… in a realizational model it is the set of morphosyntatic properties associated with a word form that determines the shape of that word form (what affixes it bears and so on), whereas in an incremental model the affixes themselves bring morphosyntactic properties which are then added as the affixes are added" (2013: 1222).

[11] In (9) I employ the glossing that corresponds to their morphological analysis of this word. Below, I will make a critical remark on an aspect of this analysis. The main point is that *háza* should be appropriately analyzed as a morphologically complex unit: *ház-a* [STEM-POSS].

(9) *háza-i-m-ban*
house-PL-1SG-INE
'in my houses'

(10) PF(<*ház*,σ: {NUM:pl, CASE:inessive, INFL:{PER:1, NUM:sg}}>)
= <*házaimban*,σ>

(10) is the result of the nested application of three realization rules, in other words, these rules operate in three ordered blocks. The realization rules have the following general format.

(11) n, X_C, τ → *f* (X)
where: n = rule-block index
X = variable over the forms to which the rule applies
C = class of forms to which the rule applies
τ = the morphosyntactic property set realized by the application of the rule
f = the morphophonological operation by which the rule realizes τ

The relevant rules for (9) and (10) are as follows (2013: 1223).

(12) a. I, X_N, {NUM:pl, INFL:{PER:*α*, NUM:*β*}} → X'*i*, where X' is the thematized stem of X.
b. II, $X_{[-V]}$, {INFL:PER:1, NUM:sg}} → X*m*
c. III, X_N, {CASE:inessive} → X*ban*

In the first block, the rule adds the special (and formally invariant) plural morph in the possessive paradigm, *-i*, to the "thematized" stem, see (12a). The problem here is that *-a* at the end of *háza* is not part of the stem. There is clear morphophonological evidence that the *-a/-e/-ja/-je* morphs encode possessivity in such words; see Bartos (2000) and Kiefer (2000), among others.[12] From this it follows that for the treatment of case-marked possessive

[12] The problem with Hungarian nouns like *ház* 'house' is as follows. When *ház* takes other inflectional suffixes, crucially the accusative marker (*-t*) and the (nonpossessive) plural marker (*-k*), it can really be argued and assumed that the stem receives a theme vowel (*-a*), see (i) and (ii) – as opposed to plural possessive forms, which should not be analyzed in the same manner, as shown in (iii).

(i) *háza-t* (ii) *háza-k* (iii) *háza-i-m*
house-ACC house-PL house-PL-1SG

However, there are (morphophonologically) classifiable nouns that have a particular theme vowel (*-o*) when the accusative marker or the plural marker is attached to

nouns like (9) a four-block system would be appropriate: POSS+NUM+INFL+CASE.

In the second block in (12) the agreement features (INFL) are encoded. This rule takes *házai* as its input, and it yields *házaim*, see (12b).

The third-block rule adds the case-marker: *házaim* → *házaimban*, see (12c).

PFM handles "zero-morpheme" or fusional cases with the following general rule (2013: 1224), without postulating "real" zero morphemes or fusional operations.

---

them, but in their plural possessive forms their alleged "theme vowel" is *-a*, compare (i)-(iii) with (v)-(vii).

(iv)	*tánc*	(v)	*tánco-t*	(vi)	*tánco-k*	(vii)	*tánca-i-m*
	dance		dance-ACC		dance-PL		dance-PL-1SG

Moreover, there is also a morphophonologically identifiable group of nouns that follow the same "accusative and (nonpossessive) plural theme vowel" pattern; however, in their possessive forms the plural marker (*-i*) is preceded by *-ja* (or *-je*, depending on the rules of vowel harmony), see (ix)-(xi).

(viii)	*bolt*	(ix)	*bolto-t*	(x)	*bolto-k*	(xi)	*boltja-i-m*
	shop		shop-ACC		shop-PL		shop-PL-1SG

On the basis of the patterns above there is a consensus in current morphological approaches (even of varying persuasions) to the effect that the *-a/-e/-ja/-je* (allo)morphs are the markers of possessivity. Consequently, the correct morphological analyses of (iii), (vii) and (xi) are as shown in (xii), (xiii) and (xiv), respectively.

(xii)	*ház-a-i-m*	(xiii)	*tánc-a-i-m*	(xiv)	*bolt-ja-i-m*
	house-POSS-PL-1SG		dance-POSS-PL-1SG		shop-POSS-PL-1SG

Given the larger picture illustrated in (i)-(xiv), it stands to reason that in the case of *ház* 'house' the correct morphological analysis is as shown in (xii), as opposed to (iii), and it simply so happens that *-a* is Janus-faced: it is either a theme vowel, see (i) and (ii), or the POSS marker, see (xii).

Also notice that the "theme vowel" analysis of (iii) is crucially challenged by the existence of the *-ja/-je* allomorphs. In this connection it is especially noteworthy that an ambiguous word form can take either *-a/-e* or *-ja/-je*, and the choice has a disambiguating function, see (xv-xx).

(xv)	*kar-t*	(xvi)	*karo-k*	(xvii)	*kar-ja-i-m*
	arm-ACC		arm-PL		arm-PL-1SG
(xviii)	*kar-t*	(xix)	*karo-k*	(xx)	*kar-a-i-m*
	faculty-ACC		faculty-PL		faculty-PL-1SG

In addition, there is also a significant degree of interspeaker variation as to the choice between *-a/-e* and *-ja/-je* even in the case of unambiguous words.

(13) Identity Function Default: *n*, X_U, {} → X

It is assumed that in all languages every rule block *n* has (13) as its default, which is the least specific member of the rule set. It applies unless it is overridden by a more specific rule in the given block, in other words, unless there is a rule operating with a morph encoding a more marked feature value (X belongs to the universal class of stems: X_U, and {} is a subset of σ). Spencer and Stump (2013) use the example of *ház* 'house' as in (14).

(14) *ház*
house.SG.NOM

This word form ("by default") realizes the following paradigm cell: <*ház*, {NUM:sg, INFL:no, CASE:nom}>.

Given the architecture and principles of LFG, an analysis along the general WP lines seems to be most appropriate. In the next section I will explore the theoretical and implementational avenues an LFG approach can take, and I will propose an account in each of these two dimensions. In the discussion I will also make some remarks on Spencer and Stump's (2013) PFM treatment of the inflectional behaviour of Hungarian nouns.

4 Developing an LFG analysis

4.1. The theoretical dimension

As I pointed out in the previous section, an LFG analysis of these Hungarian inflectional phenomena (and such inflectional phenomena across languages in general) is most appropriately developed along the WP morphological lines, given the fundamental properties of the IA, IP and WP approaches, on the one hand, and the basic architecture and principles of LFG, on the other hand.

It seems to me that in theory Spencer and Stump's (2013) PFM treatment of inflection could be directly accommodated in an LFG approach.[13] We could assume that their paradigmatic block rules are lexical redundancy rules that produce fully fledged, fully inflected lexical items. The only significant modification that would be necessary would be adding the [±POSS] feature to

[13] They extensively argue against the MP-Distributional-Morphology style syntactic analysis of inflectional phenomena in general and that of the behaviour of "inflected pronouns" in particular. They claim that morphology belongs to the lexical component of grammar.

NUM, AGR,[14] and CASE, see the relevant discussion in section 3. One of the potential problems with this accommodation pertains to the general design of morphological analysis. Inflectional morphology would be paradigmatic, without morphs having distinct lexical forms, while derivational morphology would be (because it must be) morph(eme)-based. This split, however, could be at least partially justified by claiming that it is a natural consequence of the fundamentally distinct properties of these two major types of morphological process. Despite this fact, in what follows I will argue for a morph-based approach to inflection in (what I claim) a basically paradigmatic spirit. An immediate advantage of this alternative, of course, is that both inflection and derivation can be handled on the same (morph-based) platform. If in all other respects the two approaches are equally plausible and tenable then the one with this platform uniformity should be preferred.

Consider the Hungarian possessive paradigm in (15).

(15)

STEM	**{POSS; NUM; AGR}**	**{POSS; NUM; AGR}**
	{POSS; SG; 1SG}	{POSS; PL; 1SG}
	{POSS; SG; 2SG}	{POSS; PL; 2SG}
	{POSS; SG; 3SG}	{POSS; PL; 3SG}
	{POSS; SG; 1PL}	{POSS; PL; 1PL}
	{POSS; SG; 2PL}	{POSS; PL; 2PL}
	{POSS; SG; 3PL}	{POSS; PL; 3PL}

Three remarks are in order here. (A) Recall that Spencer and Stump (2013) do not assume the POSS feature. (B) My AGR corresponds to their INFL. (C) Given that the treatment of case-marking is not relevant for the purposes of this paper, I will leave it out from the paradigmatic system to be discussed below.

The analysis to be developed here in the paradigmatic spirit is based on a sketchy proposal I made in Laczkó (2001).

Consider the examples in (3) from section 2, repeated here for convenience.

(3) a. *a* *toll-a-i-nk*
the pen-POSS-PL-1PL
'our pens'

b. *a* *toll-a-i*
the pen-POSS-PL.3SG
'her pens'

[14] AGR corresponds to their INFL. I think AGR would be a better label, given that number marking and case marking are also inflectional by nature.

c. *a* *toll-a*
the pen-POSS.3SG
'her pen'

d. *a* *hajó-i*
the ship-POSS.PL.3SG
'her ships'

My approach is morph-based (as opposed to morpheme-based approaches) and it is paradigmatic (WP vs. IA or IP). I assume that a morph (possibly in allomorphic variation) has a customary lexical form representation. However, one and the same morph can contribute partially different feature values to more than one paradigmatic slot. For instance *-a* is simply the marker of possessivity in (3a) and (3b), whereas it encodes possessivity, number and agreement in (3c). The *-i* morph simply encodes the plurality of the possessed noun in (3a), but it also marks agreement in (3b), and it represents all the relevant feature values in (3d): possessivity, number and agreement, see (16).

(16)

STEM	**{POSS; NUM; AGR}**	**{POSS; NUM; AGR}**
toll 'pen' [3a-c]	{POSS; SG; 1SG}	{POSS; PL; 1SG}
	{POSS; SG; 2SG}	{POSS; PL; 2SG}
	{POSS; SG; 3SG} ***a*** [3c]	{POSS; PL; 3SG} ***a+i*** [3b] ***i*** [3d]
hajó 'ship' [3d]	{POSS; SG; 1PL}	{POSS; PL; 1PL} ***a+i+nk*** [3a]
	{POSS; SG; 2PL}	{POSS; PL; 2PL}
	{POSS; SG; 3PL}	{POSS; PL; 3PL}

I capture this by employing functional annotational disjunctions, see (17).

(17) a. *-a* (↑ POSS) [3a,b]
((↑ POSS PERS) = 3 [3c]
(↑ POSS NUM) = SG
((↑ POSS PRED) = 'PRO'))

b. *-i* (↑ POSS) [3a]
(↑ NUM) = PL
((↑ POSS PERS) = 3 [3b,d]
(↑ POSS NUM) = SG
((↑ POSS PRED) = 'PRO'))

The default function of *-a* is to encode possessivity. I represent this with the (↑ POSS) existential constraint, which requires this function to be present in

the f-structure of the noun phrase, see the top disjunct in (17a). The ability of *-a* to additionally mark 3SG AGR is encoded in the lower disjunct in (17a). The *-i* morph always expresses the plurality of a possessed noun, hence the (↑POSS) and (↑NUM) = PL pair of annotations in the top disjunct in (17b).[15] Here, too, its ability to additionally mark 3SG AGR is encoded in the lower disjunct in (17b). The optional (↑ POSS PRED) = 'PRO' annotation in the last line in the lower disjunct in both (17a) and (17b) is the standard LFG device for handling pro-drop.

The use of a particular morph with the appropriate feature value combinations has a set of complex morphophonological constraints that must be incorporated in any approach, whether IA, IP or WP. For instance, the plural possession *-i* must always be preceded by a vowel. If the stem-final vowel is not *-a*, *-e* or *-i*, this morph is simply added to the stem, see (18).

(18) *autó-i-m*
car-POSS.PL-1SG
'my cars'

If the final vowel is *-a* or *-e*, vowel-lengthening takes place (*-a→-á*, *-e→-é*), see (19) and (20).[16]

(19) a. *fa*
tree
'tree'

b. *fá-i-d*
tree-POSS.PL-2SG
'your[SG] trees'

(20) a. *kecske*
goat
'goat'

b. *kecské-i*
goat-POSS.PL.3SG
'her goats'

If the final vowel is *-i*, *-ja/-je* must be inserted between the two *-i*-s, see (21).[17]

---

[15] The plurality of non-possessed nouns is encoded by *-k* (and its allomorphs); therefore, in its lexical form we need to use the following pair of annotations: ~(↑POSS) and (↑NUM) = PL.

[16] This, however, is not restricted to *-i* attaching to nouns ending in *-a* or *-e*. Any suffix (with either an initial consonant or an initial vowel) will trigger this process, see (ii) and (iii). From this it follows that this lengthening requirement must be encoded in the lexical forms of nouns with final *-a* or *-e*.

(i) *fa*
tree
'tree'

(ii) *fá-ban*
tree-INE
'in tree'

(iii) *fá-ért*
tree-CAUSFIN
'for tree'

(21) a. *taxi* b. *taxi-ja-i-nk*
taxi taxi-POSS-PL-1PL
'taxi' 'our taxis'

If the stem ends in a consonant, *-a/-e/-ja/-je* must be inserted before *-i*, see (22) and (23).

(22) a. *busz* b. *busz-a-i-tok*
bus bus-POSS-PL-1PL
'bus' 'your[PL] buses'

(23) a. *kert* b. *kert-je-i-m*
garden garden-POSS-PL-1SG
'garden' 'my gardens'

As these examples show, the system of morphophonological conditions and constraints is rather complex, and they need to be captured in a formally appropriate manner in any approach.[18] For a detailed and comprehensive discussion of these morphophonological phenomena; see Rebrus (2000). In an LFG approach, we need to encode this morphophonological dimension in the lexical forms of the words and the bound morphs involved.

Now (re)consider a quote from Spencer and Stump (2013) in Footnote 10, repeated here for convenience.

> "… in a realizational model it is the set of morphosyntatic properties associated with a word form that determines the shape of that word form (what affixes it bears and so on), whereas in an incremental model the affixes themselves bring morphosyntactic properties which are then added as the affixes are added" (2013: 1222).

In light of this quote, my LFG analysis proposed above may seem to be incremental rather than realizational – at first sight. After all, I use affixes (i.e. suffixes), and they are added one after the other to a word stem by dint of lexical redundancy rules, and, in the strict sense of the word they add morphosyntactic information incrementally. As usual, these redundancy rules

[17] Most probably the insertion of *-ja*/*-je* is triggered by processing factors: if two *-i*-s are adjacent, it may not be easy to acoustically identify the second *-i*, which encodes an important morphosyntactic property: the possessivity feature.

[18] As I have also pointed out, occasionally a lexically ambiguous word (with the same phonological shape) requires *-a/-e* in one reading and *-ja/-je* in another. In addition, there are also speaker-choice differences in the case of a great number of words. All such additional complications need to be appropriately captured in the lexical forms of the words involved.

create fully-fledged (i.e. fully inflected) word forms in the lexical component of the grammar It is also true that when a suffix is added, it contributes a particular morphosyntactic feature value (set).[19] The problem then would be that according to Spencer and Stump (2013) their PFM model is inferential and realizational, and if my approach is incremental rather than realizational, then its paradigmatic nature becomes questionable. My response to such concerns is as follows.

(A) The way in which I envisage the process of developing all the details of an LFG analysis of these Hungarian inflectional phenomena is absolutely paradigmatic, therefore inferential, in nature. First, the relevant paradigmatic system needs to be established, see (15) above.[20] Next, all the attested morphs or morph combinations must be associated with their respective paradigmatic cells. For a simple example, see (16). Then the individual morphs have to be exhaustively characterized with respect to their feature value contributions to (possibly) various cells in their lexical forms. For some examples, see (17). The important point is that this approach is morph-based (and not morpheme-based), and the treatment of morphs is paradigm-driven.

(B) True, the lexical redundancy rules I assume add morphs one after the other (and these morphemes contribute morphosyntactic information incrementally). If several morphs are involved, in an ordinary morphological approach this cannot really be otherwise. However, as I pointed out in (A) above, this incremental information contribution is strictly paradigm-driven.

(C) It is noteworthy in this context that as far as I can tell Spencer and Stump's (2013) PFM analysis and mine are essentially the same in nature, although the formal devices and the rules differ considerably. Their approach, too, is paradigm-driven (not surprisingly, of course, because this is its defining property). In their system paradigm cell satisfaction is carried out by strictly ordered rules that operate in blocks, and these blocks are arranged on the basis of the order of the morph types that contribute the relevant types of feature values. When these rules work one after the other, the newly added morpheme does contribute specific values towards the satisfaction of the requirements of a particular paradigm cell. In my view, this aspect of the PFM approach exhibits the same kind and degree of incrementality as the fundamentally similar aspect of my approach, see (B) above.

(D) Finally, Spencer and Stump's (2013) approach avoids the postulation of zero morphs or fusion by introducing a default rule the essence of which is that in any block of rules the default is that the input and the output forms are

[19] And, as I have shown, the same morph can realize (partially) different value sets, largely determined by its actual morpho(phono)logical environment. This is an important factor for the assessment of the (paradigmatic vs. nonparadigmatic) nature of my analysis.

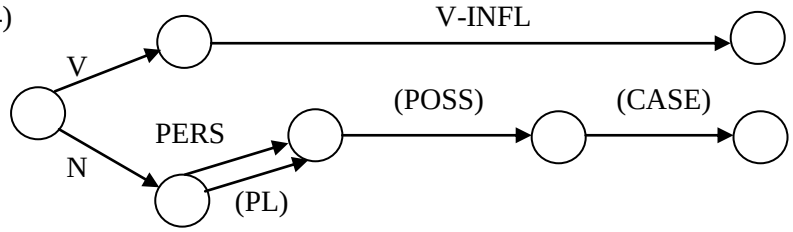
[20] Recall that in Spencer and Stump's (2013) system the POSS feature is not distinguished, see (10) in section 3.

identical morphologically (i.e. no new morph is added), see the Identity Function Default in (13) in section 3; however, the feature value set of the new form is richer: it also realizes the features that the overt morphs in that block contribute. In my system in cases of lack of full agglutination a particular morph is directly associated with the whole relevant set of feature values. Consequently, my approach is less "procedural", as it employs fewer morphological rules/processes, hence it is more paradigmatic in nature in this respect.

4.2. The implementational dimension

In his discussion of the fundamental aspects of developing the morphological component of a computational grammar of Hungarian, Prószéky (2000: 1039) schematizes the system of Hungarian inflectional suffixes on a computational platform as in (24).

(24)



We have verbal (V) and nominal (N) inflection. In the case of verbal inflection, there is just one inflectional-paradigmatic slot (V-INFL). It is filled by morph-complexes encoding value sets for the following feature-complexes {INDICATIVE; TENSE; AGR} or {CONDITIONAL; AGR} or {IMPERATIVE; AGR}.[21] In the case of nominal inflection, the fundamental contrast is that between possessive inflection and nonpossessive inflection. In (24) PERS represents the possessive line, and (PL) the nonpossessive line. PERS in this system stands for a single paradigmatic slot for the {POSS; NUM; AGR} feature complex. On the nonpossessive line NUM is encoded: SG is unmarked, PL is realized by *-k* and its allomorphs (as opposed to *-i* in the possessive paradigm). (POSS) indicates a slot for an optional special pro morph (either in the singular or in the plural) standing for a possessed noun.[22] The final slot is for case markers.

[21] In the case of the AGR feature there are two subparadigms with respect to definite and indefinite object marking in all the three alternative feature complexes.

[22] As in (ii) and (iv) [next page]. Incidentally, this also means that the paradigmatic space in the Hungarian nominal inflectional domain needs to be augmented.

An efficient implementational version of an LFG grammar of Hungarian needs to be developed along the lines of these general computational-morphological assumptions. Given that even morph complexes that are traditionally analyzed as combinations of morphs function as single, unanalyzed morphs individually filling their respective paradigm cells, the representation in (16) of the relevant morphs in the examples in (3) needs to be modified in this approach as shown in (25).

(25)

STEM	**{POSS; NUM; AGR}**	**{POSS; NUM; AGR}**
toll ‘pen’ [3a-c]	{POSS; SG; 1SG}	{POSS; PL; 1SG}
	{POSS; SG; 2SG}	{POSS; PL; 2SG}
	{POSS; SG; 3SG} ***a*** [3c]	{POSS; PL; 3SG} ***ai*** [3b] ***i*** [3d]
hajó ‘ship’ [3d]	{POSS; SG; 1PL}	{POSS; PL; 1PL} ***aink*** [3a]
	{POSS; SG; 2PL}	{POSS; PL; 2PL}
	{POSS; SG; 3PL}	{POSS; PL; 3PL}

(26) and (27) present all the possible allomorphs (as single morphological objects) in their respective paradigm cells, cf. PERS in (24).

(26)

{POSS; NUM; AGR}	STEM	MORPH
{POSS; SG; 1SG}:	*stem*	*m, am, em, om, om*
{POSS; SG; 2SG}:		*d, ad, ed, od, öd*
{POSS; SG; 3SG}:		*a(á), e(é), ja(já), je(jé)*
{POSS; SG; 1PL}:		*nk, unk, ünk*
{POSS; SG; 2PL}:		*tok, tek, tök, atok, etek, ötök*
{POSS; SG; 3PL}:		*uk, ük, juk, jük*

(i) *János* *toll-a*
John.NOM pen-POSS.SG.3SG
‘John’s pen’

(ii) *János-é*
John-POSS_PRO
‘that of John’s’

(iii) *a* *fia-m* *toll-a-i*
the son-POSS.SG.1SG pen-POSS.PL.3SG-PL
‘my son’s pens’

(iv) *a* *fia-m-é-i*
the son-POSS.SG.1SG -POSS.PRO-PL
‘those of my son’s’

(27)

{POSS; NUM; AGR}	STEM	MORPH
{POSS; PL; 1SG}:		*im, aim, eim, jaim, jeim*
{POSS; PL; 2SG}:		*id, aid, eid, jaid, jeid*
{POSS; PL; 3SG}:	*stem*	*i, ai, ei, jai, jei*
{POSS; PL; 1PL}:		*ink, aink, eink, jaink, jeink*
{POSS; PL; 2PL}:		*itok, itek, aitok, eitek, jaitok, jeitek*
{POSS; PL; 3PL}:		*ik, aik, eik, jaik, jeik*

Needless to say, this approach dramatically simplifies the task of the developer of an implemented grammar, and almost astronomically enhances speed and efficiency both in parsing and in generation in this domain of the grammar. Fundamentally, it has to deal with the morphophonological aspects (the conditions and constraints) of a single morph boundary, while a morphosyntactically transparent, fully analytical approach needs to cope with three such boundaries with significant further complications caused by instances of systematic lack of (full) agglutination. In addition, it can be argued that this radically simplified treatment with respect to the number of morphs involved is truly paradigmatic in nature. In this inflectional domain "horizontally" there is a stem and a single inflectional cell (with a particular set of morphosyntactic feature values), and each cell is filled with a single morphological object (an unanalyzed morph and its (also unanalyzed) allomorphs).[23] Of course, the cost is that it cannot formally directly capture the (otherwise attestable) morphosyntactic contributions of parts of (complex but unanalyzed) morphs. This is something an implementational grammar can (happily) live with; see Prószéky (2000), for instance. More theoretically biased approaches favour the morph-separation method; see Spencer and Stump (2013) and my proposal in section 4.1. However, at this point, let me speculate about how these implementational and theoretical biases could possibly be reconciled.

It may be the case that Hungarian native speakers store morph complexes like those in (26) and (27) in their mental lexicon, in addition to the simplex morph forms. If we want to model this, we need to have separate lexical forms for these morph complexes as well. The basic idea would be that each simplex morph (with its allomorphs) would have its own lexical form with

[23] Notice that Spencer and Stump (2013) also postulate a single paradigmatic cell combined with the stem. "For instance, the Hungarian lexeme HÁZ 'house' (root *ház*) has (<*ház*,σ: {NUM:pl, CASE:inessive, INFL:{PER:1, NUM:sg}}>) as one of its cells" (2013: 1222). Two remarks are in order here. (A) Recall again that they do not distinguish the POSS feature. (B) They put the case feature into this single cell, as opposed to Prószéky's (2000) and my view.

the specification of its basic function (e.g. *-ja:* POSS, *-i:* PL, *-m:* 1SG), and all their attested paradigmatic combinations into a complex form (with a complete set of of morphosyntactic feature values) would also have their respective lexical forms. In a significant sense (and as a probably not very far-fetched analogy) this would be similar to compounding: there are simplex words and there are compound (complex) words in the lexicon, the latter consisting of the former. The psychological plausibility of this assumption could be tested in the following way. The production and the processing of inflected possessive nouns would need to be measured with technical devices of high precision. If nouns with a simplex morph in a particular paradigmatic cell were produced and processed at exactly the same speed as nouns with a (more) complex morph in the same cell (e.g. *-i* vs. *-jei*) then this fact would lend considerable support to the assumption I made above.[24] I leave exploring this issue to future research.

5 Conclusion

In this paper, I have developed a WP analysis of Hungarian possessive inflectional phenomena in an LFG framework, by comparing it with previous alternative analyses along the three major lines of morphological investigation (IA, IP and WP).

My analysis is morph-based (and not morpheme-based). The basic idea is that a morph (in various morphological configurations) can contribute partially different sets of morphosyntactic feature values to different paradigmatic cells, largely depending on morphophonological conditions and constraints. This is encoded in its lexical form by means of functional disjunctions.

Capitalizing on Prószéky (2000), I pointed out that from an implementational perspective it is far more efficient to assume that unanalyzed complex morphs (and their allomorphs) fill each paradigmatic cell. Finally, I speculated that this "simplistic" approach can possibly be accommodated even in more theoretically biased analyses.

Acknowledgements

I thank the participants of the LFG18 conference and my two anonymous reviewers for very useful remarks and suggestions, which considerably enhanced the presentational aspects of the paper. Any errors that remain are solely mine.

[24] In addition, I think experimenting across cells would also be important, given that the possible difference in the degree of complexity could be the greatest there (1 vs. 3 elements), e.g.: *-i* vs. *-aim*.

Project no. 111918 (*New approaches in the description of the grammar of Hungarian pronominal)* has been implemented with support provided from the National Research, Development and Innovation Fund of Hungary, financed under the K funding scheme.

References

Baker, Mark. 1985. The Mirror Principle and morphosyntactic explanation. *Lingustic Inquiry* 16: 373-415.

Baker, Mark. 1988. *Incorporation: A Theory of Grammatical Function Changing.* Chicago, IL: University of Chicago Press.

Bartos, Huba. 2000. Az inflexiós jelenségek szintaktikai háttere [The syntactic background of inflectional phenomena]. In: Kiefer, Ferenc. ed. *Strukturális magyar nyelvtan 3. Morfológia.* [Structural Hungarian Grammar, Vol. 3. Morphology]. Budapest: Akadémiai Kiadó, 653-762.

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

Charters, Helen. 2014. Anchor: A DF in DP. In: Butt, Miriam & King, Tracy Holloway. eds. *The Proceedings of the LFG14 Conference*. Stanford, CA: CSLI Publications, 200-220.

Chisarik, Erika & Payne, John. 2001. Modelling possessor constructions in LFG: English and Hungarian. In: Butt, Miriam & King, Tracy Holloway. eds. *The Proceedings of the LFG01 Conference*. Stanford, CA: CSLI Publications.

Halle, Morris & Marantz, Alec. 1993. Distributed morphology and the pieces of inflection. In: Hale, Kenneth & Keyser, Samuel J. eds. *The View from Building 20. Essays in Honor of Sylvain Bromberger.* Cambridge, MA: MIT Press, 111-176.

Halle, Morris & Marantz, Alec. 1994. Some key features of Distributed Morphology. *MIT Working Papers in Linguistics* 21: 275-288.

Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *UPenn WPL* 4(2): 201-225.

Kiefer, Ferenc. 2000. A ragozás [Inflection]. In: Kiefer Ferenc. ed. *Strukturális magyar nyelvtan 3. Morfológia* [Structural Hungarian Grammar 3. Morphology]. Budapest: Akadémiai Kiadó, 569-618.

Laczkó, Tibor. 1995. *The Syntax of Hungarian Noun Phrases – A Lexical-Functional Approach. Metalinguistica 2*. Frankfurt am Main: Peter Lang.

Laczkó, Tibor. 2001. A magyar birtokos DP inflexiós morfológiájáról lexikai-funkcionális megközelítésben [Remarks on the inflectional morphology of Hungarian possessive DPs from a lexical-functional perspective]. In: Bakró-Nagy, Marianne; Bánréti, Zoltán & É. Kiss, Katalin. eds. *Újabb tanulmányok a strukturális magyar nyelvtan és a nyelvtörténet köréből. Kiefer Ferenc tiszteletére barátai és tanítványai* [New Studies on Hungarian Structural Grammar and Historical Linguistics. A Festschrift for Ferenc Kiefer by His Friends and Students]. Budapest: Osiris Kiadó, 59-77.

Laczkó, Tibor. 2010. A new account of possessors and event nominals in Hungarian. In: Alexiadou, Artemis & Rathert, Monika. eds. *The Semantics of Nominalizations across Languages and Frameworks. Interface Explorations 22*. Berlin: Mouton de Gruyter, 81-106.

Laczkó, Tibor. 2017. Modelling (in)definiteness, external possessors and (typological) variation in Hungarian possessive DPs. In: Butt, Miriam & King, Tracy Holloway. eds. *The Proceedings of the LFG17 Conference*. Stanford, CA: CSLI Publications, 243-263.

Prószéky, Gábor. 2000. A magyar morfológia számítógépes kezelése [The computational treatment of Hungarian morphology]. In: Kiefer Ferenc. ed. *Strukturális magyar nyelvtan 3. Morfológia* [Structural Hungarian Grammar 3. Morphology]. Budapest: Akadémiai Kiadó, 1021-1063.

Rebrus, Péter. 2000. Morfofonológiai jelenségek [Morphophonological phenomena]. In: Kiefer Ferenc. ed. *Strukturális magyar nyelvtan 3. Morfológia* [Structural Hungarian Grammar 3. Morphology]. Budapest: Akadémiai Kiadó, 763-947.

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

Spencer, Andrew. 2004. Morphology – an overview of central concepts. In: Sadler, Louisa & Spencer, Andrew. eds. *Projecting Morphology. Stanford Studies in Morphology and the Lexicon.* Stanford, CA: CSLI Publications, 67-109.

Spencer, Andrew J. & Stump, Gregory T. 2013. Hungarian pronominal case and the dichotomy of content and form in inflectional morphology. *Natural Language and Linguistic Theory* 31: 1207-1248.

Stump, Gregory T. 2001. *Inflectional Morphology*. Cambridge: Cambridge University Press.

Stump, Gregory T. 2002. Morphological and syntactic paradigms: Arguments for a theory of paradigm linkage. In: Booij, Geert & Marle, van Jaap. eds. *Yearbook of Morphology*. Dordrecht: Kluwer, 147-180.

Prominent Internal Possessors and Backward Possessor Raising: Norwegian *ryggen på ham* 'the back on him'

Helge Lødrup
University of Oslo

Proceedings of the LFG'18 Conference

University of Vienna

Miriam Butt, Tracy Holloway King (Editors)

2018

CSLI Publications

pages 248–267

http://csli-publications.stanford.edu/LFG/2018

Keywords: Norwegian syntax, body part nouns, possessor raising, backward raising, prominent internal possessors

Lødrup, Helge. 2018. Prominent Internal Possessors and Backward Possessor Raising: Norwegian *ryggen på ham* 'the back on him'. In Butt, Miriam, & King, Tracy Holloway (Eds.), *Proceedings of the LFG'18 Conference, University of Vienna*, 248–267. Stanford, CA: CSLI Publications.

Abstract

Norwegian has a possessor with the preposition *på* 'on' which is used with body part nouns. It shares properties with the dative external possessor of e.g. German and French, but it differs from them in that it can be a part of the noun phrase with the body part noun. I show that the noun phrase-internal possessor should be considered a so-called prominent internal possessor, and propose that it is derived by the application of possessor raising "backward".

1. Introduction[1]

The topic of this paper is Norwegian sentences such as (1)-(2).

(1) De skar dypt i ryggen på ham.
they cut deep in back.DEF on him
'They cut deep in his back.'
(2) De måtte fjerne leveren på ham
they must remove liver.DEF on him
'They had to remove his liver.'

These sentences contain body part nouns whose possessor is expressed in a PP with the preposition *på* 'on'. This possessor shares properties with the dative external possessor in e.g. German or French. Two French examples are (3)-(4).

(3) On lui a tiré dans le ventre.
one him has shot in the stomach
'We shot him in the stomach.'
(4) Je lui casse le bras.
I him break the arm
'I break his arm.'

An important difference from the dative external possessor construction is that the Norwegian possessor PP is - or can be - a part of the noun phrase with the body part noun. The purpose of this paper is to give an analysis of this construction, which was discussed in Lødrup (2009a). The proposal here is that this possessor is a so-called prominent internal possessor, which is derived by what could be seen as backward possessor raising.

[1] Versions of this work have been presented at Forum for Theoretical Linguistics (Oslo, January 2018), LFG18 (Vienna, July 2018), and Syntax of the World's Languages 8 (Paris, September 2018). I grateful for valuable input from the audiences. Special thanks to Joan Bresnan, Dag Haug, and Tanya Nikitina. I would also like to thank the reviewers and the proceedings editors for their good work.

The structure of the paper is the following: Section 2 discusses the European dative external possessor construction, which also existed in Old Norse (section 3). Section 4 shows that Modern Norwegian has the *på* possessor as a part of the body part noun phrase, or as a separate constituent. Section 5 discusses properties of the *på* possessor, and section 6 compares it with partitives. Section 7 discusses the role of affectedness. Sections 8 and 9 show how the *på* possessor should be treated in LFG, with possessor raising applying not only "forward", but also "backward".

2. Dative external possessors

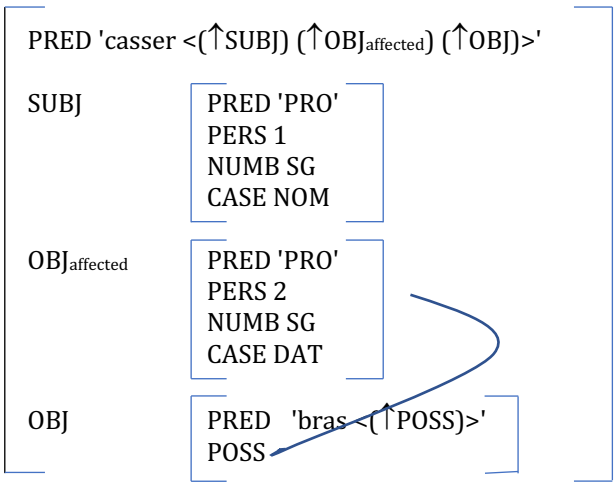
As a background for the Norwegian *på* possessor, it is useful to have a look at the Dative External Possessor - DEP - construction which is found with body part nouns in several European languages (see examples (3)-(4) above).

The DEP construction is rather similar in the European languages that have one (see e.g. König and Haspelmath (1998), Haspelmath (1999) for overviews). The dative is an "extra" argument of the verb, which is realized as an indirect object (LFG's OBJ_θ). It is understood as an affected argument, which means that the verb is understood to have an extra "affected" thematic role. At the same time, the dative is understood as the possessor of a body part noun which is not a part of the same phrase. There are two different ways to account for its relation to the body part noun phrase: A traditional idea is that the possessor is raised from the body part noun phrase to the sentence level (see e.g. Langacker 1968, Lee-Schoenfeld 2006). The alternative is to assume that the dative binds an invisible possessive element in the body part noun phrase (see e.g. Guéron (1985), Hole (2005); and Deal (2017) for general discussion). The former approach is often called possessor raising, and the present paper takes this kind of approach.

In Lexical Functional Grammar (LFG), possessor raising could be implemented as structure sharing. A complication is then that OBJ_θ and POSS do not have identical requirements concerning form - typically OBJ_θ is dative, while POSS is genitive. This fact can be stipulated using the restriction operator (see e.g. Butt et al. 2003).

The f-structure of sentence (4) above is (5), on the next page.

(5)



A terminological note: In the literature, the term possessor raising is used about two phenomena that are both similar and different. The DEP construction is primarily a European phenomenon (Haspelmath 1999). More common in the languages of the world is the kind of possessor raising that can be found in e.g. *She washed him in the face* (see e.g. Levin 1993:71-72). The possessor is realized as a direct object, or an unaccusative subject. Stump and Yadav (1988:310) call this kind of possessor raising possessor-to-host raising, because the possessor (in derivational terms) takes the underlying function of the body part noun phrase.

3. Diachrony

From a diachronic point of view, the Norwegian *på* possessor is clearly related to the European DEP construction. Old Norse had this construction, see Faarlund (2004:170-71). One example is (6.)

(6) konungr steig á bak hesti sínum.
king.NOM mounted on back.ACC horse.DAT his.REFL
'The king got on the back of his horse.'

The possessive dative was not a part of the same phrase as the body part noun (Faarlund 2004:111). The dative could precede the body part noun, or follow it (Skard 1951:13).

A PP with the preposition *á* 'on' was used as an alternative to the dative, and later replaced it. This preposition corresponds to Modern Norwegian *på* 'on' - the preposition of the modern construction. The Old Norwegian (7) (from Skard 1951:56) is from the 14. century.

(7) þu skalt ei vita fyr en ek hifuir uppi iliannar a þer.
you shall not know before than I raise up heels on you
'I will throw you upside down before you notice.'

4. The *på* possessor in Modern Norwegian

In Modern Norwegian, the *på* possessor construction is still used with nouns denoting body parts, as in (1)-(2). Swedish and Danish give the impression of being like Norwegian in relevant respects, but this has not been investigated. This possessor PP is mentioned briefly in König and Haspelmath (1998:559), Haspelmath (1999:123), Stolz et al. (2008:231-238), Dahl (2015:168), and discussed more in Lødrup (2009a), Johannessen et al. (2014).

The *på* possessor in Norwegian, Swedish and Danish is assumed to be a constituent of the sentence, and not a part of the body part noun phrase in König and Haspelmath (1998:584), Haspelmath (1999:123), and Dahl (2015:168). However, Lødrup (2009a) shows that standard constituency tests indicate that the body part noun and the *på* possessor can have an analysis as one constituent in Modern Norwegian. The PP can be a part of a phrase that is topicalized to first position, as in (8), which is sufficient evidence for constituency. Clefting is also possible, as in (9).

(8) I ryggen på ham skar de dypt.
in back.DEF on him cut they deep
'In his back, they cut deep.'
(9) Det var i ryggen på ham de skar dypt.
it was in back.DEF on him they cut deep
'It was in his back they cut deep.'

The old situation with the *på* possessor as a separate argument can still be found, however. When the body part noun is the object of the verb, as in (10), constituency tests give evidence that both options are available. In (11) and (12), the *på* possessor is topicalized and clefted with the body part noun, indicating that the sequence is one constituent. In (13) and (14), the body part noun is topicalized and clefted alone, indicating that the body part noun and the *på* possessor are two constituents.[2]

[2] Norwegian could here be compared to French, which can - to some extent - use the dative equivalent PP with *à* with a non-pronominal external possessor.
(i) On a cassé le bras à ce garçon. (Kayne 1975:143)
one has broken the arm on that boy
'We broke that boy's arm.'
Kayne (1975:143–44) argues that the PP is not a part of the phrase headed by the body part noun, using clefting and pronominalization as arguments. This view is accepted in the literature (see e.g. Vergnaud and Zubizarreta 1992: 618).

(10) De måtte fjerne leveren på ham.
they must remove liver.DEF on him
'They had to remove his liver.'
(11) Leveren på ham måtte de fjerne.
liver.DEF on him must they remove
'His liver, they had to remove.'
(12) Det var ikke bare leveren på ham de måtte fjerne.
it was not only liver.DEF on him they must remove
'It was not only his liver that they had to remove.'
(13) Leveren måtte de fjerne på ham.
liver.DEF must they remove on him
'They had to remove his liver.'
(14) Det var ikke bare leveren de måtte fjerne på ham.
it was not only liver.DEF they must remove on him
'It was not only his liver that they had to remove.'

There are also other options for the body part noun and the *på* possessor to be two constituents (Lødrup 2009a). Unaccusatives and passives can have the body part noun as a subject, as in (15)-(16). The *på* possessor is then a separate constituent or a part of the subject.

(15) Hendene skjelver på ham. / Hendene på ham skjelver.
hands.DEF shake on him / hands.DEF on him shake
'His hand are shaking.'
(16) Neglene må klippes på ham. / Neglene på ham må klippes.
nails.DEF must cut.PASS on him / nails.DEF on him must cut.PASS
'His nails must be cut.'

It is possible to find sentences in which the body part noun and the *på* possessor are not continuous. An example is (17).[3]

(17) (Det) dreide hodet trill rundt på ham.
it turned head.DEF completely around on him
'It turned his head around completely.'

When the *på* possessor is a part of the body part noun phrase, I will call the construction the internal *på* possessor construction. The external *på* possessor construction is the one with the *på* possessor and the body part noun as two separate constituents.

[3] Example (17), as well as a number of the example sentences to come, are from www texts, either directly, through Google searches, or indirectly, through the NoWaC corpus of Norwegian web texts. Some orthographic corrections have been made silently.

At some point in time, the *på* possessor must have been reanalyzed from a separate constituent to a part of the body part noun phrase. This kind of reanalysis from external to internal possessor is also known from some other languages, an example is Hungarian (Nikolaeva 2002).[4]

The evidence for two constituents might be dismissed as involving discontinuous constituents. On the other hand, it is not unexpected that there could be a "residue" from the diachronic development from two to one constituent. It will be assumed here that there is an option for a two constituent analysis. This question is not decisive to the present paper, however, because it is the one constituent construction that raises the interesing questions.

5. Properties of the *på* possessor construction

The *på* possessor construction shares important properties with the DEP construction in e.g. German or French. These properties seem to be unrelated to the status of the *på* possessor, as a separate constituent or a part of the noun phrase with the body part noun.

Restrictions on the head noun

In the DEP construction, the central type of possessum is a body-part noun. This group can be extended to some extent, varying between languages (see e.g. König and Haspelmath 1998:531-33). With the *på* possessor, body part nouns and garments - worn by the owner - are the only possible possessums.

(18) De sparket i hodet / *bilen på ham.
they kicked in head.DEF / car.DEF on him
'They kicked his head / car.'

(19) Noen stappa ting ned i buksa på ham.
some put things down in pants.DEF on him
'Somebody put things down his pants.'

[4] König and Haspelmath (1998:587) hint that this development might have taken place in Icelandic. This seems to be correct, judging from examples in Thráinsson (2007:94-95 and Stolz et al. (2008:114-16) where the body part noun and the PP occur together preceding a finite verb. Even so, Icelandic, like Norwegian, must still have an option for a two constituent analysis, because a possessor PP can precede or follow the body part noun, see Lødrup (2009a:242).

A restriction on modification

In the DEP construction, a non-restrictive adjective cannot modify the body part noun (König and Haspelmath 1998:534-36, Guéron 2006:618).[5] Example (20) is from Vergnaud and Zubizarreta (1992:603).

(20) Pierre lui a lavé les mains (*sales).
Pierre him has washed the hands dirty
'Pierre washed his (dirty) hands.'

The *på* possessor construction shows the same restriction, as (21)-(22) show.

(21) Hun vasket grundig (*den skitne) ryggen på ham.
she washed thoroughly the dirty back.DEF on him
'She washed his (dirty) back thoroughly.'
(22) (*Den skitne) ryggen på ham vasket hun grundig.
the dirty back.DEF on him washed she thoroughly
'She washed his (dirty) back thoroughly.'

A restriction on number

In the DEP construction, a body-part noun always occurs in the singular when it denotes something that we have only one of, such as the mouth (see e.g. König and Haspelmath 1998:581-83). If the possessor is plural, there is a distributive reading[6]. Example (23) is from König and Haspelmath (1998:581).

---

[5] This restriction is also known from other constructions that involve definite body part nouns. One case is the construction in (i) in which the subject is a possessor (König and Haspelmath 1998:534-35, Lødrup 2010 on Norwegian). Another case is possessor-to-host raising, as in (ii) (Lødrup 2009a:245).
(i) Han løfter (*de vakre) øynene fra boken.
he raises the pretty eyes.DEF from book.DEF
'He raises his (pretty) eyes from his book.'
(ii) Hun slo ham i (*det skitne) hodet.
she hit him in the dirty head.DEF
'She hit him in the (dirty) head.'

[6] This restriction is also known from other constructions that involve definite body part nouns. One case is the construction in which the subject is a possessor, as in (i) (König and Haspelmath 1998:581-83, Lødrup 2010 on Norwegian). Another case is possessor-to-host raising, as in (ii) (Lødrup 2009a:244).
(i) De nikket med hodet / *hodene.
they nodded with head.DEF / heads.DEF
'They nodded their heads.'
(ii) Hun slo dem i hodet / *hodene.
she hit them in head.DEF / heads.DEF
'She hit them in the head.'

(23) Tu leur as photographié la bouche / *les bouches.
you them have photographed the mouth / the mouths
'You photographed their mouths.'

This is also the case with the *på* possessor construction, as (24)-(25) show.

(24) Hun stappet kaker i munnen / *munnene på dem.
she popped cakes in mouth.DEF / mouths.DEF on them
'She popped cakes into their mouths.'
(25) I munnen / *munnene på dem stappet hun kaker.
in mouth.DEF / mouths.DEF on them popped she cakes
'Into their mouths, she popped cakes.'

A restriction on function

With the DEP construction, the body part noun is usually not a subject, except to some extent with unaccusative and passive verbs (e.g. König and Haspelmath 1998:538-39). This fact is related to the body part noun's interpretation as an affected argument of the verb. Both this restriction and its exception are reflected in the *på* possessor construction, as in (26)-(27).

(26) *Hodet på ham traff et bord. / *Hodet traff et bord på ham.
head.DEF on him hit a table / head.DEF hit a table on him
'His head hit a table.' (e.g. when he fell) [intended]
(27) Neglene på ham må klippes. / Neglene må klippes på ham.
nails.DEF on him must cut.PASS / nails.DEF must cut.PASS on him
'His nails must be cut.'

Restrictions on the predicate

In the DEP construction, there is a requirement that the possessor is affected, which has been discussed several times (see e.g. Haspelmath (1999:112), Hole (2005:220), Deal 2017, Lee-Schoenfeld (2006), Lee-Schoenfeld and Diewald (2014) - Landau (1999) says that Hebrew is different in this respect). The *på* possessor construction seems to have the same requirement, as indicated by the unaccaptability of (28)-(29) with non-affected body part nouns. The affectedness requirement is discussed further in part 7.

(28) *Legene diskuterte ryggen på dem.
doctors.DEF discussed back.DEF on them
'The doctors discussed their backs.' [intended]
(29) *Legene tenkte på ryggen på dem.
doctors.DEF thought of back.DEF on them
'The doctors thought about their backs.' [intended]

Locality

The DEP construction requires that the possessor and the body part noun phrase are arguments of the same clause (see e.g. Guéron (1985:47-49) on French). For the Norwegian case, locality works the same way in the external *på* possessor construction. In the internal construction, the body part noun phrase must be an argument of the relevant verb.

Some of the restrictions mentioned in section 5 seem to be difficult to state in a natural way. It is striking that the *på* possessor imposes the same restrictions independently of its status as external or internal. Furthermore, the internal possessor imposes restrictions on elements that are not local to it in c-structure. This possessor is a modifier of the body part noun, but it is this modifier that restricts e.g. the type of verb (examples 28-29). The analysis proposed in section 9 gives a way of solving these problems.

6. Partitives

It is necessary to distinguish between the *på* possessor and other PPs with *på* 'on'. There are PPs with *på* and a body part noun that might be seen as regular partitives, and/or locatives, with no connection to the topic of this paper. Examples are (30)-(32).

(30) Sår i underlivet på den drepte viste også at ...
wounds in abdomen.DEF on the killed showed also that ..
'Wounds in the abdomen of the murdered person also showed that ...'

(31) Ser forsatt det søte fjeset på babyen jeg drømte om.
see still the sweet face.DEF on baby.DEF I dreamed of
'(I) still see the sweet face of the baby I dreamed about.'

(32) Hodene på dem var jevnhøye med furutoppene. (Fairytale)
heads.DEF on them were on.a.level with pine.tops.DEF
'Their heads were on a level with the tops of the pine trees.' [about trolls]

In the partitive construction, the restrictions described for the *på* possessor construction in section 5 do not apply. The body part noun phrase does not have to be local to the verb (cf. 30), it can be modified non-restrictively (cf. 31), it can be plural (cf. 32), it can be a subject (cf. 32), and there is no affectedness interpretation (cf. 30-32).

7. Affectedness

In Lødrup (2009a), it was claimed that a *på* possessor in the external construction is interpreted as an affected participant in the event - just like a dative possessor - while a *på* possessor in the internal construction is not interpreted as a participant. This was criticized briefly in Eik (2014:52-53). She says that an internal possessor is interpreted as affected in relevant sentences such as (33)-(34) (from Lødrup (2009a). This observation seems to be correct.

(33) Noen stappa ting ned i buksa på ham.
some put things down in pants.DEF on him
'Somebody put things down his pants.'
(34) Jeg sprutet insektmiddel i håret på ham.
I sprayed insecticide in hair.DEF on him
'I sprayed insecticide in his hair.'

It was mentioned in section 5 that the DEP in e.g. German and French has a condition that the possessor must be affected by the action. However, there are different ways of using the term affected in the literature, as pointed out by Lee-Schoenfeld and Diewald (2014:291). Sometimes it is used of a typical patient. However, the use that is interesting in this context is a different one:

> 'taking part in the situation as an empathetic, necessarily animate co-participant', i.e. sharing some features of a typical agent, without, however, being an agent because not having control (Lee-Schoenfeld and Diewald 2014:288)

I assume that the *på* possessor is normally affected by the verbal action, independently of its realization as external or internal.[7] This is difficult to demonstrate conclusively, however, one problem being the boundary between *på* possessors and partitives. In Lødrup (2009a:237), it was claimed that example (35) is an example of a *på* possessor that is not affected by the verbal action.

(35) Det fløy en fugl over hodet på ham.
there flew a bird over head.DEF on him
'A bird flew over his head.'

[7] Some fixed expressions do not require affectedness, e.g.
(i) Jeg tåler ikke trynet på ham.
I stand not snout.DEF on him
'I cannot stand his face.' [i.e. I don't like him]

It is not clear, however, that this PP should be seen as a *på* possessor. One argument is that the noun phrases could be made plural, as in example (36) (Lødrup 2009a:229), which is is not too different in relevant respects. (Remember from section 5 that a body-part noun always occurs in the singular in the *på* possessor construction.)

(36) (De) har avfyrt varselskudd rett over hodene på demonstrantene.
they have fired warning.shots right over heads.DEF on demonstrators.DEF
'They have fired shots of warning above the heads of the demonstrators.'

Lee-Schoenfeld and Diewald (2014:291) say that the dative external possessor is the protoypical strategy for expressing inalienable possession in German. In Norwegian, the *på* possessor construction seems to be more marked and more colloquial. It is especially called for when the action is dramatic, with real consequences for the possessor, as in (37). The event does not have to concern life and death, however, as in (38). There are also more or less fixed expressions with a metaphorical meaning, cf. (39).

(37) Skjær ut innvollene på ham!
cut out guts.DEF on him
'Cut out his guts!'
(38) (De) stakk tunga i munnen på hver sin soldat.
they put tongue.DEF in mouth.DEF on each their.REFL soldier
'They each put their tongue into the mouth of a soldier.'
(39) De kloke ordene gikk rett til hjertet på ham.
the wise words.DEF went straight to heart.DEF on him
'The wise words went straight to his heart.'

A regular possessor could have been used in the above examples. However, the possessor is then not depicted as a participant in the event. For example, replacing the *på* possessor in (37) with regular possessors gives sentences which are not easy to contextualize. They give an impression that cutting out somebody's guts is an ordinary thing to do, or that the possessor's dead body is given an autopsy.

(40) Skjær ut innvollene hans!
cut out guts.DEF his
'Cut out his guts!'
(41) Skjær ut Olas innvoller!
cut out Ola's guts
'Cut out Ola's guts!'

(42) Skjær ut innvollene til Ola!
cut out guts.DEF to Ola
'Cut out Ola's guts!'

In some sentences with a *på* possessor, an alternative with a regular possessor would sound very strange. In some cases, there is also a question of how a counterpart with a regular possessive should look, if the body part noun should be singular or plural, etc. For example, in (43), it is not clear to me which possessive construction could be chosen as an alternative - none of them sound good.

(43) (Jeg vil) sette en pil i nakken på alle som kødder med skogen vår.
I will put an arrow in neck.DEF on all who mess with forest.DEF our
'I will put an arrow in the necks of all people who mess with our forest.'

8. Prominent internal possessors

The *på* possessors that are external to the body part noun phrase are similar to dative possessors, and can be treated the same way grammatically (assuming that the preposition is a kind of grammatical marker). It is the internal *på* possessors that are interesting theoretically. Even if they are not constituents of the clause, they behave as arguments in some respects. They are interpreted as a clausal argument, and they impose grammatical restrictions that apply above their local domain (restricting e.g. the type of verb, and the option of being a subject - see section 5).

Possessors that behave as arguments in some respects are well known from other languages. They are often called prominent internal possessors, and they can be found in a number of unrelated languages (see e.g. Ritchie 2016, 2017). For example, the possessor can play a part in verb agreement. Ritchie (2017:663) gives sentences (44)-(46) from Chimane (unclassified, Bolivia).

(44) Juan täj-je-'$_i$ [un mu' Sergio$_j$-s]$_i$
Juan(M) touch-CLF-3SG.F.O hand(F) the.M Sergio(M)-F
'Juan touched Sergio's hand.'

(45) Juan täj-je-te$_j$ [mu' Sergio]$_j$ [un=che']$_i$
Juan(M) touch-CLF-3SG.M.O the.M Sergio(M) hand(F)=SUPERESSIVE
'Juan touched Sergio on the hand.'

(46) Juan täj-je-bi**-te**$_j$ [un **mu' Sergio**j-s]$_i$
Juan(M) touch-CLF-POSS.APPL**-3SG.M.O** hand(F) **the.M Sergio(M)**-F
'Juan touched Sergio's hand.'

In (44), there is an object with a regular internal possessor. The possessor agrees with the head of the object 'hand'. The object decides object agreement. In (45), there is an external possessor. This possessor does not agree with 'hand'. However, it triggers object agreement on the verb. The crucial example is (46), in which the verb has an applicative suffix. The possessor agrees with the head of the object 'hand'. However, this possessor also triggers object agreement on the verb. Similar cases with the possessor triggering verb agreement can be found in other languages, see e.g. Stump and Yadav (1988), Meakins and Nordlinger (2017). Ritchie (2016) stresses that prominent internal possessor constructions "are not a homogeneous phenomenon and require different types of analysis for different languages" (Ritchie 2016:623).

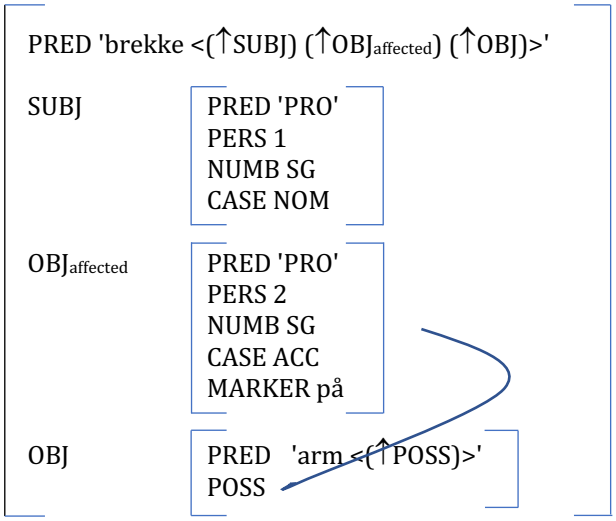
9. Forward and backward raising

The analysis of the external *på* possessor construction raises no new challenges. One can simply transfer one's favorite analysis of the DEP construction. Consider example (47), which is equivalent to (4) above.

(47) Jeg brekker armen på ham.
I break arm.DEF on him
'I break his arm.'

The f-structure of (47) with an external possessor is given in (48). It is almost identical to the f-structure of example (4), given in (5) above, except for a feature from the grammatical preposition *på* 'on'.

(48)



The question is then what kind of syntactic representation would be adequate for the internal *på* possessor construction. The simple answer seems to be that the internal and the external constructions should have the same f-structures. In both cases, there is a possessor that has a double function. In the external construction, one could talk about "forward" possessor raising - the well known type of possessor raising in which the phonologically realized affected object is also a possessor. The proposal here is that there should also be an option for "backward" possessor raising. This is what is needed in the internal construction, in which the phonologically realized possessor is also an affected object.

This proposal gives the basis for an account of the restrictions discussed in section 5. These restrictions are also relevant for other constructions with body part nouns (see notes 5 and 6), and their exact formulation is not at issue here. The point to be made concerns the problems with stating these restrictions, which were mentioned in section 5: First, the *på* possessor imposes the same restrictions independently of its status as external or internal to the body part noun phrase. Second, the internal possessor imposes restrictions on elements that are not local to it in c-structure (concerning e.g. the type of verb). These problems now disappear. When the external and internal possessor positions are structure shared, a restriction on one position is also a restriction on the other.

A comparison of possessor raising to raising and control of subjects could be enlightening. It is clear that possessor raising shares properties with raising and control of subjects of infinitives (Lødrup 2009b, Deal 2013, 2017). For most cases of possessor raising, the parallel to control is more relevant, because the raised possessor realizes a semantic role in both its positions - as a possessor in the body part noun phrase and as an affected participant at the clausal level.

Ritchie (2016, 2017) points out that a situation with a constituent in a low position that is shared with a function at a higher level has a parallel in what has been called backward control of subjects (Polinsky and Potsdam 2002, 2006). This kind of control must be assumed for languages in which the shared subject is phonologically realized in the low position, giving sentences that could be rendered as (49).

(49) tried [John to leave]

The shared argument is at the same time the subject of the main verb and of the subordinate verb. The difference from regular control is that it is phonologically realized in the subordinate position. To draw the parallel to subject control further than Ritchie does, one could say that there are languages with backward possessor raising, just as there are languages with backward control of subjects of infinitives. The important point is that the possessor has

two roles to play, as argument of the verb and as possessor of the body part noun.

The treatment of obligatory control of subjects has been discussed many times. Hornstein (1999) proposed an influential Minimalist treatment, in which the controller is moved from the controlled position to its surface position. This makes it very similar to subject-to-subject-raising; the difference is that a controller moves into a thematic position, while a raised subject moves into a non-thematic position. Polinsky and Potsdam (2002) see the existence of backward control as an argument for Hornstein's analysis; the difference between forward and backward control is only the position in which the moved element is to be pronounced.

In LFG, obligatory control and raising have been treated the same way since the theory was first introduced. The classical article is Bresnan (1982). LFG uses structure sharing, which means that two syntactic functions share one argument. Structure sharing is traditionally implemented as unification, as in the equation (50), which unifies the two subject positions in control and raising sentences.

(50) (↑ SUBJ) = (↑ XCOMP SUBJ)

Unification is a symmetric relation, which says that all properties must be shared between the functions. This means that unification does not distinguish between forward and backward raising.

To account for Norwegian possessor raising, with both forward and backward raising, a corresponding equation is all that is needed. This equation must be a part of the lexical entry for verbs whose valency has been "expanded" to include the affected object. In (51), the equation is a part of the lexical entry for *brekke* 'break' (cf. (48) above).

(51) brekke <(↑ SUBJ) (↑ $\text{OBJ}_{\text{affected}}$) (↑ OBJ)>'
(↑ $\text{OBJ}_{\text{affected}}$) = (↑ GF POSS), where GF is a local function

To avoid overgeneration, the equation should also require the presence of the grammatical preposition *på* 'on'. It could be noted that Norwegian differs from the languages with the DEP construction in that $\text{OBJ}_{\text{affected}}$ and the noun phrase internal possessor have the same form in Norwegian. There is thus no need for the restriction operator (see section 2) in the account of Norwegian possessor raising.

We see, then, that a unification analysis gives a simple account of the situation in Norwegian with both forward and backward possessor raising.[8]

10. Conclusion

I have shown that the Norwegian *på* possessor can be realized within the body part noun phrase. It is then a prominent internal possessor, which must be related to the sentence level using backward possessor raising. The *på* possessor can also be realized at sentence level, so it is necessary to assume that possessor raising can apply both backward and forward. The existence of backward possessor raising strengthens the parallel between possessor raising and obligatory control and raising of subjects, and has consequences for our general understanding of these processes.

[8] A unification analysis raises a more general question: If unification is used to account for raising of subjects and possessors, isn't the implication that all languages should have both forward and backward raising? This is clearly not the case, and there are two possible ways to handle this.

One option is a c-structure account. Concerning subject positions, one could point to the fact that e.g. an infinitival VP in Norwegian does not have a position for a phonologically realized subject. For possessor raising in e.g. French and German, one could make use of the difference in form between an affected object and a possessor in a noun phrase. The shared argument must be a dative nominal, and a dative nominal cannot be realized phonologically in the POSS position of a noun phrase.

Another option is to stay at the level of f-structure, and use subsumption instead of unification to relate the two positions. Subsumption is an assymmetrical relation in which the flow of information goes in one direction only (Zaenen and Kaplan 2002, Sells 2006). If one says that the sentential subject subsumes the infinitival subject, there will be forward control only. In the same way, one could say that the affected object subsumes the possessive in e.g. French and German, to achieve forward control only.

CORPUS

NoWaC (Norwegian Web as Corpus)
http://www.hf.uio.no/iln/om/organisasjon/tekstlab/prosjekter/nowac/index.html

REFERENCES

Bresnan, Joan. 1982. Control and complementation. In Joan Bresnan (ed.) *The Mental Representation of Grammatical Relations*. Cambridge, MA: MIT Press. 282-390.

Butt, Miriam, Tracy Holloway King, and John T. Maxwell III. 2003. Complex predicates via restriction. In Miriam Butt and Tracy Holloway King (eds.) *Proceedings of the LFG03 Conference*, Stanford, CA: CSLI Publications. 92-104.

Dahl, Östen. 2015. *Grammaticalization in the North: Noun Phrase Morphosyntax in Scandinavian Vernaculars*. Berlin: Language Science Press.

Deal, Amy Rose. 2013. Possessor raising. *Linguistic Inquiry* 37, 391-432.

Deal, Amy Rose. 2017. External possession and possessor raising. In Martin Everaert and Henk van Riemsdijk (eds.) *The Wiley Blackwell Companion to Syntax. 2nd edition*. Hoboken NJ: Wiley. 1509-1540.

Eik, Ragnhild. 2014. Setningar med possessivt objekt i norsk: Ein nykonstruksjonistisk analyse av setningar som «ho trakka han på foten». MA thesis. NTNU, Trondheim.

Faarlund, Jan Terje 2004 *The Syntax of Old Norse*. Oxford: Oxford University Press.

Guéron, Jacqueline. 1985. Inalienable possession, PRO-inclusion and lexical chains. In Jacqueline Guéron, Hans-Georg Obenauer and Jean-Yves Pollock (eds.), *Grammatical Representation*, 43-86. Dordrecht: Foris.

Guéron, Jacqueline. 2006. Inalienable possession. Martin Everaert and Henk van Riemsdijk (eds.) *The Blackwell Companion to Syntax. Volume 2*. Malden, MA: Blackwell, 589-638.

Haspelmath, Martin. 1999. External possession in a European areal perspective. Doris L. Payne and Immanuel Barshi (eds.), *External Possession*, 109-135. Amsterdam: John Benjamins.

Hole, Daniel. 2005. Reconciling "possessor" datives and "beneficiary" datives - Towards a unified voice account of dative binding in German. In Claudia Maienborn and Angelika Wöllstein (eds.) *Event Arguments: Foundations and Applications*. Tübingen: Niemeyer, 213-242.

Hornstein, Norbert. 1999. Movement and control. *Linguistic Inquiry* 30, 1, 69-96.

Johannessen, Janne Bondi, Marit Julien and Helge Lødrup. 2014. Preposisjoner og eierskapsrelasjoner: et menneskesentrert hierarki. In Janne Bondi Johannessen and Kristin Hagen (eds.) *Språk i Norge og nabolanda: Ny forskning om talespråk*, 65-98. Oslo: Novus.

Kayne, Richard S. 1975. *French Syntax: The Transformational Cycle*. Cambridge, MA: MIT Press.

König, Ekkehard and Martin Haspelmath. 1998 Les constructions à possesseur externe dans les langues de l´Europe. In Jack Feuillet (ed.) *Actance et valence dans les langues de l'Europe*, 525-606. Berlin: Mouton de Gruyter.

Landau, Idan. 1999. Possessor raising and the structure of VP. *Lingua* 107, 1-37.

Langacker, Ronald W. 1968. Observations on French possessives. *Language* 44, 1, 51-75.

Lee-Schoenfeld, Vera. 2006. German possessor datives: raised and affected. *Journal of Comperative Germanic Linguistics* 9, 2, 101-142.

Lee-Schoenfeld, Vera. 2012. Case and affectedness in German inalienable possession constructions. *Linguistische Berichte* 232 399-416.

Lee-Schoenfeld, Vera. 2016. The syntax of external and internal possessor variation in German inalienable possession. *STUF Language Typology and Universals*, 69, 1 113-129.

Lee-Schoenfeld, Vera and Gabriele Diewald. 2014. The pragmatics and syntax of German inalienable possession constructions. *Annual Meeting of the Berkeley Linguistics Society*. Vol. 40. 286-310.

Levin, Beth 1993 *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.

Lødrup, Helge. 2009a. External and internal possessors with body part nouns: The case of Norwegian. *SKY Journal of Linguistics* 22, 221-250.

Lødrup, Helge. 2009b. Looking Possessor Raising in the mouth: Norwegian Possessor Raising with unergatives. In Miriam Butt and Tracy Holloway King (eds.) *Proceedings of the LFG09 Conference*, Stanford, CA: CSLI Publications. 420-40.

Lødrup, Helge. 2010. Implicit possessives and reflexive binding in Norwegian. *Transactions of the Philological Society* 108, 2, 89-109.

Meakins, Felicity, and Rachel Nordlinger. 2017. Possessor dissension: Agreement mismatch in Ngumpin-Yapa possessive constructions. *Linguistic Typology* 21, 1, 143-176.

Nikolaeva, Irina. 2002. The Hungarian external possessor in a European perspective. In Cornelius Hasselblatt and Roger Blokland (eds.) *Finno-Ugrians and Indo-Europeans: Linguistic and Literary Contacts. Proceedings of the Symposium at the University of Groningen, November 22-24, 2001*. Maastricht: Shaker. 272-285.

Polinsky, Maria and Eric Potsdam. 2002. Backward control. *Linguistic Inquiry* 33, 2, 245-282.

Polinsky, Maria, and Eric Potsdam. 2006. Expanding the scope of control and raising. *Syntax* 9, 2, 171-192.

Ritchie, Sandy. 2016. Two cases of prominent internal possessor constructions. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway-King, Stefan Müller (eds.): *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar*. 620-40. Stanford, CA: CSLI Publications.

Ritchie, Sandy. 2017. Agreement with the internal possessor in Chimane. *Studies in Language* 41, 3, 660-716.

Sells, Peter. 2006. Using subsumption rather than equality in functional control. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG06 Conference*. Stanford CA: CSLI Publications.

Skard, Vemund. 1951. *Dativstudien: Dativus sympatheticus und Dativus comparationis in der norrönen Sprache*. Skrifter (Det Norske videnskaps-akademi i Oslo) Historisk-filosofisk klasse. Oslo: Dybwad.

Stolz, Thomas; Kettler, Sonja; Stroh, Cornelia and Urdze, Aina. 2008. *Split Possession: An Areal-linguistic Study of the Alienability Correlation and Related Phenomena in the Languages of Europe*. Amsterdam: Benjamins.

Stump, Gregory T. and Ramawatar Yadav. 1988. Maithili verb agreement and the Control Agreement Principle. In *Chicago Linguistics Society* (CLS) 24, 2, 304-321.

Thráinsson, Höskuldur. 2007. *The Syntax of Icelandic*. Cambridge: Cambridge University Press.

Vergnaud, Jean-Roger and Maria Luisa Zubizarreta. 1992. The definite determiner and the inalienable construction in French and English. *Linguistic Inquiry* 23, 595-652.

Zaenen, Annie and Ronald M. Kaplan. 2002. Subsumption and equality: German partial fronting in LFG. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG02 Conference*. Stanford CA: CSLI Publications, 408-26.

The Glue Semantics Workbench: A modular toolkit for exploring Linear Logic and Glue Semantics

Moritz Meßmer
Across Systems GmbH

Mark-Matthias Zymla
University of Konstanz

Proceedings of the LFG'18 Conference

University of Vienna

Miriam Butt, Tracy Holloway King (Editors)

2018

CSLI Publications

pages 268–282

http://csli-publications.stanford.edu/LFG/2018

Keywords: Glue semantics, linear logic, implementation, Java, workbench

Meßmer, Moritz, & Zymla, Mark-Matthias. 2018. The Glue Semantics Workbench: A modular toolkit for exploring Linear Logic and Glue Semantics. In Butt, Miriam, & King, Tracy Holloway (Eds.), *Proceedings of the LFG'18 Conference, University of Vienna*, 268–282. Stanford, CA: CSLI Publications.

Abstract

In this paper we present an easy to use, modular Glue semantic prover building on the work by Crouch & van Genabith (2000) and implemented in Java. We take inspiration from a Glue semantics parser written in Prolog as well as other existing tools such as the NLTK Glue semantics system. The architecture of our semantic parser allows us to explore the computational viability of linear logic as a mechanism for modeling compositional semantics within LFG. Furthermore, it allows researchers interested in linear logic (for computational linguistics) to research its usefulness, when applied to different syntactic models and various formal semantic frameworks. The goal of this resource is to provide an accessible entry point for both beginners and adepts in computational semantics. It thus also has prospective uses as a teaching tool for computational semantics and linear logic.

1 Introduction

In this paper we present an easy to use, modular Glue semantic prover and parser called *Glue semantics workbench* building on the work by Crouch & van Genabith (2000).[1] Thereby, we revive a Glue semantics parser written in Prolog, since this first implementation is not readily accessible anymore, due to the commercialization of the programming language.[2] Our goal is to translate the system into a more modern implementation within the Java programming language.

Glue semantics is the formalism of choice for formal semantics within the LFG framework (Dalrymple, 2001), but has since attracted interest from different venues, e.g. Asudeh & Crouch (2002); Gotham (2018); Garrette & Klein (2009); Gotham & Haug (to appear). The composition process in this framework is based on linear logic which guides semantic composition comparable to types in Montague semantics (Montague, 1970). Linear logic lends itself well to modelling compositionality due to its resource-sensitivity (Dalrymple, 2001).

The Glue prover presented in this paper is a rejuvenation of existing theoretical and practical approaches to modeling Glue semantics. More specifically, the sys-

[†]We thank the participants of the 2018 LFG conference for valuable feedback. Many thanks also for the very helpful comments by the internal and external reviewers. We are particularly grateful to Richard Crouch and Valeria de Paiva for their support in developing the Glue semantics workbench. Furthermore, we thank the researchers at the CSLI, Stanford for their assistance.

[1]The Glue semantics workbench is publicly available on https://github.com/mmessmer/GlueSemWorkbench. It is free software and distributed under the conditions of the GNU General Public License.

[2]The Prolog Glue prover has been designed as part of the Xerox Linguistic Environment (XLE). In its older iterations, this system relied on SICStus Prolog, a commercial strain of the Prolog family of programming languages. More recent iterations of XLE do not rely on SICStus Prolog anymore, however, this is at the cost of certain features of XLE, in particular the transfer system (Crouch et al., 2017) and other systems that rely on a Prolog interface, such as AKR semantics (Bobrow et al., 2007) and the mentioned Prolog Glue prover.

tem is based on a chart parser devised by Hepple (1996). The main element taken from Hepple's work is the compilation process, which is used to deconstruct Glue premises from higher-order premises into first-order premises. This simplifies the combinatory process of Glue semantic resources. The Glue prover has been further refined with ideas from Gupta & Lamping (1998) that improve efficiency by reducing unnecessary steps in the computation.

Employing these two strategies allows us to present a reasonably efficient algorithm for conducting Glue semantics computations. The modularity of our semantic parser not only allows us to continue the exploration of the computational viability of linear logic as a mechanism for modeling compositional semantics within LFG but it also allows us to explore the interoperability of linear logic with respect to other syntactic theories as well as different semantic formalisms. For this purpose, we illustrate the use of the Glue prover in interaction with both LFG and UD (universal dependency) grammars on the syntactic side and its interaction with Montague-style lambda calculus on the semantic side. To this end, we have implemented a light-weight Montague-style semantics that can be combined with the linear logic prover; however, other semantic formalisms can also be plugged-in without the need to change the overall system.

The paper is structured as follows: In Section 2 we briefly introduce the formalities of linear logic and how it can be used in the domain of compositional semantics. Readers who are familiar with the subject and are interested in the concrete implementation may jump directly to Section 3. The architecture of the system is described in Section 4, which involves the assembly of the linear logic prover with syntactic parsers and formal semantic models. This is of particular interest for readers who intend to work with the Glue semantics workbench. Section 5 concludes.

2 Glue Semantics

Glue semantics is a framework that continues to attract interest not only in the LFG community. Its elegance in terms of aligning the structure of logic proofs with the structure of semantic meaning compositions through the Curry-Howard-Isomorphism has motivated researchers to adapt it for other frameworks such as HPSG (Asudeh & Crouch, 2002), LTAG (Frank & van Genabith, 2001) and Minimalism (Gotham, 2015, 2018). Instead of relying on rules that map the syntactic structure to semantic composition rules, Glue semantics uses a fragment of linear logic to constrain the composition of meaning representations.

In this scenario, a semantic representation is a pair consisting of a linear logic side and a meaning side (in this paper: Montague-style lambda calculus). Thereby, the logic side constrains the possible combination of semantic elements. I.e. the linear logic side of a lexical entry constrains the compositional possibilities of its

meaning side. This is reflected in the Curry-Howard-Isomorphism. The isomorphism describes the correspondence between natural deduction proofs, i.e. the logic side, and computational models like lambda calculus, i.e. the meaning side. It is the foundation for the pairing of logics used in the Glue approach. More concretely, the Curry-Howard-Isomorphism states that lambda abstraction on the meaning side corresponds to $\multimap$ introduction and functional application corresponds to $\multimap$ elimination. This is illustrated in the following figure. On the left side, it is shown how the introduction of a linear implication affects the meaning side: A lambda function is generated. On the right side, the correspondence between a functional application and the combination of a linear implication with its corresponding resource is depicted. Due to this system, Glue semantics formulas can be composed and decomposed on the logic side and the meaning side in concord.

$$\frac{\begin{array}{c}[x:A]^i \\ \vdots \\ f(x):B\end{array}}{\lambda x.f(x):A \multimap B}\ \multimap_{I,i} \qquad \frac{f:A \multimap B \qquad a:A}{f(a):B}\ \multimap_E$$

Figure 1: Implication introduction and elimination

In Glue semantics proofs, the $\multimap_E$ rule is applied when *combining* two meaning constructors, while the $\multimap_I$ rule is used for introducing assumptions. In linear logic proofs, assumptions are a deduction tool for deriving a proof whose premises are not immediately compatible. For a proof to be valid, all assumptions that have been made during the deduction have to be reintroduced via implication introduction. This follows from the general principle of linear logic, which states that a valid proof needs to consume all available resources. In other words, assumptions are simply treated as additional resources that emerge during the computation.

(1) **John** $j:g$
Mary $m:h$
loves $\lambda x.\lambda y.loves(x,y):g \multimap (h \multimap f)$

$$\frac{\dfrac{\lambda x.\lambda y.loves(x,y):g \multimap (h \multimap f) \qquad j:g}{\lambda y.loves(j,y):h \multimap f} \qquad m:h}{loves(j,m):f}$$

Figure 2: Derivation of *John loves Mary.*

The propositional implicational fragment of linear logic paired with lambda-calculus is already capable of deriving simple meaning structures as shown in Figure 2. However, as soon as scope-taking expressions enter the stage and potentially

$$f\begin{bmatrix} \text{PRED} & \text{`kiss}\langle \text{SUBJ,OBJ} \rangle\text{'} \\ \text{SUBJ} & m\begin{bmatrix} \text{SPEC} & \begin{bmatrix} \text{PRED} & \text{`every'} \end{bmatrix} \\ \text{PRED} & \text{`man'} \end{bmatrix} \\ \text{OBJ} & w\begin{bmatrix} \text{SPEC} & \begin{bmatrix} \text{PRED} & \text{`a'} \end{bmatrix} \\ \text{PRED} & \text{`woman'} \end{bmatrix} \end{bmatrix}$$

Figure 3: F-structure of *Every man loves a woman.*

introduce ambiguities, this fragment of linear logic does not suffice anymore. A quantifier expression can take different constituents as its scope, therefore the labels on the linear logic side cannot be fixed to constants. Instead, the scope of a quantifier is encoded with linear logic variables. These variables are introduced by a universal quantifier which binds a variable that is instantiated to a linear logic constant in the derivation process. Consequently, we move from a propositional linear logic fragment to a higher-order predicate logic fragment of linear logic involving universal quantification over f-structure labels.[3] Consider the following sentence and its f-structure.

(2) Every man kisses a woman.

(3) $$m\begin{bmatrix} \text{SPEC} & \begin{bmatrix} \text{PRED} & \text{`every'} \end{bmatrix} \\ \text{PRED} & \text{`man'} \end{bmatrix} \qquad m_\sigma\begin{bmatrix} \text{VAR} & v\begin{bmatrix} \ \end{bmatrix} \\ \text{RESTR} & r\begin{bmatrix} \ \end{bmatrix} \end{bmatrix}$$

(4) **every**: [((SPEC $\uparrow$)$_\sigma$ VAR) $\multimap$ ((SPEC $\uparrow$)$_\sigma$ RESTR)]
$\multimap \forall$X.[((SPEC $\uparrow$)$_\sigma$ $\multimap$ X) $\multimap$ X]
man: ($\uparrow_\sigma$ VAR) $\multimap$ ($\uparrow_\sigma$ RESTR)

(4) shows the Glue side of the meaning constructor of the quantifying expression *every*. The scope part of the quantifier universally quantifies over variables of type t. This can be seen as a quantifier which expresses that any f-structure can be inserted as the scope of *every*. Deriving the two scope constellations now requires us to restructure the proof with the tools of natural deduction: eliminating implications by combining resources and introducing implications by making (temporary) assumptions.

Given the following meaning constructors with instantiated labels, we can make two logically equivalent derivations, which result in two different readings. For reasons of brevity, the quantified NPs are already combined with their restrictors.

[3]This is a fairly standard system in the Glue literature. However, it has been argued that a first-order linear logic fragment is sufficient to model natural language semantics. For discussion on this topic, see Kokkonidis (2008). Thanks to the external reviewer for bringing this to our attention.

every man	λP.$\forall$x[man(x) $\rightarrow$ P(x)]	:	$(m_\sigma \multimap f_\sigma) \multimap f_\sigma$
a woman	λQ.$\exists$y[woman(y) $\wedge$ Q(y)]	:	$(w_\sigma \multimap f_\sigma) \multimap f_\sigma$
kiss	λx.λy.kiss(x,y)	:	$m_\sigma \multimap (w_\sigma \multimap f_\sigma)$

In their original form given by the lexical entries, the quantifiers cannot combine with the verb. One (surface scope) or both (inverse scope) lambda slots of the verb have to be temporally saturated with assumptions. Assumptions need to be reintroduced into the computation later in the proof in accordance with Figure 1. As shown in Figure 4a and 4b, the assumed resources are marked with square brackets and indices. They introduce a temporary unbound variable that is bound later via lambda abstraction when the linear implication is reintroduced via $\multimap$ introduction.

The Glue semantics fragment introduced above is the foundation for the Glue semantics workbench. In the next section we show how this system is translated into a computationally viable Glue semantics parser.

3 The linear logic prover algorithm

While this system of using assumptions in the deduction process to temporarily saturate lambda binders on the meaning side is very elegant for dealing with scope ambiguities since it is independent of syntactic assumptions about the phenomenon, it is a very complex system for any automatic proving algorithm. In practice, even the implicational fragment of linear logic used here is NP-complete and may be computationally intractable once the formulas reach a certain complexity. An algorithm to circumvent some of the computational complexity was proposed by Hepple (1996) and is used for our proving algorithm as well. Additionally, our proving algorithm uses a system, based on Gupta & Lamping (1998), that distinguishes between symmetric modifier resources and asymmetric skeleton resources to further increase its efficiency. The algorithm is based on three principles to transform linear logic proofs into computationally tractable algorithms: (I) indexation of premises (II) compilation of nested implications (III) separation of modifier-type premises. In the following, the three principles will be explained and the algorithm for calculating proofs is outlined.

3.1 Basic first-order chart prover

As has been pointed out by Crouch & van Genabith (2000), Glue semantics strongly resembles categorial grammar systems and a Glue semantics proof resembles the principles of syntactic chart parsing techniques. In both systems individual items (premises in linear logic and words or constituents for chart parsers) are taken from the agenda and combined to obtain intermediate results which in turn are combined again until the agenda is empty. The proving system by Hepple (1996) makes use of this resemblance by adapting the chart parsing technique for linear logic proofs in the form of **indexation**. The Glue premises are each assigned a set of indices in Hepple's prover algorithm. Initial premises are added to an

$$\dfrac{\dfrac{\dfrac{[X : m_\sigma]^1 \qquad \lambda x.\lambda y.\text{kiss}(x,y) : m_\sigma \multimap (w_\sigma \multimap f_\sigma)}{\lambda y.kiss(X,y) : w_\sigma \multimap f_\sigma \quad \lambda Q.\exists y[\text{woman}(y) \wedge Q(y)] : (w_\sigma \multimap f_\sigma) \multimap f_\sigma}\multimap_E}{\exists y[\text{woman}(y) \wedge \text{kiss}(X,y)] : f_\sigma}\multimap_E}{\lambda P.\forall x[\text{man}(x) \rightarrow P(x)] : (m_\sigma \multimap f_\sigma) \multimap f_\sigma \quad \lambda x.\exists y[\text{woman}(y) \wedge \text{kiss}(x,y)] : m_\sigma \multimap f_\sigma}\multimap_{I,1}}{\forall x[\text{man}(x) \rightarrow \exists y[\text{woman}(y) \wedge \text{kiss}(x,y)]] : f_\sigma}\multimap_E$$

(a) Glue proof: *Every man kisses a woman* surface scope

$$\dfrac{\dfrac{\dfrac{\dfrac{[X : m_\sigma]^1 \qquad \lambda x.\lambda y.\text{kiss}(x,y) : m_\sigma \multimap (w_\sigma \multimap f_\sigma)}{\lambda y.\text{kiss}(X,y) : w_\sigma \multimap f_\sigma \qquad [Y : w_\sigma]^2}\multimap_E}{\text{kiss}(X,Y) : f_\sigma}\multimap_E}{\lambda x.\text{kiss}(x,Y) : m_\sigma \multimap f_\sigma}\multimap_{I,1}}{\lambda P.\forall x[\text{man}(x) \rightarrow P(x)] : (m_\sigma \multimap f_\sigma) \multimap f_\sigma \qquad \forall x[\text{man}(x) \rightarrow \text{kiss}(x,Y)] : f_\sigma}\multimap_E}{\lambda y.\forall x[\text{man}(x) \rightarrow \text{kiss}(x,y)] : w_\sigma \multimap f_\sigma \quad \lambda Q.\exists y[\text{woman}(y) \wedge Q(y)] : (w_\sigma \multimap f_\sigma) \multimap f_\sigma}\multimap_{I,2}}{\exists y[\text{woman}(y) \wedge \forall x[\text{man}(x) \rightarrow \text{kiss}(x,y)]] : f_\sigma}\multimap_E$$

(b) Glue proof: *Every man kisses a woman* inverse scope

Figure 4: Quantifier ambiguity in Glue

agenda and each is assigned a single index. For each combination of two premises, the index sets of the two premises are also combined and the joined index set is assigned to the newly created premise. If a premise $A : [0]$, for example, is combined with a premise $A \multimap B : [1]$ then the newly created premise is $B : [0, 1]$. In order to ensure that each possible combination is checked, the algorithm works with an agenda, containing all "fresh" premises and a database containing all the premises which have already been taken from the agenda. For each premise taken off the agenda, the algorithm checks combinatory possibilities with all premises in the database. If the current premise combines with one from the database, the newly created premise is also added to the agenda. After all checks have been made, the current premise is moved from the agenda to the database and the algorithm proceeds with the next premise from the agenda. Unnecessary or invalid steps in the computation can thus be avoided by requiring that when combining two premises, their index sets must be disjoint.

3.2 Compilation of higher-order premises

As mentioned above, this simple chart prover algorithm reaches its limits as soon as the proof contains higher-order premises. For our algorithm, higher-order linear logic formulas are nested implications where the antecedent is itself an implication.[4] In a natural deduction-style proof, these formulas require making assumptions and discharging them at some points of the proof. It would take an algorithm a great deal of computational effort to determine when it is necessary to make an assumption and when to discharge it. Therefore Hepple's prover implements a computationally feasible solution to that problem: every higher-order premise is compiled by separating its antecedent as an additional premise and adding it to the agenda, marked as an assumption.[5] The premise from which the auxiliary premise is taken is marked with a dependency on the respective auxiliary premise. This step is repeated until only first-order premises are left. In the notation of our proving algorithm, auxiliary premises are marked with $\{\}$ and their dependencies (*discharges*) are marked on the original premise with []. More concretely, these references are implemented in our code such that each premise has two lists associated with it, one for assumptions and one for discharges.[6]

(5) $(a \multimap b) \multimap c\ [0] \Rightarrow_{compile} b[a] \multimap c\ [0];$

[4]Note that this does not correspond to higher-order linear predicate logic formulas as discussed in the previous section.

[5]In Hepple (1996) the term *assumption* is used to describe these auxiliary premises. In this paper, both of these terms refer specifically to premises that have been generated via the compilation process; i.e. premises that have been cut off from a higher-order premise.

[6]In the original algorithm by Hepple, references to auxiliar premises are made via their indices. The premise from which an assumption is compiled out will from now on be called the assumption's *host* premise. For our system we decided to add references to the Glue resources themselves, as that makes the proofs more readable and is easily implemented due to Java's object-oriented programming paradigm.

$$\{a\}\ [1]$$

By adding a reference to the extracted assumption to its host premise, the algorithm prevents invalid proofs where the assumption might be used, without later discharging it. This restriction is achieved by adding two rules to the proving algorithms. First, a premise P containing a set of discharges δ may only combine with a premise whose list of assumptions α is a subset of δ. In that case, all matching assumption and discharge pairs are removed from the newly created premise. Second, if two premises contain (or are themselves) assumptions, their lists of assumptions are joined. With these modifications, a proof is now only valid if the resulting premise, besides containing all initial indices, does not have any assumptions or discharges associated with it.

$$(6) \qquad \frac{b\{a\}[1,2] \qquad b[a] \multimap C[3]}{c[1,2,3]}$$

So far, only the linear logic side has been dealt with, but of course the semantic side of a premise is affected by the compilation process as well. As mentioned before, operations on linear logic proofs and operations on lambda-expressions on the semantic side of premises are aligned via the Curry-Howard isomorphism. Implication elimination on the Glue side of a premise can therefore be seen as a functional application operation on the semantic side, while implication introduction amounts to functional abstraction. This becomes relevant when proofs contain assumptions. Auxiliary premises that are introduced into a Glue semantics proof carry unbound variables on the semantic side. When an assumption is combined with another premise, this variable is then inserted into the semantic representation of that other premise via functional application. Later in the proof, when the assumption is discharged, the assumption variable is bound by a lambda term again.

This elegant system of temporarily saturating λ-slots in semantic computation is one of the reasons why Glue semantics interests formal semanticists. It allows a system of formally resolving ambiguities without having to rely on additional abstract systems such as a logical form or Cooper-storage (Cooper, 1983). The semantic aspect of Glue proofs is covered by Hepple (1996) as well. In his algorithm, auxiliary premises created in the compilation process carry temporary variables as well, but the re-binding of the variables is done via an additional lambda binder that is functionally applied to the semantic representation of the host premise. As soon as the premise containing the unbound assumption variable combines with its host the lambda term binds the variable. The lambda term binding the variable can then be applied to the original meaning representation. In regular lambda calculus such an "accidental" binding an unbound variable by adding a lambda binder is not a legal operation. However, Hepple's algorithm uses this operation in a deliberate and controlled manner. By using different variables for each newly created formula during the compilation process, the algorithm therefore ensures that free variables are accidentally bound by the wrong lambda binder. In our prover algo-

rithm the creation of new variables is handled centrally for all formulas in a proof. This allows full control of which variables are used and inserted into formulas. The compilation of glue formulas with semantic representations is illustrated in Figure 5.

$$\dfrac{\dfrac{H[g_2] \multimap H : \lambda u.\lambda P.\forall x[\text{person}(x) \wedge P(x)](\lambda v.u) \qquad \dfrac{g_1 \multimap f : \lambda y.\text{sleep}(y) \qquad \{g_2\} : v}{f\{g_2\} : \text{sleep}(v)}}{f : \lambda P.\forall x[\text{person}(x) \wedge P(x)](\lambda v.\text{sleep}(v))}\,[\text{H/f}]}{f : \forall x[\text{person}(x) \wedge \text{sleep}(x)]}\,\beta\text{-conversion} \tag{7}$$

Figure 5: Every person sleeps. – Hepple style

3.3 Treating modifier premises

While this algorithm is already capable of handling linear logic formula of the implicational fragment, it is still rather inefficient if a proof contains *modifier*-type premises. Modifiers as defined by Gupta & Lamping (1998) are premises whose linear logic side has a certain pattern. This pattern can be seen if occurrences of linear logic atoms, or in the case of Glue semantics, type labels, are assigned a polarity depending on whether they occur in the antecedent or the consequent of a linear implication: the consequent has the same polarity as the whole implication and the antecedent has the opposite polarity. Assuming that linear logic formulas as a whole always have positive polarity, the polarity of each atom can thus be assigned:

$$((f_+ \multimap g_-)_- \multimap (f_- \multimap g_+)_+)_+ \tag{8}$$

$$((v_+ \multimap r_-)_- \multimap ((g_+ \multimap X_-)_- \multimap X_+)_+)_+ \tag{9}$$

The formula in (8) is considered to be a modifier type because all positive occurrences of Glue labels are matched up with negative occurrences. In Glue semantics, lexical entries for adjuncts are usually modifier types because they modify the meaning of an f-structure node without altering its type. Other lexical entries are mostly purely skeleton-type because they only consist of singular positive or negative occurrences of each label.

There are some cases where skeleton and modifier-type occurrences are mixed inside a formula. Quantifiers, like the one in (9), for example, are mostly skeleton-types except for the matching positive and negative occurrence of the Glue variable (X in the example above) which denotes the scope. These mixed-type quantifiers are also compiled. In general, all premises that are not pure modifiers are compiled until they are either pure skeletons or have the form $a \multimap M$, where a is an atomic type and M is a pure modifier type. The latter sort of premise is treated like a skeleton during the deduction process (until the atomic antecedent is consumed and the premise becomes a pure modifier). Such cases occur for certain kinds of modifiers. One such example would be recursive modification as described in Dalrymple (2001). In order to obtain the correct meaning for the phrase *apparently*

Swedish man, Dalrymple (2001) proposes an internally structured meaning of adjectival and adverbial modifiers. This in principle means that the meaning of these modifies is deconstructed into two separate modifiers. One constructor contributes the lexical information of the respective modifier and the second constructor contributes the structure for semantic composition, i.e. it guarantees that *apparently* modifies the complete noun phrase *swedish man* and not just *man*. This leads to the following lexical entries:

Swedish1	$\lambda x, Swedish(x)$	$: (g_v \multimap g_\sigma)$
Swedish2	$\lambda Q.\lambda P.\lambda x.Q(x) \wedge P(x)$	$: (g_v \multimap g_\sigma) \multimap ((v \multimap r) \multimap (v \multimap r))$
apparently1	$\lambda P, apparently(P)$	$: (h_v \multimap h_\sigma)$
apparently2	$\lambda Q.\lambda R.\lambda x.Q(R(x))$	$: (h_v \multimap h_\sigma) \multimap ((g_v \multimap g_\sigma) \multimap (g_v \multimap g_\sigma))$
man	$\lambda y.man(y)$	$: (v \multimap r)$
ap. sw. man	$\lambda x.apparently(swedish(x) \wedge man(x))$	$: (v \multimap r)$

In the above case the meaning constructors of *Swedish2* and *apparently2* would each be compiled once, so they have the appropriate form. The adjective *Swedish1* may then combine either directly with *Swedish2* so it can be applied to the noun; or it may first combine with the two lexical entries for *apparently* to yield the fully modified phrase. The former combination would be possible and one of the partial solutions, but not a valid one, as it does not contain all initial premises. Only the second option would be recognized as a valid solution by the prover.

4 Structure of the workbench

This section presents the overall structure of the system. Thereby, we want to give the reader a brief overview of the packages and how they are organized to allow for extension with one's own work.

When implementing this program, the intention was not only to provide an easily-accessible Glue prover, but also to create a tool that is interesting for formal linguists, especially formal semanticists and those working at the syntax-semantics interface, and also for users who are interested in NLP applications. In order to make the workbench extendable for any of these purposes, the code was modularized. Figure 6 shows the structure of the program, with arrows indicating the flow of data. Blue boxes represent packages and green boxes represent Java classes. The prover itself, as the core component of the workbench, has its own module. It takes a list of lexical entries as input and then searches for all valid solutions using the algorithm outlined above. The deduction process, as well as all valid solutions are printed and displayed to the user. Input for the *prover* module can be generated in two ways: by directly entering all lexical entries or by using an interface to XLE or to the Stanford CoreNLP dependency parser (Manning et al., 2014).[7]

[7] As of the publication of this paper, there exists no interface with non-Java libraries in the workbench. This means that the dependency structure is generated at run-time upon entering a sentence since it can use the Stanford CoreNLP Java library. On the other hand, XLE parses still need to be generated externally. We hope to integrate this functionality for future iterations of the workbench.

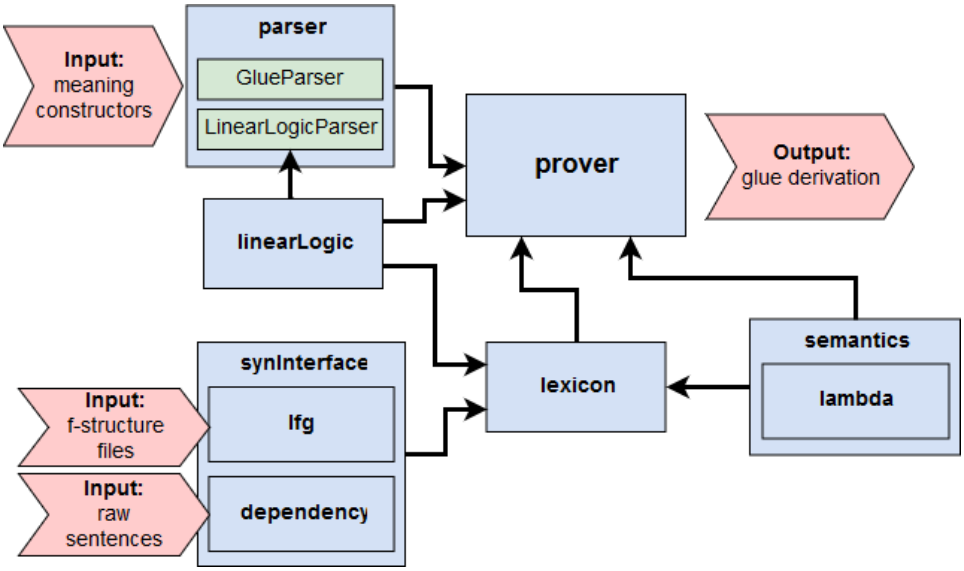


Figure 6: Module diagram

The Workbench also has a built-in parsing system by using the programming interface of the Stanford CoreNLP tools to create dependency structures. This allows users to parse single sentences and create the appropriate lexical entries from a small toy lexicon. This is discussed in more detail in section 4.1.

Either way, the user input is converted to a list of premises as input for the prover. The components for parsing lexical entries directly are situated inside the *parser* module, while the LFG and dependency structure input is handled in separate packages inside the *synInterface* module. All Java resources that are related to linear logic and the underlying proof system are part of the *linearLogic* module. Classes used for representing semantic formalisms can be found inside the *semantics* module. It contains an implementation of Montague-style lambda calculus that is used as a default semantic framework for the prover. In the *lexicon* module all classes for creating lexical entries from syntactically parsed input can be found. The distribution of the workbench contains a toy lexicon, but it can be extended, or even completely replaced, as desired by the user. In the remainder of this section, the two methods of providing input for the prover will be outlined.

4.1 Generating lexical entries

The systems for creating lexical entries are very similar for LFG and dependency structures. therefore only the generation of lexical entries based on LFG structures will be described here. The f-structure parser in the *synInterface* module reads f-structure files in Prolog syntax and generates lexical entries from the syntactic information extracted from the input file. The XLE interface is thus compatible with stand-alone XLE distributions and also with Prolog output generated by the

INESS XLE-Web service (Rosén et al., 2012). It also has a small toy lexicon integrated that can be extended and modified. In the original release of the Workbench, it contains classes for verbs (intransitive, transitive and ditransitive), common and proper nouns, determiners (including quantifiers) and adjectival modifiers.

Both syntactic frameworks access the *lexicon* module for generating lexical entries. In this module, the information given by the syntactic analysis is used to generate a semantic representation for the semantic side and a linear logic formula for the Glue side of the lexical entry. These lexical entries are then converted into premises that can be used by the prover. The generation of lexical entry objects from the input data follows the same principle in both systems. First, the root predicate and its arguments are determined. The arguments are resolved first so the appropriate template for the verb can be chosen, based on the subcategorization frame. The head of an argument f-structure is resolved first and afterwards its subordinate items, such as modifiers and determiners. Such dependents of an f-structure head include structural information about their head in their Glue meaning constructors.

(10)
$$\begin{bmatrix} \text{PRED 'man'} \\ \text{ADJ} \left\{ \begin{bmatrix} \text{PRED 'swedish'} \end{bmatrix} \right\} \end{bmatrix}$$

(11) **man** $\lambda x.man(x) : v \multimap r$
Swedish $\lambda P.\lambda y.Swedish(y) \wedge P(y) : (v \multimap r) \multimap (v \multimap r)$

Consider the lexical entries for the NP 'Swedish man' in (11) (Dalrymple, 2001). The adjectival modifier 'Swedish' is part of the f-structure of the NP and therefore uses the same Glue labels as its head. The noun itself has the semantic type $< e, t >$ and as the adjective modifies this meaning, its type is $<< e, t >, < e, t >>$. These semantic types are reflected in the Glue labels. The lexical resource we provide uses a top-down algorithm for generating meaning constructors. In the example above, ((10)), this would be the constructor for *man*. The modifier *swedish* which is subordinated from an f-structure perspective can thus access the relevant Glue elements which have been generated for the governing structure. Thus, in ((11)) the entries share the Glue constants v and r. In other words, the head of each dependent is always resolved first and therefore the necessary information of a given (partial) f-structure can be passed down to its modifiers. In the toy lexicon that is provided with the Glue Workbench modifiers only take the Glue labels of their heads as arguments, but as all lexical and functional information of the parent f-structures is available during the instantiation of the lexical entry for the modifier, other restrictions, such as semantic types could be passed as well.

4.2 Parsing Glue premises

As Glue semantics is a framework with growing interest from different semantic and syntactic backgrounds, the Workbench tries to honor that diversity by provid-

ing the possibility of directly entering and parsing Glue meaning constructors that can then be fed into the Glue prover. The "native" system of the Workbench is a Montague-style lambda calculus, but it is possible to use other semantic frameworks, such as DRT. However, by default, the compilation algorithm employed by the prover uses lambda abstraction and lambda application operations to modify the meanings accordingly. The compilation algorithm was implemented in such a way that it is possible to add a different semantic formalism (via Java interfaces).

When entering lexical entries manually, the prover will use the "default" classes for generating the meaning side. This means that all semantic representations will be treated as an atomic string of characters that is not modified. During the derivation process lambda abstractions may be added and modified, but the core meaning provided in the original lexical entry will remain untouched. That way, the Workbench allows using any kind of semantic framework as input and the semantic representations that are derived by the prover can be evaluated manually or with a beta reduction tool.

5 Conclusion

In this paper the Glue Semantics Workbench was outlined as a tool for research at the syntax/semantics interface. The Workbench is centered around a Glue proving algorithm which is able to process Glue semantic expressions that are part of the implicational subset of linear logic, commonly used in the newer style of Glue semantics. Our implementation resolves issues with the computational tractability and implements some improvements in efficiency, using the algorithms outlined by Hepple (1996) and Gupta & Lamping (1998). Due to its modularized implementation via Java packages, the Workbench allows some flexibility for the user. It offers three modes for providing lexical entries as input for the parser: entering and parsing them directly or letting the *lexicon* and *synInterface* modules derive them either from LFG f-structures or dependency parses. The modular structure allows relatively easy modification and extension of its modules.

References

Asudeh, Ash & Richard Crouch. 2002. Glue Semantics for HPSG. In *Proceedings of the 8th International HPSG Conference*, 1–19.

Bobrow, Daniel G., Bob Cheslow, Cleo Condoravdi, Lauri Karttunen, Tracy Holloway King, Rowan Nairn, Valeria de Paiva, Charlotte Price & Annie Zaenen. 2007. PARC's Bridge and Question Answering System. In *Proceedings of the GEAF 2007 Workshop*, 1–22.

Cooper, Robin. 1983. *Quantification and Syntactic Theory*, vol. 21 Synthese Language Library, Texts and Studies in Linguistics and Philosophy. Dordrecht: Springer.

Crouch, Dick, Mary Dalrymple, Ronald M. Kaplan, Tracy Holloway King, John T. Maxwell III & Paula Newman. 2017. *XLE Documentation*. Palo Alto Research Center.

Crouch, Richard & Josef van Genabith. 2000. Linear Logic for Linguists. *http://www2.parc.com/istl/members/crouch* .

Dalrymple, Mary. 2001. *Lexical Functional Grammar*, vol. 34. Academic Press.

Frank, Anette & Josef van Genabith. 2001. GlueTag-Linear Logic Based Semantics for LTAG – and What It Teaches Us About LFG and LTAG. In *Proceedings of the LFG01 Conference*, CSLI Publication.

Garrette, Dan & Ewan Klein. 2009. An Extensible Toolkit for Computational Semantics. In *Proceedings of the eighth international conference on computational semantics*, 116–127. Association for Computational Linguistics.

Gotham, Matthew. 2015. Towards Glue Semantics for Minimalist Syntax. *Cambridge Occasional Papers in Linguistics* 8. 5683.

Gotham, Matthew. 2018. Making Logical Form Type-Logical: Glue Semantics for Minimalist Syntax. *Linguistics and Philosophy* 1–46.

Gotham, Matthew & Dag Haug. to appear. Glue semantics for Universal Dependencies. In *Proceedings of the 23rd lexical functional grammar conference*, tbd.

Gupta, Vineet & John Lamping. 1998. Efficient Linear Logic Meaning Assembly. In *Proceedings of the 17th international conference on computational linguistics-volume 1*, 464–470. Association for Computational Linguistics.

Hepple, Mark. 1996. A Compilation-Chart Method for Linear Categorial Deduction. In *Proceedings of the 16th conference on computational linguistics-volume 1*, 537–542. Association for Computational Linguistics.

Kokkonidis, Miltiadis. 2008. First-Order Glue. *Journal of Logic, Language and Information* 17(1). 43–68.

Manning, Christopher, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard & David McClosky. 2014. The Stanford CoreNLP Natural Language Processing Toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 55–60.

Montague, Richard. 1970. English as a formal language. In Bruno Visentini (ed.), *Linguaggi nella societa e nella tecnica*, 188–221. Edizioni di Communita.

Rosén, Victoria, Koenraad De Smedt, Paul Meurer & Helge Dyvik. 2012. An Open Infrastructure for Advanced Treebanking. In *Meta-research workshop on advanced treebanking at lrec2012*, 22–29. Hajič, Jan.

Incorporating Conjunctions in Polish

Agnieszka Patejuk
Institute of Computer Science, Polish Academy of Sciences and University of Oxford

Proceedings of the LFG'18 Conference

University of Vienna

Miriam Butt, Tracy Holloway King (Editors)

2018

CSLI Publications

pages 283–303

http://csli-publications.stanford.edu/LFG/2018

Keywords: Polish, conjunctions

Patejuk, Agnieszka. 2018. Incorporating Conjunctions in Polish. In Butt, Miriam, & King, Tracy Holloway (Eds.), *Proceedings of the LFG'18 Conference, University of Vienna*, 283–303. Stanford, CA: CSLI Publications.

Abstract

This paper provides a formal analysis of Polish conjunctions, coordinating and subordinating, which are placed in a non-standard position – inside the item they normally precede, possibly deeply embedded.

1 Introduction

Normally Polish coordinating conjunctions occur between the conjuncts that they join, see (1). Putting the coordinating conjunction inside the second conjunct results in ungrammaticality, as shown in (2), where the conjunction *a* 'and/while' follows *Janek*, the nominal subject of the second conjunct:

(1) {[Marysia grała na gitarze], a [Janek śpiewał]}.
Marysia.NOM played.3.SG.F on guitar and Janek.NOM sang.3.SG.M
'Marysia played the guitar and/while Janek was singing.'

(2) * {[Marysia grała na gitarze], [Janek a śpiewał]}.

Similarly, subordinating conjunctions (complementisers) are normally expected to be the first element of the subordinate clause (complementiser clause, CP), see (3). Placing the complementiser in a different position inside the subordinate clause results in ungrammaticality, as shown in (4), where the complementiser *ponieważ* 'because' follows *Marysia*, the nominal subject of the subordinate clause:

(3) [Janek śpiewał, [ponieważ [Marysia grała na gitarze]]].
Janek.NOM sang.3.SG.M because Marysia.NOM played.3.SG.F on guitar
'Janek was singing, because Marysia played the guitar.'

(4) * [Janek śpiewał, [Marysia ponieważ grała na gitarze]].

However, there are certain conjunctions in Polish, both coordinating and subordinating, which violate the constraints shown above. They may be placed inside the item they are normally supposed to precede: inside the second conjunct or inside the subordinate clause, respectively. Hence their traditional name, "incorporating conjunctions"; still, they are separate, independent words – they are not clitics. In spite of the non-standard position of such conjunctions, they are interpreted in the same way as run-of-the-mill conjunctions. The effect of using incorporating conjunctions is purely stylistic: they are characteristic of highly formal style.

The following conventions are adopted in this paper: partial c-structure bracketing is provided, only selected categories are labelled. In glossed examples and free translations, the incorporating conjunction is in blue, while the item which it follows is in red. Similarly, the partial f-structure of the item which the conjunction follows is marked in red, while the partial f-structure built by the conjunction is marked in blue. Some f-structures are simplified due to limited space available.

[†]This research is partially supported by the Polish Ministry of Science and Higher Education (MNiSW) within the CLARIN ERIC programme 2016–2018 (http://clarin.eu/). It was carried out during my stay at the Centre for Advanced Study in Oslo within the SynSem project led by Dag Haug and Stephan Oepen and during the Mobilność Plus mobility grant awarded by MNiSW.

1.1 Coordination

In (5) two sentences are coordinated – the first conjunct is headed by BYĆ 'be', while the head of the second conjunct is OTOCZYĆ 'surround'. The conjunction *zaś* 'but' is incorporating – instead of being placed between the two conjuncts, before the adjectival passive participle *dołączony* 'attached', *zaś* follows it. As a result, at the level of c-structure, *zaś* belongs to the second conjunct, as shown in (5).[1]

(5) {[Tekst był jawny], [[[dołączony]$_{AP}$ zaś [tajny]
text.NOM was.3.SG.M public.NOM attached.ACC but secret.ACC
protokół] otoczono szczelną zasłoną milczenia]}.
report.ACC surrounded.IMPERS airtight.INST curtain.INST silence.GEN
'The text was public, but the attached report was shrouded in secrecy.' (NKJP)

In terms of f-structure, *dołączony* is an ADJUNCT (ADJ), marked in red, of *protokół* 'report', the object (OBJ) of *otoczono*, an impersonal form of the verb OTOCZYĆ, the head of the second conjunct – see (6). Though *zaś* belongs to the second conjunct in terms of c-structure (it follows *dołączony*), its f-structure contribution (COORD-FORM attribute, marked in blue), is in a different place than the f-structure contribution of *dołączony* (marked in red) – it is in the same place where a non-incorporating conjunction (placed between the conjuncts) would contribute its f-structure.

(6)
$$\left[\left\{\left[\begin{array}{ll}\text{PRED} & \text{'BE}\langle\boxed{2}\rangle\boxed{1}\text{'}\\ \text{SUBJ} & \boxed{1}\left[\text{PRED}\quad \text{'TEXT'}\right]\\ \text{XC-P} & \boxed{2}\left[\begin{array}{ll}\text{PRED} & \text{'PUBLIC}\langle\boxed{1}\rangle\text{'}\\ \text{SUBJ} & \boxed{1}\end{array}\right]\end{array}\right],\left[\begin{array}{ll}\text{PRED} & \text{'SURROUND}\langle\boxed{3},\boxed{4},\boxed{5}\rangle\text{'}\\ \text{SUBJ} & \boxed{3}\left[\text{PRED}\quad \text{'PRO'}\right]\\ \text{OBJ} & \boxed{4}\left[\begin{array}{ll}\text{PRED} & \text{'REPORT'}\\ \text{ADJ} & \left\{\left[\text{PRED}\quad \text{'ATTACHED'}\right],\left[\text{PRED}\quad \text{'SECRET'}\right]\right\}\end{array}\right]\\ \text{OBL} & \boxed{5}\left[\text{PRED}\quad \text{'SILENCE...'}\right]\end{array}\right]\right\},\ \text{COORD-FORM}\quad \text{BUT}\right]$$

1.2 Subordination

In (7) the main clause headed by BYĆ is modified by a subordinate clause introduced by the incorporating subordinating conjunction (complementiser) – *bowiem* 'since' follows the adjective *podstawowym* 'main' which belongs to the subordinate clause at the level of c-structure.

[1] 'NKJP' marks attested examples taken from the National Corpus of Polish (NKJP, http://nkjp.pl, Przepiórkowski et al. 2012).

(7) [Takim zawodem jest zawód lekarza weterynarii,
such.INST profession.INST is.3.SG profession.NOM doctor.GEN veterinary.GEN
[[[podstawowym]$_{AP}$ bowiem [jego] zadaniem] jest ochrona]].
main.INST since his.GEN duty.INST is.3.SG protection.NOM
'Such a profession is the profession of a veterinarian, since protection is his main duty.' (NKJP)

In terms of f-structure, *podstawowym* is an adjunct of *zadaniem* 'duty', the predicative complement (XCOMP-PRED, XC-P) of *jest*, a form of the verb BYĆ 'be' – see the f-structure in (8). In spite of its non-standard c-structure position, *bowiem* makes the same f-structure contribution as without incorporation: it introduces the predicate SINCE, which is an adjunct (ADJ) of the main clause predicate (BE), and which takes the main predicate of the subordinate clause (BE) as its complement (COMP).

(8)
$$\begin{bmatrix}
\text{PRED} & \text{'BE<}\boxed{2}\text{>}\boxed{1}\text{'} \\
\text{SUBJ} & \boxed{1}\begin{bmatrix}\text{PRED} & \text{'PROFESSION_VET'}\end{bmatrix} \\
\text{XC-P} & \boxed{2}\begin{bmatrix}\text{PRED} & \text{'PROFESSION<}\boxed{1}\text{>'} \\ \text{SUBJ} & \boxed{1} \\ \text{ADJ} & \left\{\begin{bmatrix}\text{PRED} & \text{'SUCH'}\end{bmatrix}\right\}\end{bmatrix} \\
\text{ADJ} & \left\{\begin{bmatrix}\text{PRED} & \text{'SINCE<}\boxed{3}\text{>'} \\ \text{COMP} & \boxed{3}\begin{bmatrix}\text{PRED} & \text{'BE<}\boxed{5}\text{>}\boxed{4}\text{'} \\ \text{SUBJ} & \boxed{4}\begin{bmatrix}\text{PRED} & \text{'PROTECTION'}\end{bmatrix} \\ \text{XC-P} & \boxed{5}\begin{bmatrix}\text{PRED} & \text{'DUTY<}\boxed{4}\text{>'} \\ \text{SUBJ} & \boxed{4} \\ \text{ADJ} & \left\{\begin{bmatrix}\text{PRED} & \text{'MAIN'}\end{bmatrix}, \begin{bmatrix}\text{PRED} & \text{'HIS'}\end{bmatrix}\right\}\end{bmatrix}\end{bmatrix}\end{bmatrix}\right\}
\end{bmatrix}$$

1.3 Word order, non-incorporating use

Though the incorporating conjunctions in (5) and (7) follow the first word of the relevant clause, structural distance is greater – the conjunction follows an adjunct of an argument (object or predicative complement, respectively) of the main verb of the coordinate or subordinate clause, respectively. See f-structures in (6) and (8).

While it is sometimes advised to put the incorporating conjunction after the first word of the relevant clause, it may be far more distant both in terms of linear order (number of words) as well as syntactic distance (number of spanned dependents) – as it will be shown in examples to follow, incorporating conjunctions may be embedded deep inside the clause, inside almost any of its dependents.

Finally, though there are prescriptive rules claiming that conjunctions such as *zaś* (coordinating) and *bowiem* (subordinating) are obligatorily incorporating (they

must occur in the non-standard position: non-initially),[2] corpus data shows that these are no longer valid. There are numerous non-incorporating instances of conjunctions discussed here, many of which occur in edited texts such as newspapers:

(9) {[Padało na północy i wschodzie], zaś [susza dotknęła
rained.3.SG.N at north and east but drought.NOM affected.3.SG.F
przede wszystkim Polskę południową]}.
mostly Poland.ACC southern.ACC
'It mainly rained in the north and in the east, while the drought mostly affected southern Poland.' (NKJP)

(10) [W tym roku to wyjątkowo dobre trafienie, [bowiem [wzrasta
in this year this particularly good.NOM hit.NOM because grow.3.SG
popyt na margaryny i oleje]]].
demand.NOM on margarine and oils
'This year it was a particularly good choice, because the demand on margarine and oils is growing.' (NKJP)

It seems that such prescriptive rules change over time into stylistic suggestions which strongly advise the use of incorporation (as an indication of a careful writing style), but no longer consider the non-incorporating use to be unacceptable.

The existence of such variation in how conjunctions are used provides additional motivation for adopting a consistent, unified f-structure representation of both uses – in spite of the difference at the level of c-structure (caused by the presence or absence of incorporation), the interpretation of relevant utterances remains the same.

2 More data: distance

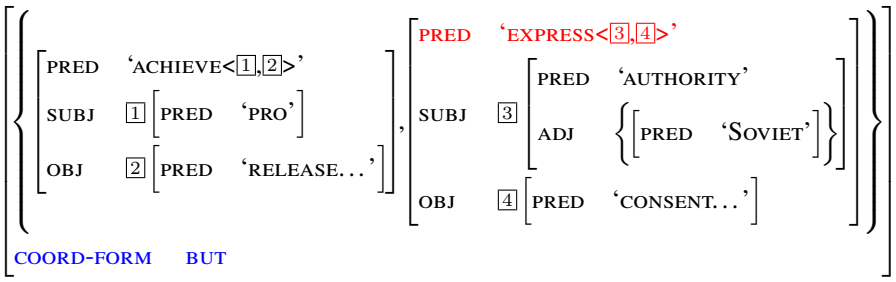
This section shows examples focusing on distance (linear and structural) between the incorporating conjunction and the position in which it is normally expected.

2.1 Coordination

In (11) two clauses are coordinated: the first conjunct is headed by UZYSKAĆ 'achieve', the head of the second conjunct is WYRAZIĆ 'express'. The coordinating conjunction *zaś* is incorporating: it is placed inside the second conjunct, it follows its third word – the verb *wyraziły* 'expressed', which is preceded by its subject, *władze sowieckie* 'Soviet authorities', so the linear distance is two phrases. See the corresponding f-structure in (12).

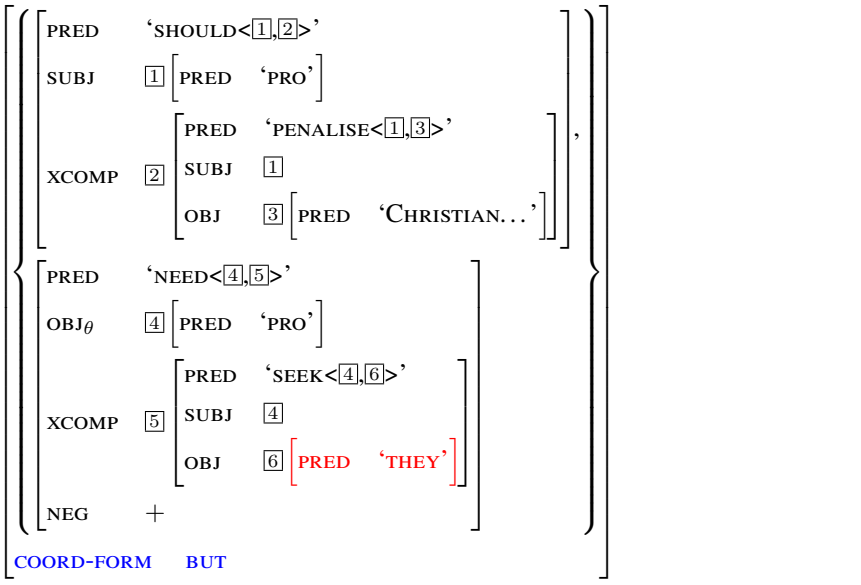
(11) {[Uzyskał zwolnienie wszystkich zakładników], [władze
achieved.3.SG.M release.ACC all.GEN hostages.GEN authorities.NOM
sowieckie [wyraziły]$_{I}$ zaś zgodę na ich powrót]}.
Soviet.NOM expressed.3.PL.F but consent.ACC on their return
'He achieved the release of all hostages, whereas the Soviet authorities agreed to their return.' (NKJP)

[2] For instance, this is the case in Świdziński 1992, where *zaś* and *bowiem* are only incorporating conjunctions, while *natomiast* and *więc* have standard, non-incorporating counterparts.

(12)


There are also 3 words before the coordinating conjunction *natomiast* 'but' in (13), but it is embedded deeper inside the second conjunct, so the structural distance is greater: *natomiast* follows *ich* 'them', the object (OBJ) of *wyszukiwać* 'seek', the infinitival complement (XCOMP) of *trzeba* 'need', the head of the second conjunct, which is preceded by negation (*nie*). See the f-structure in (14).

(13) {[Należy karać tych chrześcijan, którzy są oskarżeni
should penalise.INF those.ACC Christians.ACC who.NOM are accused.NOM
przed władzą], [nie trzeba [[ich]$_{NP}$ natomiast wyszukiwać]]}.
before authority NEG need them.GEN but seek.INF
'We should penalise those Christians who were accused by the authorities, but we do not need to search for them.' (NKJP)

(14)


2.2 Subordination

In (15) the incorporating subordinating conjunction (complementiser) *bowiem* 'since' is preceded by 4 words: the adjunct PP (*od dawna* 'long'), the main verb of the subordinate clause (*był* 'was') and its subject (*on* 'he'), which it follows – the linear distance is therefore 3 phrases. While the linear distance in (16) is just two words, the structural distance is greater – *bowiem* is embedded inside the first

phrase of the subordinate clause. It follows *wielu* ‘many’ – the numeral object of the preposition *w* ‘in’ which heads the adjunct prepositional phrase modifying *jest* ‘is’, which is in turn the head of the subordinate clause introduced by *bowiem*. It is worth noting that *bowiem* splits the numeral phrase: it separates its numeral head (*wielu*) from its nominal object (*wypadkach* ‘cases’). See the f-structure in (17).

(15) [Biskup uważał ich ciągle za swoich podwładnych,
bishop.NOM considered.3.SG.M them.ACC still for SELF subordinate
[od dawna był [on]$_{NP}$ bowiem zwierzchnikiem szkół]].
from long was.3.SG.M he.NOM since head.INST schools.GEN
‘The bishop still considered them to be his subordinates, since he has long been the head of schools.’ (NKJP)

(16) [Zrealizowanie zamierzenia spowoduje ogromne uciążliwości [...],
realising.NOM plan.GEN cause.3.SG great.ACC inconveniences.ACC
[[[w]$_{P}$ [[wielu]$_{NUM}$ bowiem [wypadkach]$_{NP}$]$_{NUMP}$]$_{PP}$ kolej jest
in many since cases railway.NOM is
jedynym środkiem lokomocji]].
only.INST means.INST transport
‘Realising this plan will cause great inconvenience, since in many cases the rail is the only means of transport.’ (NKJP)

(17)
$$\begin{bmatrix}
\text{PRED} & \text{‘CAUSE<}\boxed{1}\text{,}\boxed{2}\text{>’} \\
\text{SUBJ} & \boxed{1}\begin{bmatrix}\text{PRED} & \text{‘REALISING_PLAN’}\end{bmatrix} \\
\text{OBJ} & \boxed{2}\begin{bmatrix}\text{PRED} & \text{‘INCONVENIENCE’} \\ \text{ADJ} & \left\{\begin{bmatrix}\text{PRED} & \text{‘GREAT’}\end{bmatrix}\right\}\end{bmatrix} \\
\text{ADJ} & \left\{\begin{bmatrix}\text{PRED} & \text{‘SINCE<}\boxed{3}\text{>’} \\ \text{COMP} & \boxed{3}\begin{bmatrix}\text{PRED} & \text{‘BE<}\boxed{5}\text{>}\boxed{4}\text{’} \\ \text{SUBJ} & \boxed{4}\begin{bmatrix}\text{PRED} & \text{‘RAILWAY’}\end{bmatrix} \\ \text{XC-P} & \boxed{5}\begin{bmatrix}\text{PRED} & \text{‘MEANS_TRANSPORT<}\boxed{4}\text{>’} \\ \text{SUBJ} & \boxed{4} \\ \text{ADJ} & \left\{\begin{bmatrix}\text{PRED} & \text{‘ONLY’}\end{bmatrix}\right\}\end{bmatrix} \\ \text{ADJ} & \left\{\begin{bmatrix}\text{PRED} & \text{‘IN<}\boxed{6}\text{>’} \\ \text{OBJ} & \boxed{6}\begin{bmatrix}\text{PRED} & \text{‘MANY<}\boxed{7}\text{>’} \\ \text{OBJ} & \boxed{7}\begin{bmatrix}\text{PRED} & \text{‘CASE’}\end{bmatrix}\end{bmatrix}\end{bmatrix}\right\}\end{bmatrix}\end{bmatrix}\right\}
\end{bmatrix}$$

3 Complex data and interactions

3.1 Coordinating conjunction inside second conjunct's CP adjunct

In (18) the incorporating coordinating conjunction *natomiast* 'but', which joins sentences headed by verbs *przedstawiła* 'presented' and *musi* 'must', is placed inside the subordinate clause introduced by the complementiser *jeśli* 'if', an adjunct of *musi* – the head of the second conjunct. See the f-structure in (19).

(18) {[Komisja przedstawiła swoje stanowisko], [[[jeśli]$_{\text{COMP}}$
committee.NOM presented.3.SG.F its.ACC view.ACC if
natomiast [mamy to przedstawić Senatowi]], musi zostać
but have.1.PL this.ACC present.INF senate.DAT must.3.SG be.INF
zaprezentowany projekt uchwały]}.
presented.NOM draft.NOM resolution.GEN
'The commitee has presented its view, but, if we are to present it to the senate, the draft of the resolution must be presented.' (NKJP)

(19)
$$\left[\left\{\left[\begin{array}{ll}\text{PRED} & \text{'PRESENT<}\boxed{1}\text{,}\boxed{2}\text{>'}\\ \text{SUBJ} & \boxed{1}\left[\text{PRED}\quad\text{'COMMITTEE'}\right]\\ \text{OBJ} & \boxed{1}\left[\text{PRED}\quad\text{'VIEW'}\right]\end{array}\right],\right.$$
$$\left[\begin{array}{ll}\text{PRED} & \text{'MUST<}\boxed{3}\text{,}\boxed{4}\text{>'}\\ \text{SUBJ} & \boxed{3}\left[\text{PRED}\quad\text{'DRAFT'}\right]\\ \text{XCOMP} & \boxed{4}\left[\begin{array}{ll}\text{PRED} & \text{'PRESENT<}\boxed{3}\text{,NULL>'}\\ \text{SUBJ} & \boxed{3}\\ \text{PASSIVE} & +\end{array}\right]\\ \text{ADJ} & \left\{\left[\begin{array}{ll}\text{PRED} & \text{'IF<}\boxed{5}\text{>'}\\ \text{COMP} & \boxed{5}\left[\begin{array}{ll}\text{PRED} & \text{'HAVE<}\boxed{6}\text{,}\boxed{7}\text{>'}\\ \text{SUBJ} & \boxed{6}\left[\text{PRED}\quad\text{'PRO'}\right]\\ \text{XCOMP} & \boxed{7}\left[\begin{array}{ll}\text{PRED} & \text{'PRESENT<}\boxed{6}\text{,}\boxed{8}\text{,}\boxed{9}\text{>'}\\ \text{SUBJ} & \boxed{6}\\ \text{OBJ} & \boxed{8}\left[\text{PRED}\quad\text{'THIS'}\right]\\ \text{OBJ}_\theta & \boxed{9}\left[\text{PRED}\quad\text{'SENATE'}\right]\end{array}\right]\end{array}\right]\right\}\end{array}\right]\right\}$$
$$\left.\begin{array}{ll}\text{COORD-FORM} & \text{BUT}\end{array}\right]$$

(18) would also be grammatical if *jeśli* and *natomiast* were not adjacent (e.g.: *jeśli mamy to natomiast*, with the verb and the object of its infinitival complement between them), which shows that *jeśli natomiast* in (18) is not a multiword unit.

In (20) the incorporating coordinating conjunction *zaś* 'but' joining sentences headed by *trudno* 'hard' and *zdobywa* 'gain' is placed inside the adjunct subordinate

clause introduced by the complementiser *jeśli* 'if', which modifies *zdobywa* 'gain' – the head of the second conjunct. The two conjunctions, subordinating (*jeśli*) and coordinating (*zaś*), are separated by the impersonal marker *się* – a dependent of *czyni* 'do' (the main predicate of the subordinate clause introduced by *jeśli*). In (20) *zaś* could also follow *tego* 'this', the object of *czyni*: *jeśli się tego zaś nie czyni*, as in (21) – the f-structure in (22) corresponds to this possibility.

(20) {[Trudno wydawać pismo w regularnych odstępach czasu], [[jeśli hard publish.INF magazine.ACC in regular intervals time if [[się]$_{\text{PART}}$ zaś tego nie czyni]], nie zdobywa się stałego odbiorcy]}.
REFL but this.GEN NEG do.3.SG NEG gain.3.SG REFL regular reader.GEN
'It is hard to publish the magazine in regular intervals, but, if you do not do this, you do not gain regular readers.' (NKJP)

(21) {[Trudno wydawać pismo w regularnych odstępach czasu], [[jeśli [się [tego]$_{\text{NP}}$ zaś nie czyni]], nie zdobywa się stałego odbiorcy]}.

(22)
$$\left[\left\{\left[\begin{array}{ll}\text{PRED} & \text{'HARD<}\boxed{1}\text{,}\boxed{2}\text{>'}\\ \text{SUBJ} & \boxed{1}\left[\begin{array}{ll}\text{PRED} & \text{'PUBLISH<}\boxed{2}\text{,}\boxed{3}\text{>'}\\ \text{SUBJ} & \boxed{2}\\ \text{OBJ} & \boxed{3}\left[\text{PRED}\quad\text{'MAGAZINE'}\right]\end{array}\right]\\ \text{OBJ}_\theta & \boxed{2}\left[\text{PRED}\quad\text{'PRO'}\right]\end{array}\right],\right.$$
$$\left.\left[\begin{array}{ll}\text{PRED} & \text{'GAIN<}\boxed{4}\text{,}\boxed{5}\text{>'}\\ \text{SUBJ} & \boxed{4}\left[\text{PRED}\quad\text{'PRO'}\right]\\ \text{OBJ} & \boxed{5}\left[\begin{array}{ll}\text{PRED} & \text{'READER'}\\ \text{ADJ} & \left\{\left[\text{PRED}\quad\text{'REGULAR'}\right]\right\}\end{array}\right]\\ \text{ADJ} & \left\{\left[\begin{array}{ll}\text{PRED} & \text{'IF<}\boxed{6}\text{>'}\\ \text{COMP} & \boxed{6}\left[\begin{array}{ll}\text{PRED} & \text{'DO<}\boxed{7}\text{,}\boxed{8}\text{>'}\\ \text{SUBJ} & \boxed{7}\left[\text{PRED}\quad\text{'PRO'}\right]\\ \text{OBJ} & \boxed{8}\left[\text{PRED}\quad\text{'THIS'}\right]\\ \text{NEG} & +\\ \text{IMPERSONAL} & +\end{array}\right]\end{array}\right]\right\}\\ \text{NEG} & +\end{array}\right]\right\}$$
$$\text{COORD-FORM}\quad\text{BUT}\Bigg]$$

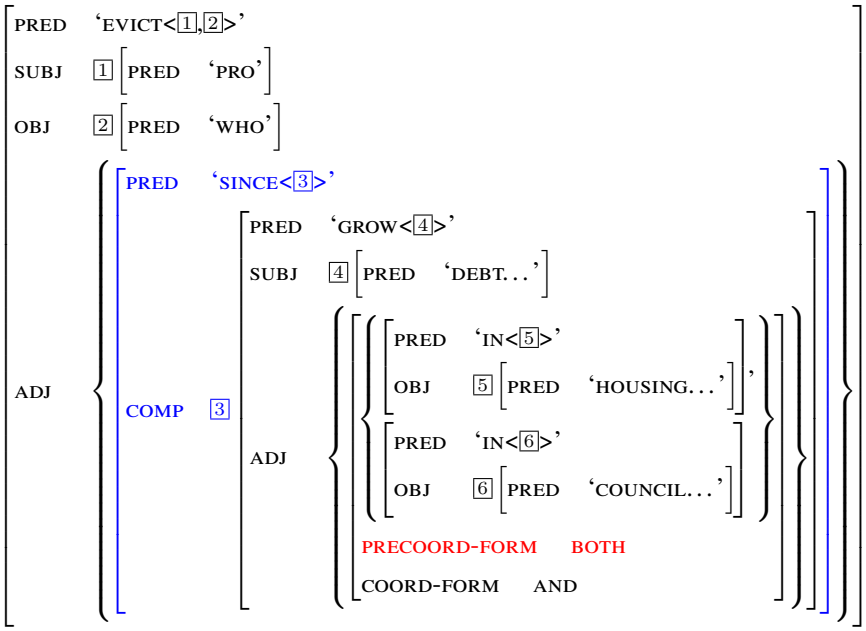
3.2 Subordinating conjunction inside coordinate phrase

This section discusses examples where the subordinating conjunction is contained in a coordinate phrase which is one of the dependents of the subordinate clause.

In a sense, this can be seen as the opposite of the configuration presented in §3.1, where the coordinating conjunction was placed inside the adjunct subordinate clause which in turn was a dependent of the second conjunct.

In (23) the subordinating conjunction *bowiem* 'since' follows *zarówno* 'both'– a preconjunction which is a part of the coordinate adjunct phrase which consists of two prepositional phrases: *w spółdzielniach mieszkaniowych* 'in housing associations' and *w kwaterunkach* 'in council housing', joined by the conjunction *jak i* 'and'. This coordinated PP is a modifier of *rośnie* 'grows' — the main predicate of the subordinate clause introduced by *bowiem*. Though the main verb is elided in (23), the corresponding f-structure in (24) represents it explicitly as EVICT – it can be recovered from context (the preceding sentence).[3]

(23) {A [byłoby kogo, [{[zarówno]$_{\text{PRECONJ}}$ bowiem [w
and would be who.ACC both since in
spółdzielniach mieszkaniowych]$_{\text{PP}}$, [jak i]$_{\text{CONJ}}$ [w kwaterunkach]$_{\text{PP}}$}
housing associations and in council housing
rośnie zadłużenie lokatorów]]}.
grow.3.SG debt.NOM tenants.GEN
'And there would be who [to evict], since the tenants' debt is growing both in housing associations and in council housing.' (NKJP)

(24)


(25) is similar to (23) in that the subordinating conjunction (*bowiem*) follows a preconjunction (*nie tylko* 'not only'), but the path to the grammatical function of the coordinate phrase in which *bowiem* is embedded is different. While in (23) the coordinate phrase corresponds to an adjunct of the subordinate clause, in (25) the coordinate phrase corresponds to the main predicate(s) of the subordinate clause. Apart

[3]Though (23) is a sentence starting with a conjunction (see §3.4), it is not represented in (24).

from the opening preconjunction, it consists of two verbal phrases, *będę mówił o sobie bez wstydu* 'I will speak about myself without shame' and *będę się chełpił swoją formą fizyczną* 'I will boast about my physical fitness', joined by the conjunction *ale* 'but (also)'. See the corresponding simplified f-structure in (26).

(25) [Tutaj popełnię podwójny grzech, [{[nie tylko]$_{\text{PRECONJ}}$ bowiem
here commit.1.SG double.ACC sin.ACC not only since
[będę mówił o sobie bez wstydu]$_{\text{IP}}$, [ale]$_{\text{CONJ}}$ [będę się
will.1.SG speak about myself without shame but will.1.SG REFL
chełpił swoją formą fizyczną]$_{\text{IP}}$}]].
boast own form.INST physical
'I will commit a double sin, since not only will I speak about myself without shame, but also I will boast about my physical fitness.' (NKJP)

(26)
$$\begin{bmatrix}
\text{PRED} & \text{'COMMIT<}\boxed{1}\text{,}\boxed{2}\text{>'} \\
\text{SUBJ} & \boxed{1}\begin{bmatrix}\text{PRED} & \text{'PRO'}\end{bmatrix} \\
\text{OBJ} & \boxed{2}\begin{bmatrix}\text{PRED} & \text{'SIN'} \\ \text{ADJ} & \left\{\begin{bmatrix}\text{PRED} & \text{'DOUBLE'}\end{bmatrix}\right\}\end{bmatrix} \\
\text{ADJ} & \left\{\begin{bmatrix}
\text{PRED} & \text{'SINCE<}\boxed{3}\text{>'} \\
\text{COMP} & \boxed{3}\begin{bmatrix}
\left\{\begin{bmatrix}
\text{PRED} & \text{'SPEAK<}\boxed{4}\text{,}\boxed{5}\text{>'} \\
\text{SUBJ} & \boxed{4}\begin{bmatrix}\text{PRED} & \text{'PRO'}\end{bmatrix} \\
\text{OBL} & \boxed{5}\begin{bmatrix}\text{PRED} & \text{'SELF'} \\ \text{PFORM} & \text{ABOUT}\end{bmatrix} \\
\text{ADJ} & \left\{\begin{bmatrix}\text{PRED} & \text{'WITHOUT_SHAME'}\end{bmatrix}\right\}
\end{bmatrix}, \right. \\
\left.\begin{bmatrix}
\text{PRED} & \text{'BOAST<}\boxed{4}\text{,}\boxed{6}\text{>'} \\
\text{SUBJ} & \boxed{4} \\
\text{OBJ} & \boxed{6}\begin{bmatrix}\text{PRED} & \text{'FITNESS'} \\ \text{ADJ} & \left\{\begin{bmatrix}\text{PRED} & \text{'OWN'}\end{bmatrix}\right\}\end{bmatrix}
\end{bmatrix}\right\} \\
\text{PRECOORD-FORM} & \text{NOT_ONLY} \\
\text{COORD-FORM} & \text{BUT_ALSO}
\end{bmatrix}\right\}
\end{bmatrix}$$

3.3 Gapping

Incorporating conjunctions also occur with gapping – a special type of coordination where the predicate of the second conjunct is elided. In such examples – rather than between the conjuncts – the conjunction joining clauses is placed inside the gapped conjunct (typically the second conjunct). The f-structure representations provided below follow the analysis of gapping offered in Patejuk and Przepiórkowski 2017,

which accounts for differences in agreement features, independent structural case assignment and unlike category coordination.

In (27) the coordinating conjunction *zaś* follows *ona* 'she' which is the subject of the gapped second conjunct (which would be a form of MIEĆ 'have'), see (28):

(27) {[W chwili ich poznania miał lat 20], [[ona]$_{NP}$ zaś 36]}.
in time they.GEN meeting had.3.SG.M years 20.ACC she.NOM but 36.ACC
'At the time of meeting them he was 20, and she (was) 36.' (NKJP)

(28)
$$\left[\left\{\left[\begin{array}{ll}\textsc{pred} & \text{'HAVE<}\boxed{1}\text{,}\boxed{2}\text{>'}\\ \textsc{subj} & \boxed{1}\left[\textsc{pred}\quad \text{'PRO'}\right]\\ \textsc{obj} & \boxed{2}\left[\begin{array}{ll}\textsc{pred} & \text{'20<}\boxed{3}\text{>'}\\ \textsc{obj} & \boxed{3}\left[\textsc{pred}\quad \text{'YEAR'}\right]\end{array}\right]\end{array}\right]\boxed{0},\left[\begin{array}{ll}\textsc{pred} & \text{'HAVE<}\boxed{4}\text{,}\boxed{5}\text{>'}\\ \textsc{subj} & \boxed{4}\left[\textsc{pred}\quad \text{'SHE'}\right]\\ \textsc{obj} & \boxed{4}\left[\textsc{pred}\quad \text{'36'}\right]\end{array}\right]\right\}\right]$$
$$\textsc{coord-form}\quad \text{BUT}$$
$$\textsc{local}\ \boxed{0}$$

In (29) the coordinating conjunction *natomiast* follows *ucznia* 'schoolboy', which is the head of the oblique complement (OBL) of the gapped clause, see (30):[4]

(29) {[Do każdej uczennicy Herr Poliffka zwracał się per
to every.F schoolgirl Herr.NOM Poliffka.NOM addressed.3.SG.M REFL as
"Franciszka"], [[[do]$_P$ [[każdego]$_{AP}$ [ucznia]$_{NP}$ natomiast]]$_{PP}$ per "Alojzy"]}.
Franciszka.F to every.M schoolboy but as Alojzy.M
'Herr Poliffka addressed every schoolgirl as "Franciszka" and every schoolboy as "Alojzy".' (NKJP)

(30)
$$\left[\left\{\left[\begin{array}{ll}\textsc{pred} & \text{'ADDRESS<}\boxed{1}\text{,}\boxed{2}\text{,}\boxed{3}\text{>'}\\ \textsc{subj} & \boxed{1}\left[\textsc{pred}\quad \text{'HERR_POLIFFKA'}\right]\\ \textsc{obl} & \boxed{2}\left[\begin{array}{ll}\textsc{pred} & \text{'SCHOOLGIRL'}\\ \textsc{pform} & \text{TO}\end{array}\right]\\ \textsc{obl2} & \boxed{3}\left[\begin{array}{ll}\textsc{pred} & \text{'FRANCISZKA'}\\ \textsc{pform} & \text{AS}\end{array}\right]\end{array}\right]\boxed{0},\left[\begin{array}{ll}\textsc{pred} & \text{'ADDRESS<}\boxed{1}\text{,}\boxed{4}\text{,}\boxed{5}\text{>'}\\ \textsc{subj} & \boxed{1}\\ \textsc{obl} & \boxed{4}\left[\begin{array}{ll}\textsc{pred} & \text{'SCHOOLBOY'}\\ \textsc{pform} & \text{TO}\end{array}\right]\\ \textsc{obl2} & \boxed{5}\left[\begin{array}{ll}\textsc{pred} & \text{'ALOJZY'}\\ \textsc{pform} & \text{AS}\end{array}\right]\end{array}\right]\right\}\right]$$
$$\textsc{coord-form}\quad \text{BUT}$$
$$\textsc{local}\ \boxed{0}$$

[4]Though OBL2 is used in (30) as the grammatical function corresponding to PPs with *per* 'as', it should perhaps be a manner oblique (OBL-MOD). As shown below, a manner adverb (*normalnie* 'normally') can be coordinated with a PP consisting of *per* and a nominative/vocative nominal:

(i) Zwracali się do niego {[normalnie], [per "pan"]}.
addressed.3.PL.M REFL to him normally as sir.NOM
'They addressed him normally, as "sir".' (NKJP)

Still, the grammatical functions in (30) have no bearing on general points made in this paper.

3.4 In stand-alone clauses

Though many people are taught that they should never start a sentence with a conjunction, this is not considered a rule by authoritative sources such as Poradnia językowa PWN,[5] whose stance is supported by corpus data which abounds in examples starting with a conjunction, many of which come from edited sources.

(31) {Ale [myślicie prawidłowo]}.
but think.2.PL correctly
'But your thinking is right.' (NKJP)

(32) [Że [istnieje Europa]]!
that exist.3.SG Europe.NOM
'That Europe exists!' (NKJP)

(33) [Jeśli [nie liczyć coraz mniejszej ilości włosów]].
if NEG count.INF ever smaller.GEN amount.GEN hair.GEN
'If you do not consider the ever smaller amount of hair.' (NKJP)

(31) starts with the coordinating conjunction *ale* 'but', followed by the stranded second conjunct. (32)–(33) begin with the subordinating conjunction (complementiser): *że* 'that' in (32) introduces an argument subordinate clause (subcategorised), while *jeśli* 'if' in (33) introduces an adjunct subordinate clause (not subcategorised).

Let us proceed to examples with incorporating conjunctions. In (34) the subordinating conjunction *bowiem* 'since' follows *będą* 'will' – the copula heading the stand-alone subordinate clause introduced by *bowiem*. See the f-structure in (35).

(34) [Przelewy z unijnej kasy nie [będą]$_{1}$ bowiem automatyczne].
transfers.NOM from EU cash box NEG will.3.PL since automatic.NOM
'Since transfers of EU money will not be automatic.' (NKJP)

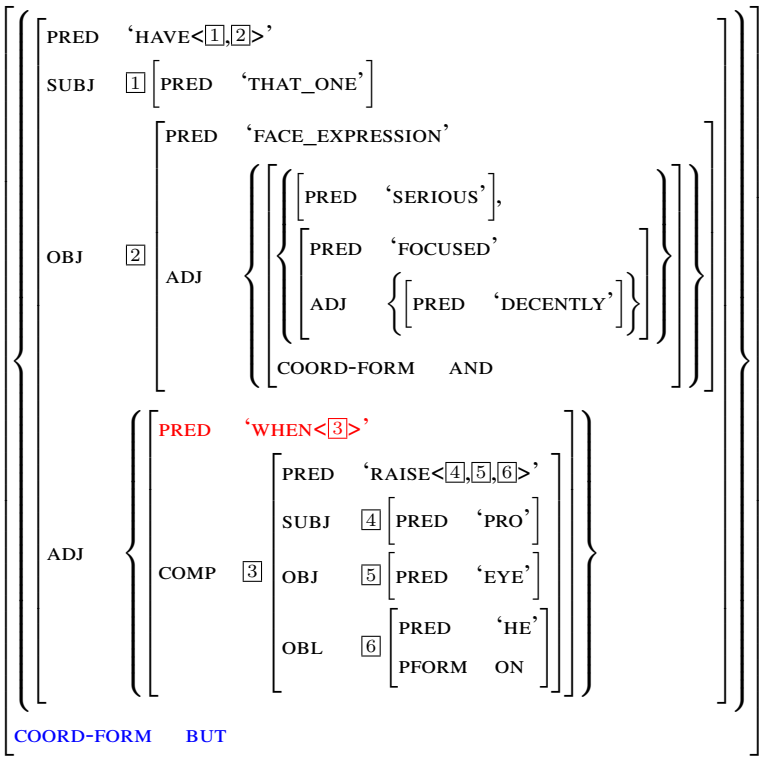
(35)
$$\begin{bmatrix} \text{PRED} & \text{'SINCE<}\boxed{1}\text{>'} \\ \text{COMP} & \boxed{1}\begin{bmatrix} \text{PRED} & \text{'BE<}\boxed{3}\text{>}\boxed{2}\text{'} \\ \text{SUBJ} & \boxed{2}\begin{bmatrix}\text{PRED} & \text{'TRANSFER'}\end{bmatrix} \\ \text{XC-P} & \boxed{3}\begin{bmatrix}\text{PRED} & \text{'AUTOMATIC<}\boxed{2}\text{>'} \\ \text{SUBJ} & \boxed{2}\end{bmatrix} \\ \text{NEG} & + \end{bmatrix} \end{bmatrix}$$

(36) is analogous to examples discussed in §3.1, where the incorporating coordinating conjunction joining two clauses is placed inside the adjunct CP modifying the second conjunct. However, the first conjunct is not present in (36) – the coordinating conjunction *zaś* 'but' follows the subordinating conjunction (complementiser) *gdy* 'when', which introduces the subordinate clause modifying the verb *miał* 'had', which is the head of the stranded second conjunct. See the f-structure in (37).

[5] https://sjp.pwn.pl/poradnia/haslo/spojnik-na-poczatku-zdania;2374.html: "There is, however, no rule that would ban using a conjunction at the beginning of a sentence." ("Nie ma jednak reguły, która zabraniałaby użycia spójnika na początku zdania.").

(36) {[[[Gdy]$_{\text{COMP}}$ zaś [podniósł nań oczy]], tamten miał
when but raised.3.SG.M on him eyes.ACC that one.NOM had.3.SG.M
minę poważną i przyzwoicie skupioną]}.
face expression.ACC serious.ACC and decently focused.ACC
'But when he raised his eyes on him (== he looked at him), that one's face expression was serious and decently focused.' (NKJP)

(37)

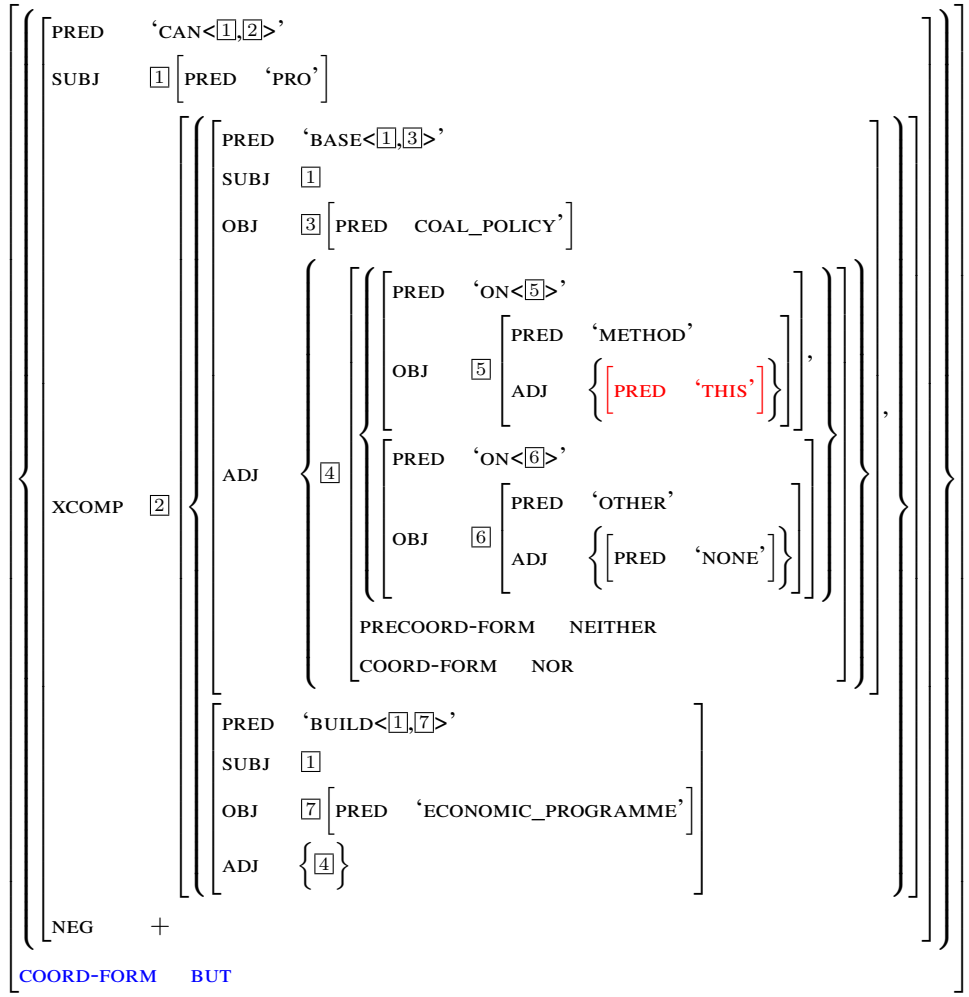


While §3.2 discussed examples where an incorporating subordinating conjunction (complementiser) is placed inside a coordinate phrase which is a dependent of the subordinate clause, in (38) the incorporating coordinating conjunction *zaś* 'but' is put inside a coordinate phrase which is a dependent of the stranded second conjunct. In (38) *zaś* follows *tej* 'this' – an adjectival modifier of *metodzie* 'method', which is the nominal head of the first conjunct of a coordinate phrase consisting of two prepositional phrases (*{[Na tej metodzie], ani [na żadnej innej]}*), which is a shared oblique complement of coordinated verbs *opierać* 'base' and *budować* 'build' (each of which has its own, additional dependents), which are a complement of the verb *można* 'can, may', which is the head of the stranded second conjunct joined by *zaś*. See the corresponding f-structure in (39).

(38) {[{[[Na]$_{\text{P}}$ [[tej]$_{\text{AP}}$ zaś [metodzie]$_{\text{N}}$]$_{\text{NP}}$]$_{\text{PP}}$, ani [na żadnej innej]$_{\text{PP}}$}
on this but method neither on none other
[...] nie można opierać ani polityki węglowej, ani też budować
NEG can base.INF neither policy.GEN coal.GEN nor also build.INF

długofalowego programu gospodarczego.]}
long-term.GEN programme.GEN economic.GEN
'But neither the coal policy, nor building a long-term economic programme can be based on this method, nor on any other.' (NKJP)

(39)



4 Analysis and formalisation

Since sentences with incorporating conjunctions (coordinating and subordinating) are interpreted in the same way as when incorporation is not involved, the proposed analysis assumes that while incorporating conjunctions are placed in a non-standard c-structure position (inside the item they join, rather than before it), the f-structure of utterances containing incorporating conjunctions is the same as when their non-incorporating counterparts are used. The proposed analysis relies exclusively on existing LFG formal devices – it does not require introducing any new mechanisms or modifications. It was implemented and verified in XLE (Crouch et al. 2011).

The analysis aims to capture two basic insights: first, that the incorporating conjunction is always non-initial, it always follows some category (see (40)) – it

may be the first word of the joined phrase, coordinate or subordinate, but it may also be more distant both in terms of linear distance as well as structural distance. The second insight is that the incorporating conjunction builds its f-structure from its non-standard c-structure position – from inside the item which it joins (second conjunct or a subordinate clause), from inside the category which it immediately follows. As a result, while the c-structure position of incorporating conjunctions is non-standard, the resulting f-structure is fully parallel to f-structures with conjunctions occupying a standard position – following the way they are interpreted (as mentioned earlier, the difference between these two uses is purely stylistic).

Drawing on these insights, the basic idea behind the formal analysis of this phenomenon is to minimally affect the organisation of the rest of the grammar, putting the burden of producing the appropriate analysis on elements related to incorporating conjunctions. The minimal additions discussed in the following sections make it possible to account for this phenomenon without affecting the general grammar – using standard formal devices available in LFG, without making any changes to the formalism. This analysis can be seen as a simple overlay on the existing grammar.

4.1 METARULEMACRO

The first insight mentioned above, namely that incorporating conjunctions (coordinating and subordinating) always follow some other category, is formalised using the metarule mechanism described in the XLE documentation,[6] which seems to suit this purpose perfectly: "The METARULEMACRO is useful for expressing generalizations that operate across all the rules of the grammar, such as coordination, brackets, parentheses, and linear precedence." The XLE documentation (ibid.) defines METARULEMACRO as follows: "The effective right-hand side of each rule in the grammar is taken to be the result of applying that macro to three parameters: the mother category of that rule, the base name of the mother category (for the complex categories of parameterized rules), and the specified right-hand side of the rule." The relevant part of the METARULEMACRO used in the proposed analysis is provided in (40) – it has 3 parameters: CAT is the mother category, BASECAT is the base name of the mother category (not discussed here) and RHS is the right-hand side of the rule to which METARULEMACRO applies. The first disjunct of (40) is RHS – it ensures that every rule in the grammar trivially rewrites to itself, so it is not modified by METARULEMACRO. The last disjunct of (40) contains a call to the macro METARULEMACRO-INK defined in (41), which applies relevant changes.

(40) METARULEMACRO(CAT BASECAT RHS) ≡
{ RHS | ... | @(METARULEMACRO-INK CAT) }

(41) METARULEMACRO-INK(CAT) ≡ CAT { CONJINK | COMPINK }

CAT in (41) is a variable corresponding to any category defined in the grammar; it is followed by an incorporating conjunction, either coordinating (CONJINK) or subordinating (COMPINK). For example, when CAT is an NP, the output of (41) are two

[6] http://ling.uni-konstanz.de/pages/xle/doc/notations.html#N3.5

rules: NP → NP CONJINK and NP → NP COMPINK. Since there is no f-structure annotation in (41), all right-hand side elements are co-heads (with the default ↑=↓ annotation). By default the METARULEMACRO-INK applies to all categories defined in the grammar. However, if need be, its application can be restricted to selected categories using an additional constraint, as shown in (42).

(42) METARULEMACRO-INK(CAT) ≡ CAT { CONJINK | COMPINK }
ϵ: CAT $\in_c$ {NP AP, ADVP, COMP, NUM, PART, I, V, PRECONJ...}

NP is used in examples (13), (15), (21) and (27) (pronouns) and (29) (noun), AP in (5) (passive participle), (7) and (38) (adjectives), NUM in (16) (numeral), PART (particle) in (21), C (complementiser, subordinating conjunction) in (18) and (36), I in (11) and (34), PRECONJ (preconjunction) in (23) and (25).

4.2 Templates in lexical entries

The second, main part of the analysis is formalised in the lexical entries of incorporating conjunctions – they implement the insight that such conjunctions, despite being embedded inside the conjunct or inside the subordinate phrase, build their f-structure bottom up, higher in the f-structure than the place where they are located in terms of c-structure. This is achieved in templates provided in (44)–(45) with the help of constraints relying on inside-out functional uncertainty, both of which use the GF variable defined in (43):[7]

(43) GF ≡ {SUBJ|OBJ|OBJ$_\theta$|OBL|XCOMP|XCOMP-PRED|COMP|ADJUNCT ∈}

(44) CONJ-INK(P) ≡
(∈ GF* ↑)=%G
(%G COORD-FORM)= P
¬(ADJUNCT %G)

(45) COMP-INK(P) ≡
(COMP GF* ↑)=%B
(%B PRED)='P<(%B COMP)>'

The template in (44) is called inside lexical entries of coordinating incorporating conjunctions such as *zaś* and *natomiast* – the P parameter corresponds to the lemma of the conjunction. The first line of (44), (∈ GF* ↑)=%G, is a definition of the path in which the f-structure of the incorporating conjunction is built – it uses an inside-out equation coupled with functional uncertainty, allowing the conjunction to build its structure going up the path consisting of any sequence (including zero)[8] of GF defined in (43), with an obligatory set element at the very end of the path (because coordinate structures are modelled as sets). This path is assigned to the %G local name[9] (variable) used in the remaining constraints: the second constraint, (%G COORD-FORM)= P, introduces P, the conjunction's lemma, as the value of the

[7](43), (44) and (47) use full names of grammatical functions, so ADJUNCT and XCOMP-PRED instead of ADJ and XC-P, respectively, used in f-structures in order to save space.

[8]E.g. when the conjunction follows the main verb, as in (11), or its co-head, as in (20).

[9]http://ling.uni-konstanz.de/pages/xle/doc/notations.html#N4.1.6

COORD-FORM attribute (coordinating conjunction form) in %G path, which is the f-structure which contains the set with the conjuncts – yielding the hybrid f-structure typical of coordination in LFG. Finally, the third constraint, ¬(ADJUNCT %G), ensures that the f-structure introduced by the coordinating incorporating conjunction is not placed inside the ADJUNCT grammatical function – its value is also a set, so it would also satisfy the condition of having a set element at the end of the path in %G. However, this simple negative constraint precludes this. As a result, the constraints stated in (44) are only satisfied by coordinate structures, as intended.

The idea behind the template in (45), which handles incorporating subordinating conjunctions (complementisers) such as *bowiem* 'since', is roughly similar to (44) but the details are different. This is because incorporating subordinating conjunctions (heading adjunct CPs),[10] unlike coordinating conjunctions, introduce a PRED attribute, which takes a COMP argument containing the subordinate clause. As a result, the first constraint in (45), (COMP GF* ↑)=%B, defines an inside-out path which passes through any sequence (including zero)[11] of GF and ends with a COMP grammatical function. This path is assigned to %B local name (variable). It is used twice in the second constraint, (%B PRED)='P<(%B COMP)>', which introduces the PRED value of the subordinating conjunction (complementiser): it consists of P, which corresponds to its lemma, and (%B COMP), which is the closed clausal complement required by this subordinating conjunction (complementiser).

4.3 C-structure rules

The last element necessary for this analysis to work are c-structure rules. The interaction of c-structure rules presented in this section, the metarule in (41)/(42) and the lexical entries of incorporating conjunctions, coordinating and subordinating, which call templates in (44) and (45), respectively, results in creating appropriate dependencies in f-structure despite the non-standard c-structure position of such conjunctions – the functional uncertainty used in (44) and (45) is constrained by the f-structure built by the rest of the sentence.

The rule in (46) handles coordination with an incorporating conjunction embedded somewhere inside the second conjunct (including gapping, see §3.3).

(46) S → S COMMA S
  ↓∈↑ ↓∈↑

This is the same rule that is used for asyndetic coordination, where no coordinating conjunction is present – the conjuncts are only separated by the comma (COMMA). Unlike under asyndetic coordination, under incorporating coordination the coordinating conjunction is present (CONJINK), but it is placed inside the second conjunct (rather than between the conjuncts). The metarule handling incorporating

[10] In subcategorised CPs the subordinating conjunction (complementiser) does not contribute its PRED, which is contributed by the main verb of the subordinate clause, but instead it contributes the COMP-FORM attribute which hosts the form of the complementiser (subordinating conjunction).

[11] E.g. when the subordinating conjunction (complementiser) follows the main verb, as in (34), or its co-head (for instance, the marker *się*, or an auxiliary).

conjunctions (see (41)) makes sure that their c-structure position is non-initial – they must follow some category (see the constraint in (42)). The template in (44), called inside the lexical entry of the incorporating coordinating conjunction, ensures that it builds an appropriate f-structure from its non-standard c-structure position.

The following two rules account for examples with an incorporating subordinating conjunction (complementiser) embedded somewhere inside the subordinate clause which it introduces.

(47) S → S CP-INK
$\downarrow \in (\uparrow \text{ADJUNCT})$

(48) CP-INK → S
$(\uparrow \text{COMP}) = \downarrow$
$(\uparrow \text{PRED}) =_c \text{BOWIEM}$

(47) is the top-level rule, where CP-INK, a CP with an incorporating subordinating conjunction, is added as a modifier (ADJUNCT) of the preceding sentence (S). (48) is the rule which builds the subordinate CP containing the incorporating subordinating conjunction (complementiser). It contains the $(\uparrow \text{COMP}) = \downarrow$ annotation which matches the f-structure built by the template in (45) called by the incorporating subordinating conjunction (which requires a COMP). The constraint on its PRED attribute ($(\uparrow \text{PRED}) =_c \text{BOWIEM}$) ensures that it is contains the relevant subordinating conjunction – since there is no COMP(INK) preceding S in (48), the subordinating conjunction must be embedded somewhere inside S. The metarule handling incorporating conjunctions in (41) ensures that these are always non-initial – they always follow some category (as defined in (42)), so the incorporating subordinating conjunction (complementiser) must be in a non-standard c-structure position.

Finally, rules in (49)–(50) make it possible to account for incorporating conjunctions in stand-alone clauses (without the first conjunct or without the main clause modified by the subordinate clause) discussed in §3.4.

(49) ROOT → S
$\downarrow \in \uparrow$
$(\uparrow \text{COORD-FORM}) \in_c \{\text{ZAŚ}, \text{NATOMIAST}\}$

(49) creates a coordination structure with a singleton set containing only the stranded second conjunct (see examples in (36) and (38)). The constraint on the form of the coordinating conjunction ($(\uparrow \text{COORD-FORM}) \in_c \{\text{ZAŚ}, \text{NATOMIAST}\}$) ensures that it must be present. Similarly as in (48), since there is no coordinating conjunction category (CONJ(INK)) in the rule in (49), it must be embedded somewhere in S. The metarule handling incorporating conjunctions in (41) ensures that these must follow some category (see the constraint in (42)) – as a result, such conjunctions are never initial and appear instead in a non-standard c-structure position.

By contrast, the rule in (50) builds a structure where the incorporating subordinating conjunction (complementiser), which is the head of the stand-alone, stranded subordinate clause, serves as the main predicate (as in (34)).

(50) ROOT → CP-INK

Constraints ensuring that there is an incorporating subordinating conjunction (complementiser) in CP-INK are imposed in the rule in (48), which interacts with relevant templates and metarules, as explained above when discussing (47)–(48).

5 Conclusion

This paper presented an implemented LFG analysis of Polish incorporating conjunctions, both coordinating and subordinating, which have a non-standard c-structure position – they are embedded inside the c-structure of the item that they are normally expected to precede: the second conjunct or the adjunct subordinate clause. Despite this difference in c-structure, the proposed analysis successfully accounts for the f-structure contribution of incorporating conjunctions which is fully consistent with corresponding sentences without incorporation. The motivation for adopting such an analysis comes from the fact that that incorporating conjunctions, despite the their non-standard c-structure position, are interpreted in the same way as their non-incorporating counterparts (the difference is purely stylistic).

The proposed analysis avoids making any changes to the general grammar (understood as all parts of the grammar before extending it so as to account for incorporating conjunctions). Instead, using standard, existing LFG mechanisms, the proposed analysis puts the burden of producing appropriate f-structures on incorporating conjunctions (both coordinating and subordinating) – they build relevant f-structure fragments from their non-standard c-structure position in a bottom up manner, using functional uncertainty.

Though the proposed analysis of incorporating conjunctions is very simple, it successfully accounts for complex interactions, which include: embedding the incorporating coordinating conjunction inside the adjunct CP which modifies the second conjunct (§3.1), embedding the incorporating subordinating conjunction (complementiser) inside one of its coordinate phrase dependents (§3.2) and gapping (a variety of coordination where the predicate in the second conjunct is elided, §3.3). It also covers the occurrence of incorporating conjunctions in stand-alone clauses (§3.4), where the first conjunct is missing or where there is no main clause on which the subordinate adjunct CP normally depends.

References

Crouch, Dick, Dalrymple, Mary, Kaplan, Ron, King, Tracy, Maxwell, John and Newman, Paula. 2011. XLE Documentation. http://www2.parc.com/isl/groups/nltt/xle/doc/xle_toc.html.

Patejuk, Agnieszka and Przepiórkowski, Adam. 2017. Filling the Gap. In Miriam Butt and Tracy Holloway King (eds.), *The Proceedings of the LFG'17 Conference*, pages 327–347, Stanford, CA: CSLI Publications.

Przepiórkowski, Adam, Bańko, Mirosław, Górski, Rafał L. and Lewandowska-Tomaszczyk, Barbara (eds.). 2012. *Narodowy Korpus Języka Polskiego*. Warsaw: Wydawnictwo Naukowe PWN.

Świdziński, Marek. 1992. *Gramatyka formalna języka polskiego*. Warsaw: Wydawnictwa Uniwersytetu Warszawskiego.

Predicative Constructions with Infinitival and Clausal Subjects in Polish

Agnieszka Patejuk
Institute of Computer Science, Polish Academy of Sciences and University of Oxford

Adam Przepiórkowski
Institute of Computer Science, Polish Academy of Sciences and Institute of Philosophy, University of Warsaw

Proceedings of the LFG'18 Conference

University of Vienna

Miriam Butt, Tracy Holloway King (Editors)

2018

CSLI Publications

pages 304–324

http://csli-publications.stanford.edu/LFG/2018

Keywords: Polish, predicative constructions

Patejuk, Agnieszka, & Przepiórkowski, Adam. 2018. Predicative Constructions with Infinitival and Clausal Subjects in Polish. In Butt, Miriam, & King, Tracy Holloway (Eds.), *Proceedings of the LFG'18 Conference, University of Vienna*, 304–324. Stanford, CA: CSLI Publications.

Abstract

This paper offers a formal syntactic analysis of predicative constructions in Polish (with and without a copula), where the item predicated of is a clause: a complementiser phrase or an infinitival phrase. The predicative item may be a noun, an adjective or, a possibility previously denied, an adverb.

1 Introduction

This paper offers a formal syntactic analysis of predicative copular constructions in Polish, where the item predicated of (the subject of predication) is a clause – a complementiser phrase (CP, see (1)) or an infinitival phrase (InfP, as in (2)). This paper takes into account uncontroversial predicative items such as nominals (§2.1) and adjectives (§2.2; see (1)–(2)),[1] but also adverbs (§2.3), which is a novel contribution, challenging the widely accepted assumption that adverbs cannot be predicative.

(1) [Że musi ich być na to stać] wydaje się być
that must.3.SG them.ACC be.INF on this afford.INF seem.3.SG RM be.INF
oczywiste.
obvious.NOM.SG.N
'It seems to be obvious that they must be able to afford it.' (NKJP)

(2) Ciekawe jest [odpowiadać na znane sobie pytania]?
interesting.NOM.SG.N is.3.SG answer.INF on known self.DAT questions
'Is answering questions known to oneself interesting?' (NKJP)

There have been discussions of non-canonical subjects (such as clauses in Polish, Dziwirek 1990; see (3)–(4) with a CP, from Patejuk 2015) and of predicative complements (including non-canonical cases, where the predicative item is a CP, a gerund or an infinitival clause, see (5)–(7) from Dalrymple et al. 2004).

(3) Naszych gości dziwiło, [że mamy tak dużo obowiązków].
our guests.ACC puzzled.3.SG.N that have.1.PL so many.ACC duties.GEN
'(The fact) that we have so many duties puzzled our guests.' (NKJP)

(4) Cieszyło ją, [że mam tak oryginalne zainteresowania].
made.happy.3.SG.N she.ACC that have.1.SG so original interests.ACC
'(The fact) that I have so original interests made her happy.' (NKJP)

(5) The problem is that they appear. (Dalrymple et al. 2004: ex. (1d))

[†] We are very grateful for valuable comments provided by anonymous reviewers and Mary Dalrymple. Unfortunately, due to limited space, not all of these comments could be taken into account. The research reported here is partially supported by the Polish Ministry of Science and Higher Education (MNiSW) within the CLARIN ERIC programme 2016–2018 (http://clarin.eu/). This work was completed during the Mobilność Plus mobility grant awarded to Agnieszka Patejuk by MNiSW.

[1] 'NKJP' and 'Google' mark attested examples taken from the National Corpus of Polish (NKJP, http://nkjp.pl, Przepiórkowski et al. 2012) and the Internet.

(6) The problem is their appearing. (Dalrymple et al. 2004: ex. (1e))

(7) The problem is to leave before 6:00. (Dalrymple et al. 2004: ex. (1f))

However, these issues have only been discussed separately so far – to the best of our knowledge, no previous LFG work focuses on predicative complements whose subject (or controller – the item it predicates of) is a CP or an infinitival clause.

This is the case despite the fact that Polish copular constructions have been investigated earlier (e.g. Citko 2004, Bondaruk 2013): while Citko 2004 does not mention such examples at all, Bondaruk 2013 does present data in (8), but does not include such examples in her analysis of Polish copular clauses.

(8) [Uratować się] było nieprawdopodobieństwem.
save.INF REFL was.3.SG.N improbability.INST.SG.N
'To be saved was improbable.' (Bondaruk 2013: ex. (8) after Kallas 1993)

Interestingly, examples in (9)–(10), which perfectly match the topic of this paper, were used by Dziwirek 1990 to show Polish default subject-verb agreement triggered by the clausal non-canonical subject (infinitive and CP, respectively). However, Dziwirek 1990 did not discuss predicative complements in these examples, focusing instead exclusively on default agreement between the subject and the verb.

(9) [Mówić prawdę] było twoim obowiązkiem.
speak.INF truth.ACC was.3.SG.N your.INST.SG.M duty.INST.SG.M
'To speak the truth was your duty.' (Dziwirek 1990: ex. (17a))

(10) [Że Janek kochał Ewę] było dla wszystkich oczywiste.
that Janek loved Ewa.ACC was.3.SG.N for all obvious.NOM.SG.N
'That Janek loved Ewa was obious to everyone.' (Dziwirek 1990: ex. (17b))

The current paper discusses default agreement on the predicative adjective (singular number, neuter gender in (10)) triggered by the non-canonical subject such as an infinitive or a complementiser clause, which has not been noticed earlier. So far default agreement in Polish has only been discussed in the context of subject-verb agreement (Dziwirek 1990), where it is also triggered by non-canonical subjects.

Another novel claim made in this paper is that there exist predicative adverbs in Polish examples such as (11), where the subject is non-canonical (clausal), though neither Bondaruk 2013 nor Kallas 1993 analyse this example in this way:

(11) [Uratować się] było trudno.
save.INF REFL was.3.SG.N difficult.ADV
'To be saved was difficult.' (Bondaruk 2013: ex. (7) after Kallas 1993)

This proposal contradicts claims that adverbs cannot be predicative, e.g. Rothstein 2001: 129 on English: "I assume that the absence of a predication relation is because adverbs are just not syntactic predicates. They never appear in a position in which they can be predicated of events, since even if the argument denotes an event, it cannot have an adverb predicated of it. The examples in [(12)] are all unacceptable with adverbial predicates, though the corresponding adjectives are all OK."

(12) a. The destruction of the city was *brutally/brutal.
b. The reading of the verdict was *slowly/slow.
c. John considered [the running *slowly/slow].

While this claim might be true for English, it will be shown that this is not the case in Polish, where an unambiguously adverbial[2] predicative complement may be used when the subject is clausal: an infinitive or a CP.

Finally, this paper discusses the issue of control into infinitival subjects by the optional dative dependent of a predicative adverb, as shown in (13):

(13) [Uratować się] było nam trudno.
save.INF REFL was.3.SG.N us.DAT difficult.ADV
'To be saved was difficult for us.' (Bondaruk 2013: ex. (10))

Bondaruk and Szymanek 2007 discuss examples analogous to (13) – they consider (14), asking the question whether the infinitival phrase *przegrywać mecze* 'lose matches' is the "extraposed subject clause" or "a complement clause". They decide that the subject hypothesis is "implausible in the light of the fact that extraction out of non-finite clauses like [(14)] is possible" – it breaches the Constraint of Extraction Domains proposed by Huang 1982, which bans extraction out of subjects. Bondaruk and Szymanek 2007 provide (15) as supporting evidence, where *Co* 'what' is extracted from the infinitival phrase headed by *przegrywać* 'lose':

(14) Jest mu smutno [przegrywać mecze].
is.3.SG he.DAT sad.ADV lose.INF matches.ACC
'It is sad for him to lose matches.' (Bondaruk and Szymanek 2007: ex. (24))

(15) Co jest mu smutno przegrywać?
what.ACC is.3.SG he.DAT sad.ADV lose.INF
'What is it sad for him to lose?' (Bondaruk and Szymanek 2007: ex. (25))

However, it can be shown that it is not true that extraction from Polish subjects is banned, making the argument against the subjecthood of the infinitive in (14) void:

(16) Czyje przyszły dzisiaj zakupy?
whose.NOM.PL.M arrived.3.PL.M today shopping.NOM.PL.M
'Whose shopping order arrived today?'

Moreover, it is increasingly clear that apparently syntactic island constraints are largely a matter of cognition and processing (see, e.g., Hofmeister and Sag 2010 and further work by Hofmeister and colleagues).

2 Data and analysis

Before proceeding to the discussion of data, it is worth noting that constructions discussed in this paper – predicative complements whose subject is clausal (CP, infinitival phrase) – are very frequent, commonly seen in various genres, including carefully edited texts, which shows that this phenomenon is systematic in Polish.

[2] In Polish adverbs are morphologically distinct from adjectives, their distribution is also different.

In many examples below the predicative complement is placed sentence initially. However, the word order in Polish is free – the tendency for placing the clausal subject sentence-finally is attributable to its relative constituent weight: it is heavy. If the predicative complement is made sufficiently heavy, or the subject sufficiently light, it becomes possible and natural to switch the word order, putting the clausal subject first and the predicative complement last.

2.1 Predicative nominals

Let us start with the discussion of predicative nominals, which are perhaps the least controversial. In Polish, the predicative nominal is typically marked for instrumental case and it neither has to agree with its subject in number nor in gender, as shown in (17), where the subject *rolnicy* 'farmers' is plural masculine, while the predicative complement *ofiarą* 'victim' is singular feminine.

(17) Rolnicy są ofiarą systemu.
farmer.NOM.PL.M are.3.PL victim.INST.SG.F system.GEN.SG.M
'Farmers are the victim of the system.' (Google)

There are numerous examples with an instrumental predicative NP whose subject is non-canonical. Let us start with examples with a CP subject:

(18) Moją pierwszą myślą było, [że nie
my.INST.SG.F first.INST.SG.F thought.INST.SG.F was.3.SG.N that NEG
powinienem tego podpisać].
should.1.SG.M this.GEN sign.INF
'My first thought was that I should not sign this.' (NKJP)

(19) Ciekawostką jest, [że w akumulatorach [...] jako paliwo
interesting fact.INST.SG.F is.3.SG that in accumulators as fuel
używany będzie alkohol].
used will.3.SG alcohol.NOM
'An interesting fact is that alcohol will be used as fuel in accumulators.' (NKJP)

There are also multiple examples where the subject of the instrumental predicative NP is an infinitival phrase (earlier examples include (8) and (9)):

(20) Grzechem jest [oglądać ten film].
sin.INST.SG.F is.3.SG watch.INF this.ACC film.ACC
'To watch this film is a sin.' (NKJP)

(21) Najwyższą karą było - [zostać
highest.INST.SG.F punishment.INST.SG.F was.3.SG.N - become.INF
usuniętym ze strajku].
removed.INST.SG.M from strike
'The worst punishment was to be removed from the strike.' (NKJP)

Though it is not a decisive argument, non-canonical subjects can in principle be substituted for a gerund, which can uncontroversially act as the subject – compare the following two examples to (19)–(20) above:

(22) Ciekawostką jest [używanie w akumulatorach
interesting fact.INST.SG.F is.3.SG using.NOM.SG.N in accumulators
alkoholu jako paliwo].
alcohol.GEN as fuel
'Using alcohol as fuel in accumulators is an interesting fact.'

(23) Grzechem jest [oglądanie tego filmu].
sin.INST.SG.F is.3.SG watching.NOM.SG.N this.GEN film.GEN
'Watching this film is a sin.'

Furthermore, predicative nominals taking a clausal subject also occur with predicates other than the copula, including WYDAWAĆ SIĘ 'seem'. Assuming that the instrumental dependent is predicative (as in cases with a nominal subject), the clause should be analysed as its subject to avoid stipulating special valency requirements.

(24) Najbardziej rozsądnym przypuszczeniem wydaje się, [że
most reasonable.INST.SG.N presumption.INST.SG.N seem.3.SG REFL that
miał zostać [...] zabity [...]].
had.3.SG.M become.INF killed.NOM
'The most reasonable presumption seems to be that he was to be killed.' (NKJP)

(25) Wobec takiej perspektywy ekonomicznie lepszą opcją
against such perspective economically better.INST.SG.F option.INST.SG.F
wydaje się [[unikać legalnego zatrudnienia], [pracować na
seem.3.SG REFL avoid.INF legal.GEN employment.GEN work.INF on
czarno], [pobierać pomoc] i [później liczyć na umorzenie długu]].
black take.INF support.ACC and later count.INF on remission debt
'From this perspective, the economically better solution seems to be to avoid legal employment, work illegally, claim benefits and later count on (hope for) debt remission.' (NKJP)

2.2 Predicative adjectives

This section presents examples with predicative adjectives whose subject is clausal – earlier examples include (2), where the subject is an infinitival phrase, and (1), where the subject is a CP. The latter also includes raising, which, as explained earlier, only occurs with the subject in Polish. All examples presented in this section feature the raising verb WYDAWAĆ SIĘ 'seem'. The adjective can be specified for any degree (positive, comparative, superlative) – synthetically or analytically (with the help of the adverb BARDZO 'very' in an appropriate degree).

Unlike predicative nominals, predicative adjectives agree with their nominal subject in number and gender:

(26) Facet był miły.
guy.NOM.SG.M was.3.SG.M kind.NOM.SG.M
'The guy was kind.'

(27) Kobieta była miła.
woman.NOM.SG.F was.3.SG.F kind.NOM.SG.F
'The woman was kind.'

(28) (Trzy) kobiety były miłe.
three.NOM.PL.F woman.NOM.PL.F were.3.PL.F kind.NOM.PL.F
'(Three) women were kind.'

As shown in (29), this also applies to numeral subjects which require a nominal in genitive case (unlike in (28)). Because the numeral subject in (29) is not nominative (but accusative: Przepiórkowski 1999), it triggers default agreement on the verb (third person, singular, neuter). Still, full agreement (in number, gender and case) is required between the subject and the predicative complement – it can either agree in case with the accusative numeral head, or with the genitive nominal.

(29) Pięć kobiet było miłe/miłych.
five.ACC.PL.F woman.GEN.PL.F was.3.SG.N kind.ACC/GEN.PL.F
'Five women were kind.'

However, when the subject of a predicative adjective is clausal (a CP or an infinitive, see (1)–(2) and examples in (30)–(35) below), it triggers default agreement both on the verb and on the predicative adjective – the latter must be marked for singular number and neuter gender. This can be observed in all relevant examples presented in this paper (see e.g. (44)–(45), (54)–(57) discussed later).

Finally, while predicative nominals discussed in §2.1 consistently appear in instrumental case (the so called instrumental of predication), predicative adjectives with a clausal subject are typically nominative (as in most examples), but not always – an example with instrumental case is provided in (35). While this option seems to be generally available as an alternative (such examples are grammatical), the nominative is clearly more frequent and seems to be the default agreement case marking on predicative adjectives with a clausal subject (a CP or an infinitival phrase).

Examples with a CP subject are given in (30)–(32) – (30) features synthetic comparative degree, while (31) shows analytic comparative degree:

(30) Dziwniejsze wydawało się, [że M.M. miała
strange.NOM.SG.N.COMPAR seemed.3.SG.N REFL that M.M.NOM had.3.SG.F
[...] większą sprzedaż]!
higher.ACC sale.ACC
'It seemed more strange that M.M. had higher sales.' (NKJP)

(31) Bardziej prawdopodobne wydawało się, [że ktoś jej pomógł].
more likely.NOM.SG.N seemed.3.SG.N REFL that sb.NOM she.DAT helped
'It seemed more likely that somebody helped her.' (NKJP)

(32) Wdowie logiczne wydawało się, [że nabyła
widow.DAT logical.NOM.SG.N seemed.3.SG.N REFL that gained.3.SG.F
prawo do przejęcia mieszkania po zmarłym mężu].
right.ACC to take over flat after deceased husband
'It seemed logical to the widow that she gained the right to take over her husband's flat.' (NKJP)

Examples with an infinitival subject are provided in (33)–(35):

(33) Rozsądne wydawało się więc [poczekać].
sensible.NOM.SG.N seemed.3.SG.N REFL so wait.INF
'So it seemed sensible to wait.' (NKJP)

(34) Nierealne wydaje się [spodziewać takich działań].
unrealistic.NOM.SG.N seem.3.SG REFL expect.INF such.GEN actions.GEN
'It seems unrealistic to expect such actions.' (NKJP)

(35) Naturalnym wydaje się [[...] spróbować uratować tę drogę].
natural.INST.SG.N seem.3.SG REFL try.INF save.INF this road.ACC
'It seems natural to try to save this road.' (NKJP)

Similarly as with predicative nominals, it is possible to construct examples where the non-canonical clausal subject is substituted with a gerund, which makes an uncontroversial nominal subject: *odpowiadanie* 'answering' in (36) corresponds to the infinitival subject (*odpowiadać* 'answer') in (2), while *nabycie* 'gaining' in (37) corresponds to the CP subject (*że nabyła...* 'that gained') in (32).

(36) Ciekawe jest [odpowiadanie na pytania]?
interesting.NOM.SG.N is.3.SG answering.NOM.SG.N on questions
'Is answering questions interesting?'

(37) Wdowie logiczne wydawało się [nabycie prawa].
widow.DAT logical.NOM.SG.N seemed.3.SG.N REFL gaining.NOM.SG.N right
'Gaining the right seemed logical to the widow.'

2.3 Predicative adverbs

Counter to claims that adverbs are never predicates (e.g., Rothstein 2001: 129 cited in §1), in Polish an adverb in any degree (positive, comparative, superlative) may serve as the predicative complement (in principle, subject to semantic constraints).

As mentioned earlier, Polish adverbs are morphologically distinct from adjectives – adjectives are marked for features such as case, number or gender, these are not applicable to adverbs. Hence adverbs and adjectives have different distribution.

Examples in (38)–(39) contain predicative adverbs whose subject is a CP:

(38) Dobrze jest, [że czują pewien respekt].
good.ADV is.3.SG that feel.3.PL some.ACC respect.ACC
'It is good that they feel some respect.' (NKJP)

(39) Maciusiowi bardzo przyjemnie było, [że królewski poseł nie
Maciuś.DAT very pleasant.ADV was.3.SG.N that royal.NOM envoy.NOM NEG
mówił w zagranicznym języku].
spoke.3.SG.M in foreign language
'That the royal envoy did not speak in foreign language was very pleasant to Maciuś.' (NKJP)

Examples in (40)–(41) contain predicative adverbs whose subject is an infinitival phrase. Additionally, (40) involves coordination of predicative adverbs.

(40) Najłatwiej i najtaniej było [upłynnić ziarno czy ziemniaki].
easy.ADV.SUP and cheap.ADV.SUP was.3.SG.N sell.INF grain or potatoes
'It was easiest and cheapest to sell grain or potatoes.' (NKJP)

(41) Oczywiście autorowi najtrudniej było [uzyskać szczegóły].
obviously author.DAT difficult.ADV.SUP was.3.SG.N get.INF details.ACC
'Obviously, to get the details was the most difficult for the author.' (NKJP)

Unlike in the case of predicative nominals (§2.1) and adjectives (§2.2), it is not possible to construct examples with predicative adverbs whose subject is a gerund.

2.4 Without a copula

The copula in constructions involving a clausal subject is optional and it may be dropped – the predicative item serves then as the main predicate, just as in standard copular constructions in Polish, involving a nominal subject, see (42)–(43):

(42) Zaczęło mu odbijać, bo *dzień* taki *długi*.
started he.DAT freak out.INF because day.NOM.SG.M so long.NOM.SG.M
'He started to freak out, because the day is so long.' (NKJP)

(43) Powiedz to całemu piekłu, bo moje *słowa* nie *wiatr*!
tell.2.SG this entire hell because my words.NOM NEG wind.NOM
'Tell this to all the hell, because my words are not wind (= are not to be ignored)!' (NKJP)

Both examples above include a subordinate clause with the complementiser *bo* 'because', where the predicative item serves as the main predicate in the absence of the copula. In (42) it is the predicative adjective *długi* 'long', whose subject is *dzień* 'day', while in (43) it is the predicative noun *wiatr* 'wind' (modified using negation), whose subject is *słowa* 'words'.

Examples below feature predicative adjectives (both superlative) controlled by an infinitival phrase (*być dobrym premierem* 'to be a good PM') in (44) or by a CP (*że dojedzie się do celu* 'that one will reach the destination') in (45). As explained in §2.2, predicative adjectives controlled by clausal subjects are in the default agreement form: singular neuter. It seems, however, that when the predicative adjective is the main predicate (there is no copula), it must be marked for nominative case – by contrast, when the copula is present, the instrumental case is rare, but possible.

(44) Najważniejsze [być dobrym premierem].
important.NOM.SG.N.SUP be.INF good.INST PM.INST
'To be a good PM [is] the most important.' (Google)

(45) Najważniejsze, [że dojedzie się do celu].
important.NOM.SG.N.SUP that reach.3.SG REFL to destination
'That one will reach the destination [is] the most important.' (NKJP)

Examples below show predicative adverbs (superlative or positive) controlled by a clausal subject: an infinitival phrase (*pogodzić się z tym* 'to come to terms with this') in (46) or by a CP (*że nie udało się uratować sosen* 'that we did not manage to save the pines') in (47).

(46) Najtrudniej [pogodzić się z tym] ludziom młodym.
difficult.ADV.SUP reconcile.INF REFL with this people.DAT young.DAT
'To come to terms with this [is] most difficult for young people.' (NKJP)

(47) Przykro, [że nie udało się uratować sosen].
sad.ADV that NEG managed.3.SG.N REFL save.INF pines.GEN
'That we did not manage to save the pines [is] sad.' (NKJP)

Finally, it seems that it is also possible (though perhaps only marginally – such constructions are not very frequent with nominal controllers in Polish) to use predicative nominals without a copula with a clausal controller – in (48) it is a CP (*że pielęgniarki zarabiają tak mało* 'that nurses earn so little'), while in (49) it is an infinitival phrase (*olewać własny naród* 'to disregard your own nation'):

(48) Absolutny skandal, że pielęgniarki tak mało zarabiają.
absolute.NOM.SG.M scandal.NOM.SG.M that nurses.NOM so little earn.3.PL
'It is an absolute scandal that nurses earn so little.' (Google)

(49) Skandal - [tak olewać własny naród].
scandal.NOM.SG.M so disregard.INF own.ACC nation.ACC
'It is a scandal to disregard your own nation in this way.' (NKJP)

2.5 With verbs taking an object-controlled predicative complement

This section presents examples with predicates UWAŻAĆ 'consider (imperfective)' and UZNAĆ 'consider (perfective), acknowledge', which take a predicative complement (inside a PP) which is not controlled by a subject, but by a different dependent – an object (it can passivise, as discussed in (58)–(59)). This amply shows that clauses may serve as controllers of predicative complements in Polish, which means that they should also be able to be the subject of sentences with a copula.

Examples with a predicative nominal controlled by a clausal object:

(50) Naprawdę uważasz za regułę, [że ludzi dotykają
really consider.2.SG as rule.ACC.SG.F that people.ACC touch.3.PL
choroby dlatego, że ich nie unikają]?
diseases.NOM because that they.GEN NEG avoid.3.PL

‘Do you really consider it a rule that diseases affect people because they do not avoid them?’ (NKJP)

(51) Uznał za sukces, [że władze zgodziły
considered.3.SG.M as success.ACC.SG.M that authorities.NOM agreed.3.PL
się na rozmowę].
REFL for talk
‘He considered it a success that the authorities agreed for a talk.’ (NKJP)

Examples with a predicative nominal controlled by an infinitival complement:

(52) Koledzy z klubu uważali za dyshonor
colleagues.NOM from club considered.3.PL.M as dishonour.ACC.SG.M
[przegrywać z kobietą].
lose.INF with woman
‘Colleagues from the club considered it a dishonour to lose against a woman.’ (NKJP)

(53) Lekarz poczuł się winny i uznał za obowiązek
doctor felt.3.SG.M REFL guilty and considered.3.SG.M as duty.ACC.SG.M
[przeprosić pacjentkę].
apologise.INF patient.ACC
‘The doctor felt guilty and he considered it his duty to apologise to the patient.’ (NKJP)

Examples with a predicative adjective controlled by a clausal complement:

(54) Uważam za prawdopodobne, [że wirus ten może się
consider.1.SG as probable.ACC.SG.N that virus this may REFL
rozprzestrzenić].
spread.INF
‘I consider it probable that this virus may spread.’ (NKJP)

(55) Jeżeli uznasz za prawdziwe, [że masz rękę i oko].
if consider.2.SG as true.ACC.SG.N that have.2.SG hand.ACC and eye.ACC
‘If you consider it true that you have a hand and an eye.’ (NKJP)

Examples with a predicative adjective controlled by an infinitival complement:

(56) Uważał za stosowne [zademonstrować swój sprzeciw].
considered.3.SG.M as suitable.ACC.SG.N demonstrate.INF his objection.ACC
‘He considered it appropriate to express his objection.’ (NKJP)

(57) Uznała za słuszne [skierować środki na likwidowanie
considered.3.SG.F as fair.ACC.SG.N direct.INF assets.ACC for removing
negatywnych skutków].
negative effects
‘She considered it fair to use assets to remove negative effects.’ (NKJP)

As mentioned earlier, it seems that examples with the predicative complement of UWAŻAĆ and UZNAĆ may be passivised, which results in the non-canonical object controller becoming the subject – the constructed passive examples in (58) and (59) roughly correspond to active examples in (50) and (54), respectively.

(58) Przez wielu jest uważane za regułę, [że ludzi
by many is.3.SG considered.NOM.SG.N as rule.ACC.SG.F that people.ACC
dotykają choroby dlatego, że ich nie unikają].
touch.3.PL diseases.NOM because that they.GEN NEG avoid.3.PL
'It is considered a rule by many that diseases affect people because they do not avoid them.'

(59) Jest uważane przez ekspertów za prawdopodobne, [że
is.3.SG considered.NOM.SG.N by experts as probable.ACC.SG.N that
wirus ten może się rozprzestrzenić].
virus this may REFL spread.INF
'It is considered probable by experts that this virus may spread.'

Finally, because the predicative complement of verbs discussed above is a prepositional phrase, where the preposition ZA assigns accusative case to its dependent, it is not surprising that there are no such examples with a predicative adverb – adverbs do not have case.

2.6 Coordination

Let us consider (60) – it provides additional evidence from coordination which shows that the clausal dependent of UWAŻAĆ in examples in §2.5 has the same grammatical function as the corresponding nominal dependent (object in active voice):

(60) Uważając [[za wariactwo] [dać się złapać hitlerowcom]],
considering as madness.ACC.SG.N give.INF REFL catch.INF Nazis.DAT
a [[samobójstwo] [za tchórzostwo]].
and suicide.ACC.SG.N as cowardice.ACC.SG.N
'Considering [[letting Nazis catch you] [madness]], and [[suicide] [cowardice]].' (NKJP)

(60) involves non-constituent coordination (Maxwell and Manning 1996) – two pairs of dependents of the predicate UWAŻAĆ 'consider' are coordinated, with each pair containing a predicative PP and its object controller. The first conjunct consists of the predicative PP *za wariactwo* 'as madness' and its infinitival phrase controller (*dać się złapać hitlerowcom* 'let Nazis catch you'). In the second conjunct the order is reversed: it contains the (structural) accusative nominal controller (*samobójstwo* 'suicide') followed by the predicative PP *za tchórzostwo* 'as cowardice'.

2.7 Control

The basic syntactic analysis of examples involving a predicative adverb controlled by an infinitival phrase or a CP is complicated by the fact that adverbial predicates

may occur with a dative experiencer acting as the controller of the subject of the infinitival phrase: see (41) with a copula and (46) without a copula, repeated below for convenience as (61) and (62).

(61) Oczywiście autorowi najtrudniej było [uzyskać szczegóły].
obviously author.DAT difficult.ADV.SUP was.3.SG.N get.INF details.ACC
'Obviously, to get the details was the most difficult for the author.' (NKJP)

(62) Najtrudniej [pogodzić się z tym] ludziom młodym.
difficult.ADV.SUP reconcile.INF REFL with this people.DAT young.DAT
'To come to terms with this [is] most difficult for young people.' (NKJP)

It is worth noting that adverbial predicates differ in their propensity to occur with dative experiencers – for example, CIEKAWIE 'interesting.ADV' combines with dative experiencers perhaps marginally, but see the attested (63):

(63) Jako że mieszkam nad samą Odrą, ciekawie mi było [o
as that live.1.SG over self Oder interesting.ADV I.DAT was.3.SG.N about
niej poczytać] ;)
her read.INF
'As I live by the Oder, it was interesting for me to read about it.' (Google)

Since the dative experiencer acts as a controller of the subject of the subject infinitival phrase in (63), it is an instance of control into a subject.

While rare, control into the subject was discussed in Arka and Simpson 1998 for Balinese, where it occurs under objective voice, as shown in (64):

(64) [teka mai prajani] ane orahin tiang Nyoman
come here immediately REL OV.ask 1 Nyoman
'Coming here immediately is what I asked Nyoman to do.'
(Arka and Simpson 1998: ex. (17))

The main verb in this example, *orahin* 'ask', is in objective voice (marked in glosses as OV). Its subject is the clause *teka mai prajani* 'come here immediately' – the head of this clause, *teka* 'come', is obligatorily controlled by the object of the main verb, *Nyoman*. The relevant functional control equation is (↑ SUBJ SUBJ)=(↑ OBJ).

Polish provides independent evidence supporting the need for control into the subject – it also occurs with the verb UDAĆ SIĘ 'manage', which can either take a nominal subject (it agrees with the verb, so the dative dependent is not the subject), as in (65), or an infinitival subject, as in (66) (as well as in (47)).

(65) Nie udał im się [rozruch ciągnika].
NEG manage.3.SG.M they.DAT REFL start.NOM.SG.M tractor.GEN
'They did not succeed in starting the tractor.'

(66) Nie udało im się [uruchomić ciągnika].
NEG manage.3.SG.N they.DAT REFL start.INF tractor.GEN
'They did not manage to start the tractor.' (NKJP)

The control equation needed in examples with an infinitival subject such as (47) and (66) is provided in the lexical entry in (67).[3]

(67) (↑ PRED)='MANAGE<(↑ SUBJ)(↑ OBJ$_\theta$)>'
(↑ OBJ$_\theta$)=(↑ SUBJ SUBJ)

This equation in not used in examples such as (65) – the subject in (65) does not have a controlled subject (its subject is closed), so using the equation in (67) would result in an incoherent f-structure.

3 Formalisation

Though the data is rich and complex, the formalisation of the proposed analysis is simple and it was verified in an XLE (Crouch et al. 2011) implementation.

It requires minimal modifications of c-structure rules: (68) allows CP/InfP to be a subject or an object (accounting for object-controlled predicative complements, see §2.5), while (69) allows AdvP to be a predicative complement. When DEP is attached as a dependent of a predicate, it bears the co-head annotation (↑=↓).

(68) DEP → { CP | InfP }
↓ = (↑ {SUBJ|OBJ})

(69) DEP → AdvP
↓ = (↑ XCOMP-PRED)

The functional annotation of the predicative complement in (69) uses the XCOMP-PRED grammatical function – this is because this paper assumes that predicative complements are open and so they require a subject which is the item predicated of, as opposed to the closed PREDLINK analysis of predicative complements (Butt et al. 1999), where the item predicated of is not structure-shared with the subject of the predicative item. However, the current analysis could easily be converted into a closed PREDLINK analysis – the CONTROLLER attribute (Patejuk and Przepiórkowski 2014) could be introduced to store the controller locally to the predicative item.

The lexicon is modified by introducing predicative entries for adverbs, which also allow for dative experiencers (OBJ$_\theta$) – see the schematic lexical entry in (70). Its first disjunct (INF) is used in (13), (41), (46) and (63), while the second disjunct (COMP-FORM) is used in (38), (39) and (47). There is no need to add predicative entries for nominals and adjectives, since they already exist in the grammar.

(70) (↑ PRED)='LEMMA<(↑ SUBJ)(↑ OBJ$_\theta$)>'
[[(↑ SUBJ CAT)=$_c$ INF ∧ (↑ OBJ$_\theta$)=(↑ SUBJ SUBJ)] ∨ (↑ SUBJ COMP-FORM)]
(↑ OBJ$_\theta$ CASE)=$_c$ DAT

Note that, as explained in §2.4, predicative items may also act as the main predicate in the absence of the copula. Since predicative nominals and adjectives are normally allowed to function as the main predicate, the relevant rule is present in the grammar – with NP and AP disjuncts. This rule need only be modified by adding the AdvP disjunct – see (71). The functional annotation of the disjunctive right hand side contains the head annotation (↓ = ↑), which contributes the predicate introduced by the predicative item as the main predicate. It also ensures that the relevant element is predicative using an existential equation (cf. (↓ PREDICATIVE)).

[3] While (67) uses functional control, further research is needed to determine suitable control type.

(71) S → { NP | AP | AdvP }
↓ = ↑
(↓ PREDICATIVE)

The lexical entry of the copula, (72), is standard – it structure-shares its subject with the subject of the open predicative complement (XCOMP-PRED):

(72) (↑ PRED)='BE<(↑ XCOMP-PRED)>(↑ SUBJ)'
(↑ SUBJ)=(↑ XCOMP-PRED SUBJ)

As shown in (70), lexical entries of predicative adverbs contain the control equation stating that the dative experiencer is the controller of the infinitival subject of the predicative adverb ((↑ OBJ$_\theta$)=(↑ SUBJ SUBJ)). As a result, there is no need to create a special lexical entry for the copula with a dative experiencer – the copula has one standard two-argument lexical entry shown in (72). This yields a unified analysis of control with predicative adverbs: it is the same regardless of whether the copula is present (as in (41)) or not (see (46)) – in both cases the dative experiencer is a dependent of the predicative adverb, resulting in a fully consistent representation.

Finally, agreement with predicative adjectives is handled using templates called inside the lexical entry of the adjective – see (73). It imposes appropriate agreement with clausal agreement targets – a CP has the attribute COMP-FORM hosting the complementiser, while an infinitive has the attribute CAT with value INF. It ensures that the adjective is in the default agreement form (singular number, neuter gender) and it allows it to be marked for nominative or instrumental case.[4]

(73) PRIPRED-ADJ-CLAUSAL ≡
[(↓ SUBJ CAT)=$_c$ INF ∨ (↓ SUBJ COMP-FORM)]
(↓ CASE)∈$_c$ {NOM, INST} (↓ NUM)=$_c$ SG (↓ GEND)=$_c$ N

The following subsections present f-structure representations (sometimes simplified; XC-P=XCOMP-PRED) resulting from the formalisation presented above. Due to space limits, it is only possible to present f-structures of selected examples.

3.1 Nominals

(74) corresponds to (20), where the predicative nominal is controlled by an InfP:

(74)
$$\begin{bmatrix} \text{PRED} & \text{'BE<}\boxed{2}\text{>}\boxed{1}\text{'} & \\ \text{SUBJ} & \boxed{1}\begin{bmatrix} \text{PRED} & \text{'WATCH<}\boxed{3}\text{,}\boxed{4}\text{>'} & \\ \text{SUBJ} & \boxed{3}\begin{bmatrix}\text{PRED} & \text{'PRO'}\end{bmatrix} & \\ \text{OBJ} & \boxed{4}\begin{bmatrix} \text{PRED} & \text{'FILM'} & \\ \text{ADJ} & \left\{\begin{bmatrix}\text{PRED} & \text{'THIS'}\end{bmatrix}\right\} & \end{bmatrix} & \end{bmatrix} & \\ \text{XC-P} & \boxed{2}\begin{bmatrix} \text{PRED} & \text{'SIN<}\boxed{1}\text{>'} & \\ \text{SUBJ} & \boxed{1} & \\ \text{CASE} & \text{INST} & \end{bmatrix} & \end{bmatrix}$$

[4] OT marks can be used to formalise the observation that nominative is clearly preferred as the default agreement case with clauses, while instrumental is a rare and rather marked alternative.

(75) corresponds to (18), where the predicative nominal is controlled by a CP:

(75)
$$\begin{bmatrix}
\textsc{pred} & \text{`BE<}\boxed{2}\text{>}\boxed{1}\text{'} \\
\textsc{subj} & \boxed{1}\begin{bmatrix}
\textsc{pred} & \text{`SHOULD<}\boxed{3}\text{,}\boxed{4}\text{>'} \\
\textsc{subj} & \boxed{3}\begin{bmatrix}\textsc{pred} & \text{`PRO'}\end{bmatrix} \\
\textsc{xcomp} & \boxed{4}\begin{bmatrix}
\textsc{pred} & \text{`SIGN<}\boxed{3}\text{,}\boxed{5}\text{>'} \\
\textsc{subj} & \boxed{3} \\
\textsc{obj} & \boxed{5}\begin{bmatrix}\textsc{pred} & \text{`THIS'}\end{bmatrix}
\end{bmatrix} \\
\textsc{comp-form} & \textsc{that} \quad \textsc{neg} +
\end{bmatrix} \\
\textsc{xc-p} & \boxed{2}\begin{bmatrix}
\textsc{pred} & \text{`THOUGHT<}\boxed{1}\text{>'} \\
\textsc{subj} & \boxed{1} \\
\textsc{adj} & \left\{\begin{bmatrix}\textsc{pred} & \text{`MY'}\end{bmatrix}, \begin{bmatrix}\textsc{pred} & \text{`FIRST'}\end{bmatrix}\right\} \\
\textsc{case} & \textsc{inst}
\end{bmatrix}
\end{bmatrix}$$

3.2 Adjectives

(76) corresponds to (2), where the predicative adjective is controlled by an InfP:

(76)
$$\begin{bmatrix}
\textsc{pred} & \text{`BE<}\boxed{2}\text{>}\boxed{1}\text{'} \\
\textsc{subj} & \boxed{1}\begin{bmatrix}
\textsc{pred} & \text{`ANSWER<}\boxed{3}\text{,}\boxed{4}\text{>'} \\
\textsc{subj} & \boxed{3}\begin{bmatrix}\textsc{pred} & \text{`PRO'}\end{bmatrix} \\
\textsc{obl} & \boxed{4}\begin{bmatrix}
\textsc{pred} & \text{`QUESTION'} \\
\textsc{adj} & \left\{\begin{bmatrix}\textsc{pred} & \text{`KNOWN'}\end{bmatrix}\right\}
\end{bmatrix}
\end{bmatrix} \\
\textsc{xc-p} & \boxed{2}\begin{bmatrix}
\textsc{pred} & \text{`INTERESTING<}\boxed{1}\text{>'} \\
\textsc{subj} & \boxed{1} \\
\textsc{case} & \textsc{nom} \quad \textsc{num} \ \textsc{sg} \quad \textsc{gend} \ \textsc{n}
\end{bmatrix}
\end{bmatrix}$$

(77) corresponds to (1), where the predicative adjective is controlled by a CP; it also includes a raising verb (WYDAWAĆ SIĘ 'seem'):

(77)
$$\begin{bmatrix}
\textsc{pred} & \text{`SEEM<}\boxed{2}\text{>}\boxed{1}\text{'} \\
\textsc{subj} & \boxed{1}\begin{bmatrix}
\textsc{pred} & \text{`MUST<}\boxed{3}\text{>'} \\
\textsc{xcomp} & \boxed{3}\begin{bmatrix}
\textsc{pred} & \text{`AFFORD<}\boxed{4}\text{,}\boxed{5}\text{>'} \\
\textsc{obj} & \boxed{4}\begin{bmatrix}\textsc{pred} & \text{`THEY'}\end{bmatrix} \\
\textsc{obl} & \boxed{5}\begin{bmatrix}\textsc{pred} & \text{`THIS'}\end{bmatrix}
\end{bmatrix} \\
\textsc{comp-form} & \textsc{that}
\end{bmatrix} \\
\textsc{xc-p} & \boxed{2}\begin{bmatrix}
\textsc{pred} & \text{`OBVIOUS<}\boxed{1}\text{>'} \\
\textsc{subj} & \boxed{1} \\
\textsc{case} & \textsc{nom} \quad \textsc{num} \ \textsc{sg} \quad \textsc{gend} \ \textsc{n}
\end{bmatrix}
\end{bmatrix}$$

3.3 Adverbs

(78) corresponds to (41), where the predicative adverb is controlled by an InfP. The dative dependent of the adverb (AUTHOR, 3) is the subject of the infinitive (GET, 1):

(78)
$$\begin{bmatrix} \text{PRED} & \text{`BE<2>1'} \\ \text{SUBJ} & 1 \begin{bmatrix} \text{PRED} & \text{`GET<3,4>'} \\ \text{SUBJ} & 3 \\ \text{OBJ} & 4 \begin{bmatrix} \text{PRED} & \text{`DETAILS'} \end{bmatrix} \end{bmatrix} \\ \text{XC-P} & 2 \begin{bmatrix} \text{PRED} & \text{`DIFFICULT<1,3>'} \\ \text{SUBJ} & 1 \\ \text{OBJ}_\theta & 3 \begin{bmatrix} \text{PRED} & \text{`AUTHOR'} \end{bmatrix} \\ \text{ADJ} & \left\{ \begin{bmatrix} \text{PRED} & \text{`MOST'} \end{bmatrix} \right\} \end{bmatrix} \end{bmatrix}$$

(79) corresponds to (38), where the predicative adverb is controlled by a CP. The nominal complement of FEEL (RESPECT, 4), is an OBL because it does not passivise.

(79)
$$\begin{bmatrix} \text{PRED} & \text{`BE<2>1'} \\ \text{SUBJ} & 1 \begin{bmatrix} \text{PRED} & \text{`FEEL<3,4>'} \\ \text{SUBJ} & 3 \begin{bmatrix} \text{PRED} & \text{`PRO'} \end{bmatrix} \\ \text{OBL} & 4 \begin{bmatrix} \text{PRED} & \text{`RESPECT'} \\ \text{ADJ} & \left\{ \begin{bmatrix} \text{PRED} & \text{`SOME'} \end{bmatrix} \right\} \end{bmatrix} \\ \text{COMP-FORM} & \text{THAT} \end{bmatrix} \\ \text{XC-P} & 2 \begin{bmatrix} \text{PRED} & \text{`GOOD<1,5>'} \\ \text{SUBJ} & 1 \\ \text{OBJ}_\theta & 5 \begin{bmatrix} \text{PRED} & \text{`PRO'} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

3.4 Without a copula

The f-structures in this section represent examples where, in absence of a copula, the predicative item acts as the main predicate. Compare these to f-structures in earlier sections where the copula is present – they correspond to the XC-P function.

(80) corresponds to (49), where the predicative nominal is controlled by an InfP. Compare (80) to its counterpart with a copula, (74).

(80)
$$\begin{bmatrix} \text{PRED} & \text{`SCANDAL<1>'} \\ \text{SUBJ} & 1 \begin{bmatrix} \text{PRED} & \text{`DISREGARD<2,3>'} \\ \text{SUBJ} & 2 \begin{bmatrix} \text{PRED} & \text{`PRO'} \end{bmatrix} \\ \text{OBJ} & 3 \begin{bmatrix} \text{PRED} & \text{`NATION'} \\ \text{ADJ} & \left\{ \begin{bmatrix} \text{PRED} & \text{`OWN'} \end{bmatrix} \right\} \end{bmatrix} \end{bmatrix} \\ \text{CASE} & \text{NOM} \end{bmatrix}$$

(81) corresponds to (45), where the superlative predicative adjective is controlled by a CP. Compare (81) to its counterpart with a copula, (77).

(81)
$$\begin{bmatrix} \textsc{pred} & \text{`IMPORTANT<}\boxed{1}\text{>'} \\ \textsc{subj} & \boxed{1}\begin{bmatrix} \textsc{pred} & \text{`REACH<}\boxed{3},\boxed{4}\text{>'} \\ \textsc{subj} & \boxed{3}\left[\textsc{pred} \quad \text{`PRO'}\right] \\ \textsc{obl} & \boxed{4}\left[\textsc{pred} \quad \text{`DESTINATION'}\right] \\ \multicolumn{2}{l}{\textsc{comp-form that} \quad \textsc{impersonal} +} \end{bmatrix} \\ \textsc{adj} & \left\{\left[\textsc{pred} \quad \text{`MOST'}\right]\right\} \\ \multicolumn{2}{l}{\textsc{case nom} \quad \textsc{num sg} \quad \textsc{gend n}} \end{bmatrix}$$

(82) corresponds to (46), where the superlative predicative adverb is controlled by an InfP. The dative dependent of the adverb (PEOPLE, 2) is the subject of the infinitive (RECONCILE, 1). Compare (82) to its counterpart with a copula, (78).

(82)
$$\begin{bmatrix} \textsc{pred} & \text{`DIFFICULT<}\boxed{1},\boxed{2}\text{>'} \\ \textsc{subj} & \boxed{1}\begin{bmatrix} \textsc{pred} & \text{`RECONCILE<}\boxed{2},\boxed{3}\text{>'} \\ \textsc{subj} & \boxed{2} \\ \textsc{obl} & \boxed{3}\left[\textsc{pred} \quad \text{`THIS'}\right] \end{bmatrix} \\ \textsc{obj}_\theta & \boxed{2}\begin{bmatrix} \textsc{pred} & \text{`PEOPLE'} \\ \textsc{adj} & \left\{\left[\textsc{pred} \quad \text{`YOUNG'}\right]\right\} \end{bmatrix} \\ \textsc{adj} & \left\{\left[\textsc{pred} \quad \text{`MOST'}\right]\right\} \end{bmatrix}$$

3.5 With verbs taking an object-controlled predicative complement

(83) corresponds to (50), where the predicative nominal (inside a PP) is controlled by a CP which is the object (2) of the main predicate (CONSIDER):

(83)
$$\begin{bmatrix} \textsc{pred} & \text{`CONSIDER<}\boxed{1},\boxed{3}\text{>}\boxed{2}\text{'} \\ \textsc{subj} & \boxed{1}\left[\textsc{pred} \quad \text{`PRO'}\right] \\ \textsc{obj} & \boxed{2}\begin{bmatrix} \textsc{pred} & \text{`AFFECT<}\boxed{4},\boxed{5}\text{>'} \\ \textsc{subj} & \boxed{4}\left[\textsc{pred} \quad \text{`DISEASE'}\right] \\ \textsc{obj} & \boxed{5}\left[\textsc{pred} \quad \text{`PEOPLE'}\right] \\ \textsc{adj} & \left\{\left[\textsc{pred} \quad \text{`BECAUSE...'}\right]\right\} \\ \multicolumn{2}{l}{\textsc{comp-form} \quad \textsc{that}} \end{bmatrix} \\ \textsc{xc-p} & \boxed{3}\begin{bmatrix} \textsc{pred} & \text{`RULE<}\boxed{2}\text{>'} \\ \textsc{subj} & \boxed{2} \\ \multicolumn{2}{l}{\textsc{case acc} \quad \textsc{pform as}} \end{bmatrix} \end{bmatrix}$$

(84) corresponds to (54), where the predicative adjective (inside a PP) is controlled by a CP which is the object (2) of the main predicate (CONSIDER):

(84)
$$\begin{bmatrix}
\text{PRED} & \text{'CONSIDER<1,3>2'} \\
\text{SUBJ} & 1\begin{bmatrix}\text{PRED} & \text{'PRO'}\end{bmatrix} \\
\text{OBJ} & 2\begin{bmatrix}
\text{PRED} & \text{'MAY<5>4'} \\
\text{SUBJ} & 4\begin{bmatrix}\text{PRED} & \text{'VIRUS'}\end{bmatrix} \\
\text{XCOMP} & 5\begin{bmatrix}\text{PRED} & \text{'SPREAD<4>'} \\ \text{SUBJ} & 4\end{bmatrix} \\
\text{COMP-FORM} & \text{THAT}
\end{bmatrix} \\
\text{XC-P} & 3\begin{bmatrix}
\text{PRED} & \text{'PROBABLE<2>'} \\
\text{SUBJ} & 2 \\
\text{CASE NOM} & \text{NUM SG} \quad \text{GEND N} \quad \text{PFORM AS}
\end{bmatrix}
\end{bmatrix}$$

(85) corresponds to (59), where the predicative adjective (inside a PP) is controlled by a CP which is the subject (1) of the main predicate (CONSIDER). This is because it is the passive rough[5] counterpart of (54), represented in (84) above. As a result of the difference in voice, while in (84) the predicative adjective is controlled by the active object, in (85) it is controlled by the passive subject.

(85)
$$\begin{bmatrix}
\text{PRED} & \text{'CONSIDER<2,3>1'} \\
\text{SUBJ} & 1\begin{bmatrix}
\text{PRED} & \text{'MAY<5>4'} \\
\text{SUBJ} & 4\begin{bmatrix}\text{PRED} & \text{'VIRUS'}\end{bmatrix} \\
\text{XCOMP} & 5\begin{bmatrix}\text{PRED} & \text{'SPREAD<4>'} \\ \text{SUBJ} & 4\end{bmatrix} \\
\text{COMP-FORM} & \text{THAT}
\end{bmatrix} \\
\text{OBL-AG} & 2\begin{bmatrix}\text{PRED} & \text{'EXPERT'}\end{bmatrix} \\
\text{XC-P} & 3\begin{bmatrix}
\text{PRED} & \text{'PROBABLE<1>'} \\
\text{SUBJ} & 1 \\
\text{CASE NOM} & \text{NUM SG} \quad \text{GEND N} \quad \text{PFORM AS}
\end{bmatrix} \\
\text{PASSIVE} & +
\end{bmatrix}$$

3.6 With coordination

(86) corresponds to (60), which is particularly demanding because it involves a complex case of coordination, namely non-constituent coordination. What is interesting is that the first conjunct consists of a predicative nominal (MADNESS, 3) controlled by an infinitival phrase (LET_BE_CAUGHT,[6] 2), while the second conjunct contains a predicative nominal (COWARDICE, 7) controlled by another nominal (SUICIDE, 6).

[5]The *by*-phrase in (59), *by experts*, does not correspond to the implicit first person subject in (54).

[6]This is a placeholder for a far more complex structure – (86) is complex enough.

(86)
$$\left[\left\{\left(\begin{bmatrix}\text{PRED} & \text{`CONSIDER<}\boxed{1}\text{,}\boxed{3}\text{>}\boxed{2}\text{'} \\ \text{SUBJ} & \boxed{1}\begin{bmatrix}\text{PRED} & \text{`PRO'}\end{bmatrix} \\ \text{OBJ} & \boxed{2}\begin{bmatrix}\text{PRED} & \text{`LET_BE_CAUGHT<}\boxed{4}\text{,}\boxed{5}\text{>'} \\ \text{SUBJ} & \boxed{4}\begin{bmatrix}\text{PRED} & \text{`PRO'}\end{bmatrix} \\ \text{OBJ}_\theta & \boxed{5}\begin{bmatrix}\text{PRED} & \text{`NAZI'} \\ \text{CASE} & \text{DAT}\end{bmatrix}\end{bmatrix} \\ \text{XC-P} & \boxed{3}\begin{bmatrix}\text{PRED} & \text{`MADNESS<}\boxed{2}\text{>'} \\ \text{SUBJ} & \boxed{2} \\ \text{CASE ACC} & \text{PFORM AS}\end{bmatrix}\end{bmatrix}, \right.\right.$$
$$\left.\left.\begin{bmatrix}\text{PRED} & \text{`CONSIDER<}\boxed{1}\text{,}\boxed{7}\text{>}\boxed{6}\text{'} \\ \text{SUBJ} & \boxed{1} \\ \text{OBJ} & \boxed{6}\begin{bmatrix}\text{PRED} & \text{`SUICIDE'} \\ \text{CASE} & \text{ACC}\end{bmatrix} \\ \text{XC-P} & \boxed{7}\begin{bmatrix}\text{PRED} & \text{`COWARDICE<}\boxed{6}\text{>'} \\ \text{SUBJ} & \boxed{6} \\ \text{CASE ACC} & \text{PFORM AS}\end{bmatrix}\end{bmatrix}\right)\right\}, \begin{bmatrix}\text{COORD-FORM} & \text{AND}\end{bmatrix}\right]$$

4 Conclusion

This paper presented a vast array of data from Polish illustrating predicative complements with a non-canonical, clausal controller – a complementiser phrase (CP) or an infinitival phrase (InfP). It discussed three types of predicative complements: nominals, adjectives and adverbs, the last of which has been rejected previously.

It presented arguments showing that clauses are subjects in the relevant predicative constructions. It also discussed examples where the copula is not present, when the predicative item functions as the main predicate – just as in standard copular constructions (with a nominal controller). Furthermore, it presented examples which involve a predicative complement controlled by an object, together with their passive variant, where the predicative complement is controlled by the subject. The aim was to argue convincingly that clauses can control predicative items.

While default agreement triggered by clausal subjects has been previously discussed in the context of subject-verb agreement, this paper discussed default agreement observed on predicative adjectives controlled by a clause.

Counter to previous claims, this paper argued for the existence of predicative adverbs, which can be predicated of clauses (CP, InfP) in Polish. It also discussed the issue of control into the infinitival subject by the dative argument of the predicative adverb, providing a unified analysis of such control.

Finally, this paper provided a simple LFG formalisation of the presented analysis, implemented and tested in XLE.

References

Arka, I Wayan and Simpson, Jane. 1998. Control and complex arguments in Balinese. In Miriam Butt and Tracy Holloway King (eds.), *The Proceedings of the LFG'98 Conference*, University of Queensland, Brisbane: CSLI Publications.

Bondaruk, Anna. 2013. *Copular Clauses in English and Polish*. Lublin: Wydawnictwo KUL.

Bondaruk, Anna and Szymanek, Bogdan. 2007. Polish Nominativeless Constructions with Dative Experiencers: Form, Meaning and Structure. *Studies in Polish Linguistics* 4, 61–97.

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

Citko, Barbara. 2004. Small Clauses Reconsidered: Not So Small and Not All Alike. *Lingua* 118, 261–295.

Crouch, Dick, Dalrymple, Mary, Kaplan, Ron, King, Tracy, Maxwell, John and Newman, Paula. 2011. XLE Documentation. http://www2.parc.com/isl/groups/nltt/xle/doc/xle_toc.html.

Dalrymple, Mary, Dyvik, Helge and King, Tracy Holloway. 2004. Copular Complements: Closed or Open? In Miriam Butt and Tracy Holloway King (eds.), *The Proceedings of the LFG'04 Conference*, pages 188–198, Stanford, CA: CSLI Publications.

Dziwirek, Katarzyna. 1990. Default Agreement in Polish. In Katarzyna Dziwirek, Patrick Farrell and Errapel Mejías-Bikandi (eds.), *Grammatical Relations: A Cross-Theoretical Perspective*, Stanford, CA: CSLI Publications.

Hofmeister, Philip and Sag, Ivan. 2010. Cognitive Constraints and Island Effects. *Language* 86, 366–415.

Huang, C.-T. J. 1982. *Logical Relations in Chinese and the Theory of Grammar*. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA.

Kallas, Krystyna. 1993. Kilka uwag na temat charakterystyki akomodacyjnej i konotacyjnej leksemów BYĆ. *Acta Universitatis Nicolai Copernici: Filologia Polska* XI, 27–39.

Maxwell, III, John T. and Manning, Christopher D. 1996. A Theory of Non-constituent Coordination based on Finite-State Rules. In Miriam Butt and Tracy Holloway King (eds.), *The Proceedings of the LFG'96 Conference*, Grenoble, France: CSLI Publications.

Patejuk, Agnieszka. 2015. *Unlike coordination in Polish: an LFG account*. Ph.D. dissertation, Institute of Polish Language, Polish Academy of Sciences, Cracow.

Patejuk, Agnieszka and Przepiórkowski, Adam. 2014. In Favour of the Raising Analysis of Passivisation. In Miriam Butt and Tracy Holloway King (eds.), *The Proceedings of the LFG'14 Conference*, pages 461–481, Stanford, CA: CSLI Publications.

Przepiórkowski, Adam. 1999. *Case Assignment and the Complement-Adjunct Dichotomy: A Non-Configurational Constraint-Based Approach*. Ph.D. dissertation, Universität Tübingen, Tübingen.

Przepiórkowski, Adam, Bańko, Mirosław, Górski, Rafał L. and Lewandowska-Tomaszczyk, Barbara (eds.). 2012. *Narodowy Korpus Języka Polskiego*. Warsaw: Wydawnictwo Naukowe PWN.

Rothstein, Susan. 2001. *Predicates and Their Subjects*. Springer.

A COMP-less Approach to Hungarian Complement Clauses

Péter Szűcs
University of Debrecen

Proceedings of the LFG'18 Conference

University of Vienna

Miriam Butt, Tracy Holloway King (Editors)

2018

CSLI Publications

pages 325–342

http://csli-publications.stanford.edu/LFG/2018

Keywords: complement clauses, LMT, grammatical functions

Szűcs, Péter. 2018. A COMP-less Approach to Hungarian Complement Clauses. In Butt, Miriam, & King, Tracy Holloway (Eds.), *Proceedings of the LFG'18 Conference, University of Vienna*, 325–342. Stanford, CA: CSLI Publications.

Abstract

This paper engages in the (X)COMP debate in LFG. It argues that the view from Hungarian supports a "reductionist" position, as Hungarian complement clauses are easily amenable to an analysis with non-COMP functions. I also remark on the wider picture and side with those who would like to maintain a parsimonious inventory of grammatical functions in LFG, in conjunction with a reworked theory of functional and anaphoric control.

1. Introduction: the (X)COMP debate[1]

There has been a debate in LFG about the grammatical function(s) (GFs) that complement clauses may have. The necessity of the COMP function has been in the heart of the debate. In their seminal paper, Dalrymple & Lødrup (2000) argue that finite complement clauses in English may have either OBJ or COMP grammatical function, depending on the lexical properties of the given predicates. They motivate this bifurcation with differing grammaticality patterns with regards to alternation with NP/DP objects, passivization and coordination, among others. Thus for them, the data in (1)-(6) justifies an analysis where the complement clause of *believe* is an OBJ, while that of *hope* is a COMP. Similar views are expressed in Lødrup (2012) and Belyaev (2017).

(1) a. *I believe that Kate is the winner.*

b. *I believe the story/ it.*

(2) *I believe the story and that it means a lot to you.*

(3) *That Kate won was believed by no one.*

(4) a. *I hope that Kate is the winner.*

b. **I hope the story/it.*

(5) **That Kate would win was hoped by no one.*

There is a more "reductionist" alternative proposal, put forward by Alsina et al. (2005), according to which the COMP function should be dropped from the inventory of grammatical functions in LFG and every finite complement clause should receive some other GF. This position is supported by Forst (2006) and Patejuk & Przepiórkowski (2014, 2016). In this view, the lack of direct object nominals for *hope* means that the complement function is actually an OBL_{θ}. From this perspective it is not surprising that we find PP-alternatives to it.

[1] The Project no. 111918 (*New approaches in the description of the grammar of Hungarian pronominals)* has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the K funding scheme.

(6) *Kate hopes for a better result next time.*

Less attention has been paid to nonfinite complements. Alsina et al. (2005: 41) mentions that as "XCOMP may be considered a special case of COMP, XCOMP should probably go the same way as COMP", but no detailed investigation is carried out. Patejuk & Przepiórkowski (2014, 2016) in their argumentation for a radical reduction of GFs in LFG explicitly push for the elimination of (X)COMP. They cite examples like (7) to argue that OBJ can also be controlled. They also show that there is a way to implement such an analysis in XLE.

(7) *I just want friends and to be happy.*

From an entirely different ("expansionist") perspective, Falk (2005) proposes that the inventory of grammatical functions in LFG should be enriched, to properly model the cross-linguistic category-function correlations outlined in (8). (Note that Patejuk & Przepiórkowski 2016 explicitly argue against the existence/significance of such correlations.) Falk (2005) puts forward an expanded version of Lexical Mapping Theory (see Table 1), in which he posits two new open functions, $XOBL_{\theta}$ and $XOBJ_{\theta}$ for the complements like the ones in (9) and (10), respectively.

(8) a. NP, DP – OBJ b. PP – OBL

c. S, IP, CP – COMP d. InfP[2] – XCOMP

(9) *The transformationalist strikes me <u>as crazy</u>.*

(10) *The transformationalist stayed <u>crazy</u>.*

		-r	*+r*	
		+s		*-s*
-c	*-o*	SUBJ	OBL_{θ}	$XOBL_{\theta}$
	+o	OBJ	OBJ_{θ}	$XOBJ_{\theta}$
+c	*+/-o*		COMP	XCOMP

Table 1.
Grammatical functions in Falk (2005). (*r*: restricted, *o*: objective, *c*: complement function, *s*: saturated).

The goal of this paper is to add Hungarian to the set of languages that are considered from these perspectives. Overall, I align myself more with the

[2] The syntactic category of infinitival clauses may be IP, CP or VP, depending on the particular analysis. I remain neutral on this issue, so the abbreviation "InfP" is used as a shorthand throughout the paper.

"reductionist" camp in that I argue that in Hungarian the (X)COMP function is not justified: any possible occurrences may be reduced to non-COMP functions. Whether Falk's (2005) new functions are justified depends on one's take on the nature of functional control and the already mentioned category-function correlations. Theoretical and cross-linguistic considerations weigh rather against than for Falk's (2005) "expansionist" view.

2. Closed complement functions in Hungarian

I this section a discuss those cases of Hungarian whereby the complement is functionally complete, so no GF is predicated from the outside. This happens because the subject/object/oblique argument is a simple nominal or a clause with its own subject.

The basic pattern is this: Hungarian complement clauses may function either as SUBJ, OBJ or OBL_{θ} arguments of their respective predicates. The primary evidence for this is that there is a systematic alternation in Hungarian whereby the respective grammatical function is realized as a) a lexical noun; b) a pronoun; c) a *that*-clause; d) an infinitival clause.

Let us take a look at the case of SUBJ first. The pattern described above is illustrated in (11) below. That is, *derogál* ('feels derogatory') has the subcategorization frame outlined in (12). Parallel examples could be construed with *kellemetlen* ('unpleasant'), *sikerül* ('successfully works out'), *bejön* ('be appealing'), etc.

(11) a. *A vereség derogál Katinak.*
the defeat feels.derogatory Kate.DAT
'The defeat feels derogatory to Kate.'

b. *Az derogál Katinak, hogy vereséget szenvedett.*
that feels.derogatory Kate.DAT that(c)[3] defeat.ACC suffered.
'It feels derogatory to Kate that she was defeated.'

c. *Derogál Katinak, hogy vereséget szenvedett.*
feels.derogatory Kate.DAT that(c) defeat.ACC suffered.
'That she was defeated feels derogatory to Kate.'

d. *Derogál Katinak vereséget szenvedni.*
feels.derogatory Kate.dat defeat.ACC suffer.INF
'To be defeated feels derogatory to Kate.'

(12) *derogál* <(SUBJ)(OBL)>

[3] The "c" stands for "complementizer". This is to avoid any confusion with the demonstrative in such sentences. If not indicated otherwise, nominative case and present tense assumed in the glosses.

As pointed out by Rákosi & Laczkó (2005), this pattern makes straightforward sense if we assume that the underlined constituents uniformly function as the SUBJ of *derogál* ('feels derogatory to'), regardless of their categorial status.

This is quite straightforward in the case of (11a). In (11b) the subject is the nominative pronoun, and the *that*-clause is in an appositive relation to this, functioning as an ADJUNCT.[4] If there is no pronoun, just a *that*-clause, as in (11c), the clause itself is the SUBJ argument.

In (11d) the infinitival clause itself functions as the subject of the main predicate. The understood subject of the infinitival is obligatorily controlled by the second, dative argument of *derogál* ('feels derogatory to'). This is a major difference compared to the English translation, where the infinitival subject has arbitrary reference. The contrast may be seen from the fact that an explicit subject may be added in English in the form of a *for*-phrase, but not in Hungarian (Rákosi 2006: 212).

(13) *For Peter to be defeated feels derogatory to Kate.*

(14) **Derogált* *Katinak* *Péternek* *vereséget* *szenvedni.*
felt.derogatory.3SG Kate.DAT Peter.DAT defeat.ACC suffer.INF

One might suggest that (14) is ungrammatical because there is simply no structural place in the infinitival clause for the overt subject *Peter*. However, as known since Szabolcsi (2009), Hungarian infinitival clauses do provide a slot for overt subjects, as long as they are pronominal in form and co-referent with the controller. That is, the 'pro' subject of the infinitival may be overt as long as it conforms to the normal requirements of the obligatory anaphoric control relations. Szabolcsi (2009) discusses regular, nonsubject clauses, but the argument smoothly carries over to subject infinitivals (the subject of these bears dative case). The additional requirement is that the overt pronominal has to be under the scope of some discourse or quantificational operator. This is just the standard requirement for overt pronominals in such positions in a pro-drop language like Hungarian.

(15) *Derogált* *Katinak*$_i$ *csak* *neki*$_{i/*j}$ *vereséget*
felt.derogatory.3SG Kate.DAT only her.DAT defeat.ACC
szenvedni.
suffer.INF
'Only for her to be defeated felt derogatory to Kate.'

Interestingly, the extraposition-version (as in the translation of (11b)) does not work in Hungarian for the infinitival. Rákosi & Laczkó (2005) explains this by

[4] This claim is related to the debate in Hungarian linguistics about the status of such pronouns. Here they are treated as contentful demonstratives and not expletives (contra Kenesei's (1994) more or less standard account). For a detailed argumentation for this position, see Szűcs (2015) and references therein.

stating a requirement that the clause functioning as the adjunct for the demonstrative cannot be headed an infinitival. Hence the contrast in (16) vs. (11b). In sum, the various structures in (11) are realizations of the basic schema shown in (12), the underlined parts of (11) being the SUBJ argument of *derogál* ('feel derogatory to').

(16) **Az derogált Katinak vereséget szenvedni.*
that felt.derogatory.3SG Kate.DAT defeat.ACC suffer.INF

Moving on to object clauses, a parallel pattern emerges. The object argument of a verb like *akar* ('want') may be realized as an NP/DP (pronoun, lexical noun), a finite clause or an infinitival. The straightforward approach here is also to posit a single lexical entry. Similar examples could be construed with a *próbál* ('try'), *utál* ('hate'), *szeret* ('like'), etc.

(17) a. *Kati ételt akar.*
Kate food.ACC wants.
'Kate wants food.'

b. *Kati azt akarja, hogy együnk.*
Kate that.ACC wants that(c) eat.3PL.SBJV
'Kate wants (it) that we eat.'

c. *Kati akarja, hogy együnk.*
Kate wants that(c) eat.3SG.SBJV
'Kate wants that we eat.'

d. *Kati enni akar.*
Kate eat.INF wants
'Kate wants to eat.'

(18) *akar* <(SUBJ)(OBJ)>

Finally, the same pattern emerges with OBL_{θ} complements: *fél* ('fear') may occur with a lexical noun or pronoun marked with ablative case, a finite or a non-finite complement clause. Other example verbs are *készül* ('prepare'), *törekszik* ('strive') or *vonakodik* ('be reluctant').

(19) a. *Kati fél a kutyáktól.*
Kate fears the dogs.from
'Kate fears dogs.'

b. *Kati attól fél, hogy a kutya megharapja.*
Kate that.from fears that(c) the dog bites.DEF
'Kate fears that the dog may bite her.'

c. *Kati fél, hogy a kutya megharapja.*
Kate fears that(c) the dog bites.DEF
'Kate fears that the dog may bite her

d. *Kati* *fél* *kutyát tartani*.
Kate fears dog.ACC keep.INF
'Kate fears keeping a dog.'

(20) *fél* <(SUBJ)(OBL_θ)>

An interesting contrast between OBJ and OBL_θ infinitives may be observed in the so-called "long-distance object definiteness agreement"-phenomenon in Hungarian (first described by É. Kiss 1989 and Kálmán et al. 1989). Szécsényi & Szécsényi (2017) observes that a finite verb may agree in definiteness with the object of its infinitival clause, but this only happens if the main verb is what Szécsényi & Szécsényi (2017) calls an "agreeing verb".[5] The distance between the agreement trigger (the embedded object) and the agreement target (the finite matrix verb) may be arbitrarily long as long as the path only contains "agreeing verbs". This distinction finds a natural home in an LFG setting as the "agreeing" category shows a near perfect correlation with OBJ-taking verbs while the "non-agreeing" category may be equated with OBL_θ-verbs.[6] The phenomenon is illustrated in (21)-(22) below.

(21) a. *Kati* *akar* *olvasni* *egy* *könyvet.*
Kate wants.INDEF read.INF a book.ACC
'Kate wants to read a book.

b. *Kati* *akar-ja* *olvasni* *a* *könyvet.*
Kate wants-DEF read.INF the book.ACC
'Kate wants to read the book.'

(22) a. *Kati* *fél* *olvasni* *egy* *könyvet.*
Kate fears read.INF a book.ACC
'Kate fears reading a book.

b. *Kati* *fél(*-i)* *olvasni* *a* *könyvet.*
Kate fears(-DEF) read.INF the book.ACC.
'Kate fears reading the book.'

In (21a) the object of the infinitive (*a book*) is indefinite and the matrix verb is in the default indefinite conjugation. In contrast, the definite object in (21b) (*the book*) triggers definite conjugation on *akar* ('want'). No such variation may be observed with *fél* ('fears'): regardless of the definiteness of the embedded object, it is in the default indefinite paradigm. This may be modelled with the following lexical entry on definiteness suffixes in Hungarian. (23) ensures that the agreement path may traverse through OBJ functions, but an

[5] See Bárány (2015) for a detailed investigation about definiteness-agreement in Hungarian.

[6] The picture is slightly blurred by the fact that some auxiliaries also participate in a long-distance agreement process (e.g. *fog* ('will'), *talál* ('happen'), etc., see also example (40)). I leave this complication to further research.

intervening OBL will block it. If (23) is not satisfied, the default indefinite paradigm appears (as elsewhere case).

(23) (↑OBJ^{+} DEF)=$_{c}$ +

The above solution has the drawback of introducing non-locality into agreement, which is theoretically dispreferred. An alternative would be to posit a feature-sharing agreement mechanism, as Haug & Nikitina (2016) suggests for Latin dominant participles. This essentially means that the definiteness feature from the most embedded object "percolates" up to the infinitival itself if it bears the OBJ function and the main verb agrees in definiteness with the infinitival. This may happen in an arbitrary number of steps and locality is ensured. Under this approach the following equation would be available on every infinitival verbal lexical item as an option.[7] See Figure 1, for (21).

(24) (↑DEF) = (↑OBJ DEF)

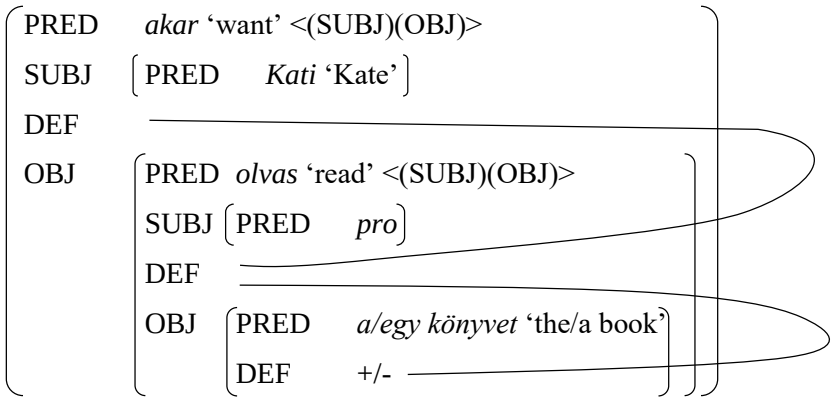


Figure 1.
F-structure for (21), with feature-sharing.

It is to be noted that in all the scenarios above, the different manifestations of the respective grammatical functions may be coordinated, which provides evidence for the uniform functional analysis. Some examples demonstrating this are shown below: in (25) an infinitival is coordinated with a pronoun, in

[7] As this is not the main concern of this paper, the ramifications for the overall system of Hungarian agreement are left for further research. For example, finite clauses are not "transparent", they always count as definite, regardless of their object:

(i)	*Ígére-m/*-k*	*elolvasok*	*egy*	*könyvet.*
	promise-DEF/*INDEF.1SG	read.1SG.INDEF	a	book.ACC
	'I promise I read a book.'			

(26) a lexical noun is coordinated with an infinitival, in (27) a lexical noun is coordinated with a finite *that*-clause. Other combinations are also possible.

(25) *Derogál Katinak vereséget szenvedni és az,*
feels.derogatory Kate.DAT defeat.ACC suffer.INF and that
hogy ez ilyen gyakran megtörténik.
that(c) this so often happens
'To be defeated and that it happens so often feels derogatory to Kate.'

(26) *Kati ételt és azzal jóllakni akar.*
Kate food.ACC and that.with satisfied.become.INF wants
'Kate wants food and to be satisfied with it.'

(27) *Kati fél a kutyáktól és hogy azok*
Kate fears the dogs.from and that(c) those
megharapják.
bite.3PL
'Kate fears dogs and that they might bite her.'

The general conclusion to be drawn from this section is that the COMP function need not be invoked in the analysis of Hungarian that-clauses. In every case, they are straightforwardly amenable to an analysis in terms of SUBJ, OBJ or OBL_{θ}. The systematic alternation and the coordination possibilities make the alternative, COMP-based alternative unlikely.

3. Open complement functions in Hungarian

An open argument function contains a grammatical function (usually the SUBJ) which is the target of a functional control equation, i.e. it is predicated from outside. The stock example for this is the raising construction, where the non-thematic matrix subject is functionally identified with the subject of the infinitival clause.[8] In standard LFG, the infinitival bears the XCOMP grammatical function.

(28) *Kate seems to be happy.*

As often noted in the literature, Hungarian seems to make a restricted use of InfP in such raising structures. The literal equivalent of (28) is ungrammatical and the state of being happy is expressed as a case-marked adjective.[9] (A finite clause along the lines of *It seems that Kate is happy* is also an option (see 36c below), but that is irrelevant at this point.)

[8] Though it is less recognized, such raising structures do occur with finite clauses as well. For an overview, see Ademola-Adeoye (2010).
[9] See also Laczkó (2012: 50) for similar points about raising in Hungarian.

(29) a. **Kati boldog lenni tűnik.*
Kate happy be.INF seems.

b. *Kati boldog-nak tűnik.*
Kate happy-DAT seems.
'Kate seems happy.'

It must be noted that some examples of the pattern *verb*.INF+*tűnik* may be found in the Hungarian National Corpus. However, this is quite limited: the Hungarian National Corpus returns 41 hits (on closer investigation, even some of these are irrelevant examples). In comparison, the *adj*.DAT+*tűnik* pattern returns 4210 sentences. A the closely related *látszik* (approx. 'appears') shows a much more balanced distribution (ca. 3000 hits with both patterns).

Nominals are also acceptable if they are predicative and not referential, as shown in (30). (31) is an example with a transitive main verb ("raising to object").

(30) *Kati (*az) okos lány-nak tűnik.*
Kate the smart girl-DAT seems
'Kate seems a smart girl.'

(31) *Kati-t boldog-nak/ zseni-nek tartom.*
Kati-ACC happy-DAT genius-DAT consider.1SG
'I consider Kate happy/ a genius.'

Falk (2005: 139) notes that in English, prepositional phrases with an adjectival meaning may be complements of *seem*. This seems to be barred in Hungarian (*magán kívül van* 'to be outside of oneself' is a fixed expression in Hungarian, meaning 'to be mad/dazzled').

(32) ?*Kate seems out of his mind.*

(33) **Kati magán kívül látszik / tűnik.*
Kate herself outside seems appears

Furthermore, English seems to allow non-adjectival PPs as in (34), but Hungarian lacks this option as well.

(34) *I want you out of the room.*

(35) **Ki/ Kint akarlak téged a szobából.*
out outside want.1SG you.ACC the room.from

What can we distill from this distribution? My position is that the XCOMP function may be dispensed with, regardless of our decision of "reductionist" (Alsina et al. 2005, Patejuk & Przepiórkowski 2014, 2016) or the "expansionist" path (Falk 2005).

In Falk's (2005) approach, the grammatical function of raising infinitivals would be $XOBJ_{\theta}$, as their most natural realization is AP and NP. I suggest that

even the InfPs may be analyzed as this GF. This should not be a controversial idea since the strict correlation of InfPs with a grammatical function has already been broken in the previous section (there they are SUBJ, OBJ or OBL_{θ}).

The reductionist take is that any of the standard grammatical functions may be functionally controlled. In other words, there is "XSUBJ", "XOBJ", "$XOBL_{\theta}$", and "$XOBJ_{\theta}$", but instead of supplying the "X" label, we just need an appropriate theory of functional control. I will briefly look into these matters in the next section.

At any rate, the $(X)OBJ_{\theta}$ seems to be an appropriate function for raising in Hungarian and XCOMP is not needed. Now we have eliminated both "comp" functions from the inventory of the GFs in Hungarian. This again could make sense from both the "reductionist" and the "expansionist" perspective. This is trivial for the "reductionist" camp, but Falk (2005) also mentions that the presence of the +/-*c* feature could be a matter of cross-linguistic variation, suggesting that Hebrew is a language without +*c* functions and according to Falk 2005 (referring to Dalrymple and Lødup 2000) Norwegian also makes a very restricted use of COMP and XCOMP.

4. Argument-structure

Now that I have outlined my position on the general situation in Hungarian, now it is possible to elaborate on some details of the emerging general picture. In particular, I comment on how the lexical entries may be handled in terms of argument structure. Two issues arise: uniformity of the lexical entries required for the various realizations of the GFs, and the perspective of Lexical Mapping Theory (LMT).

4.1 Lexical uniformity

Under the conclusions reached in sections 2-3, a problem with the functional subcategorization of the lexical entries emerges: how to attribute the same lexical entry to the controlled (the infinitival) and the uncontrolled (CP, DP, NP) manifestations of the respective predicates? In standard LFG, f-structural identities are encoded by annotations of identity for raising (36 (=29b)) or co-reference for equi (37(=17d)). But then, such annotations are clearly inoperative in (36c) and (37c (=17a)) and would result in invalid f-structures for these sentences. (Note the parallel in the English translations.)

(36) a. *Kati* *boldognak* *tűnik*.
Kate happy.DAT seems
'Kate seems happy.'

b. *tűnik* <(XOBJ$_\theta$)>SUBJ
(SUBJ)= (XOBJ$_\theta$ SUBJ)

c. *Úgy tűnik, hogy Kati boldog.*
so seems that(c) Kate happy
'It seems that Kate is happy.'

(37) a. *Kati enni akar.*
Kate eat.INF wants
'*Kate wants to eat*.'

b. *want* <(SUBJ)(OBJ)>
(SUBJ INDEX) = (OBJ SUBJ INDEX)[10]

c. *Kati ételt akar.*
Kate food.ACC wants
'Kate wants food.'

To maintain a uniform analysis, a modification is needed in how to establish the control relationship. I find the ideas expressed in Alsina (2008) attractive in this matter.[11]

He argues that LFG should abandon the lexically encoded annotations of the kind expressed in (36b) and (37b) and the identity-relations should be the results of general constraints like the ones in (38). in addition to the well-established Completeness and Coherence conditions of LFG.

[10] As *want* is a control verb, there is a referential identity between the main clause subject and the implicit ("PRO") subject of the embedded predicate. Thus, for *want*, I subscribe to anaphoric control. For an illuminating discussion on functional and anaphoric control in equi-type constructions, see Falk (2001: 136-139).

[11] While I largely agree with the spirit of Alsina's (2008) account, certain aspects of it seem too restrictive, e.g. forbidding structure sharing into a finite clause (his SUBJ Binding Condition). As already mentioned (footnote 8), finite raising does exist. Also, finite equi-like structures also seem to be possible, see e.g. Ince (2006) on Turkish. A possible way to reconcile these with Alsina (2008) is to rely on constraint-ranking, whereby certain constraints allowing finite control outrank the SUBJ Binding Condition.

Additionally, Alsina's (2008) account makes a strict correlation between raising and functional control on the one hand, and equi and anaphoric control on the other (as only non-thematic arguments may be structure-shared). However, it is likely that certain equi-verbs establish functional control (see Falk 2001: 136-139). The resolution of these issues is a task for the future.

(38) a. **Subject Condition**[12]
Every verbal f-structure must have a SUBJ and no f-structure may have more than one SUBJ.

b. **Nonthematic Condition on GF Identification**
Structure-sharing of GF s is well-formed only if, in the minimal f-structure containing two structure-shared GFs, one of them:
a) is nonthematic and
b) is more f-prominent than any GF identified with it.

The nonfinite clauses at hand are predicative f-structures, so they must have a SUBJ. This SUBJ must be provided with a PRED-value, otherwise the f-structure becomes semantically incomplete.[13] Hence, structure-sharing (functional identification) is mandated. This is what happens with raising sentences like (36a). However, the same process would violate the nonthematic condition in equi-structures like (37a), since both subjects (the matrix and the infinitival) are thematic arguments. In Alsina's (2008) view, this triggers the appearance of the dummy "pro" PRED value for the embedded subject, bypassing direct structure-sharing in favor of anaphorically binding this 'pro'. While some aspects of the theory will have to be modified to capture the full range of the data (see footnote 10), the main point is that there is a possibility in the LFG framework to posit uniform lexical entries, by getting rid of the equality-annotations in (36)-(37).

4.2 Lexical Mapping Theory

Standard LMT is trivially incompatible with the standard inventory of grammatical functions, as it only provides four options (with *r* and *o* specifications), leaving COMP and XCOMP out of the picture. So either the inventory has to be reduced or LMT has to be augmented. In this section I briefly examine these two options.

---

[12] Note that the "verbal" part in (38a) may well be too narrow, given the existence of nonverbal raising structures, see e.g. (32) and (34) above. Also, a reviewer raised the issue of possibly subjectless verbal clauses in German and Polish. This could mean that the Subject Condition is a matter of parametric variation. Alsina (2008, footnote 7) suggests that this may be modelled with an Optimality Theory-based approach to constraint satisfaction.
Alternatively, as the reviewer noted, it may well be that the Subject Condition is superfluous, given that the Coherence and Completeness conditions are satisfied.

[13] Note that this does not mean that *every* predicator must have a *thematic* subject. For instance, raising verbs subcategorize only for a propositional argument, which may be realized as a finite IP/CP. In this case, Completeness is satisfied and an expletive is only inserted because of the Subject Condition (e.g. (36c), *it seems that Kate is happy*). The nonfinite clauses in (36a, 37a) do not contain a raising predicate, so this is not an option for them.

In the "reductionist" system, nothing special needs to be added. All the GFs are standardly available: SUBJ, OBJ, OBL_{θ}, OBJ_{θ}. The "open" versions of these are handled by the system outlined in the previous section.

With a wider array of GFs, changes obviously have to be made. As already shown in Table 1, Falk (2005) adds the features *c* (complement function) and *s* (saturated) to make room for the extra functions and sets up a fairly complex mapping system to accommodate the various subcategorizations.

A more mainstream conception of LMT is Kibort's (2007) system, which works with a fixed valency template and a single mapping principle: map the argument to the least marked available grammatical function, markedness defined as having + specifications in the feature-space. As noted, this is entirely compatible with the "reductionist" approach.

It is not at all straightforward how Falk's *c* and *s* would fit into this system. Crucially, the main problem is that as long as there is a distinction between open and closed functions, no matter how one places the features into Kibort's (2007) valency frame, the controlled and the noncontrolled lexical entries ((36a) and (37a) vs. (36c) and (37c)) will always represent two separate lexical entries at the functional level. This is a clear disadvantage compared to the "reductionist" position.

Another problematic aspect of Falk's (2005) expanded LMT is its asymmetry in two respects. For instance, *c* is neutral with respect to *o*. Falk justifies this by pointing out that COMP alternates with OBJ, OBJ_{θ} and OBL_{θ}. However, even though COMP also alternates with XCOMP ((36a), (36c)) the very same argument is not used by Falk (2005) to argue that *c* is neutral with respect to *s* as well. Thus the argument from alternation is only selectively employed, as an artificial barrier from having to postulate further grammatical functions in the +*c* realm.

The empirical side also seems to militate against the exclusion of open SUBJ and OBJ. Arka & Simpson (1998) analyze certain subject clauses in Balinese as functionally controlled.[14] Furthermore, Patejuk & Przepiórkowski (2014) argue that Polish contains functionally controlled OBJ clauses, as in (39), where the controlled infinitive is coordinated with a direct object.[15]

(39) *Chcę* *pić* *i* *papierosa.*
want.1SG drink.INF and cigarette.ACC
'I want to drink and (I want) a cigarette.'

There is also some data in Hungarian which point in the direction of functionally controlled OBJ clauses. Based on the long-distance definiteness

[14] See Falk (2006) for a differing view.

[15] Based on case transmission facts, Patejuk & Przepiórkowski (2014) argue that control in Polish is functional.

agreement facts discussed in section 3, one may reach the conclusion that the complement of *kezd* ('begin') is an OBJ. The key fact is that *kezd* ('begin') has a nonthematic subject in this example, as evident from the English translations.[16] Thus, the identification is functional, yielding a raising structure.

(40) a. *János kezd szeretni egy könyvet.*
John begins.INDEF like.INF a book.ACC
'John is beginning to like a book.'

b. *János kezd-i szeretni a könyvet.*
John begins-DEF like.INF the book.ACC
'John is beginning to like the book.'

Overall, it seems to me that the "reductionist" approach is theoretically more elegant and is also better equipped to handle cross-linguistic data.

5. Conclusion and future perspectives

In this paper I examined the landscape of complement clauses, from the perspective of Hungarian. I argued that complement clauses in Hungarian do not necessitate the (X)COMP function. Finite and non-finite complement clauses are analyzable in terms of SUBJ, OBJ, OBL_{θ} and $(X)OBJ_{\theta}$. In my investigation, I surveyed recent trends in LFG's approach toward the possible grammatical functions and while I cannot say that the debate is settled, the overall picture seems to favor the "reductionist" approach.

As one of my reviewers notes, a potential avenue for future research is the extension of the discussion to the analysis of copular clauses. It is important to recognize that copular sentences are not a unitary phenomenon, but several subtypes are to be distinguished, possibly with different versions of the copula (see e.g. Laczkó (2012) and references therein). Some instances lend themselves for a straightforward analysis in terms of OBJ. According to Falk (2005), (41a) is to be analyzed as including a COMP, but given the NP/DP alternative, OBJ is an equally likely option.

(41) a. *The problem is <u>that the hamster will eat the cat</u>.*

b. *The problem is <u>the cat</u>.*

[16] The following alternative, with an expletive subject, makes the non-thematic nature of *begin*'s subject in (40) explicit: 'it is beginning to be the case that John likes a/the book'. This is equivalent in meaning to the sentences in (40).
Note that *begin* also has a use with a thematic subject, as in (i). Crucially, here the embedded predicate is agentive. (For a scope-based argument on this issue, see Szabolcsi (2009: 254-255)).

(i) *John began to run.* (≠*It began to be the case that John ran*.)

Following Laczkó (2012), existential and locational sentences like (42) might include an OBL (this might be implicit in the case of existentials).

(42) a. *There are witches (on Earth).*

b. *The cat is in the room.*

In attributive sentences, the copula might be a pure formative, without subcategorized grammatical functions.

(43) *The cat is hungry.*

However, Dalrymple et al. (2004: 193) contends that the PREDLINK function is better suited for sentences like (43) in English. Laczkó (2012) also argues for a PREDLINK-analysis of certain copular constructions in Hungarian. Both Falk (2005) and the "reductionist" approach are reluctant to recognize this GF as a distinct entity, as neither one can naturally accommodate it. Only a careful consideration of the cross-linguistic data and the theoretical consequences can settle this issue.

References

Ademola-Adeoye, Feyisayo Fehintola. 2010. A Cross-linguistic Analysis of Finite Raising Constructions. Doctoral dissertation, University of KwaZulu-Natal.

Alsina, Alex. 2008. In Uyechi, Linda, Lian-Hee Wee (eds.), *Reality Exploration and Discovery: Pattern Interaction in Language and Life*. Stanford: CSLI Publications.

Alsina, Alex, KP Mohanan, Tara Mohanan. 2005. How to get rid of the COMP. In Butt, Miriam, Tracy H. King (eds.), *Proceedings of the LFG05 Conference*. Stanford: CSLI Publications. 21-41.

Arka, I Wayan, Jane Simpson.1998. Control and complex arguments in Balinese. In Butt, Miriam, Tracy H. King (eds.), *The proceedings of the LFG'98 conference*, University of Queensland, Brisbane: CSLI Publications.

Bárány András. 2015. *Differential object marking in Hungarian and the morphosyntax of case and agreement*. Doctoral dissertation, University of Cambridge.

Belyaev, Oleg. 2017. In Defense of COMP Complementation in Moksha Mordvin. In Butt, Miriam, Tracy H. King (eds.), *Proceedings of the LFG17 Conference*. Stanford: CSLI Publications. 84-103.

Dalrymple, Mary & Helge Lødrup. 2000. The grammatical functions of complement clauses. In Butt, Miriam, Tracy H. King (eds.), *Proceedings of the LFG00 Conference*. Stanford: CSLI Publications. 104-121.

Dalrymple, Mary, Helge Dyvik & Tracy H. King. 2004. Copular complements: Closed or open? In Butt, Miriam & Tracy H. King (eds.) *Proceedings of the LFG04 Conference.* Stanford: CSLI Publications. 188-198.

É. Kiss Katalin. 1989. Egy főnévi igeneves szerkezetről. In Kiefer Ferenc (ed.), *Általános Nyelvészeti Tanulmányok XVII*. Budapest: Akadémiai Kiadó. 153–69.

Falk, Yehuda N. 2001. *Lexical-Functional Grammar*. Stanford: CSLI Publications.

Falk, Yehuda N. 2005. Open argument functions. In Butt, Miriam, Tracy H. King (eds.), *Proceedings of the LFG05 Conference*. Stanford: CSLI Publications. 136-153.

Falk, Yehuda N. 2006. *Subjects and Universal Grammar: An explanatory theory*. Cambridge: Cambridge University Press.

Haug, Dag, Tatiana Nikitina. 2016. Feature-sharing and agreement. *Natural Language & Linguistic Theory* 34(3): 865-910.

Ince, Atakan. 2006. Direct Complement Clauses as Object Control Structures in Turkish. In Bainbridge, Erin, Brian Agbayani (eds.), *Proceedings of the thirty-fourth Western Conference On Linguistics*. Department of Linguistics, California State University, Fresno. 208-221.

Kálmán C. György, Kálmán László, Nádasdy Ádám,, Prószéky Gábor. 1989. A magyar segédigék rendszere. In Kiefer Ferenc (ed.), *Általános Nyelvészeti Tanulmányok XVII*. Budapest: Akadémiai Kiadó. 49–103.

Kenesei István. 1994. Subordinate clauses. In Kiefer Ferenc, É. Kiss Katalin (eds.), *The syntactic structure of Hungarian*. San Diego: Academic Press. 141-165.

Kibort, Anna. 2007. Extending the applicability of Lexical Mapping Theory. In Butt, Miriam, Tracy H. King (eds.), *Proceedings of the LFG08 Conference*. Stanford: CSLI Publications. 250–270.

Laczkó Tibor. 2012. On the (un)bearable lightness of being an LFG style copula in Hungarian. In Butt, Miriam, Tracy H. King (eds.), *Proceedings of the LFG12 Conference*. Stanford: CSLI Publications. 341-361.

Lødrup, Helge. 2012. In search of a nominal COMP. In Butt, Miriam, Tracy H. King (eds.), *Proceedings of the LFG12 Conference* Stanford: CSLI Publications. 383-404.

Rákosi György, Laczkó Tibor. 2005. Verbal category and nominal function – Evidence from Hungarian Subject Clauses. In Butt, Miriam and Tracy H. King (eds.), *Proceedings of the LFG05 Conference*. Stanford: CSLI Publications. 353-370.

Patejuk, Agnieszka, Adam Przepiórkowski. 2014. Control into selected conjuncts. In Butt, Miriam and Tracy H. King (eds.), *Proceedings of the LFG14 Conference*. Stanford: CSLI Publications. 448-460.

Patejuk & Przepiórkowski. 2016. Reducing grammatical functions in LFG. In Arnold, Doug, Miriam Butt, Berthold Crysmann, Tracy H. 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. 2006. *Dative experiencer predicates in Hungarian*. Doctoral dissertation, University of Utrecht.

Szabolcsi Anna. 2009. Overt nominative subjects in infinitival complements in Hungarian. In den Dikken, Marcel, Robert M. Vago (eds.), *Approaches to Hungarian 11*. John Benjamins. 251-276.

Szécsényi Kriszta, Szécsényi Tibor. 2017. Definiteness agreement in Hungarian multiple infinitival constructions. In Emonds, Joseph, Markéta Janebová (eds.), *Language Use and Linguistic Structure: Proceedings of the Olomouc Linguistics Colloquium*. Olomouc: Palacký University. 75-89.

Szűcs Péter. 2015. On pronouns in Hungarian complex sentences. *Argumentum* 11: 292-313.

Operator Fronting in Hungarian

Péter Szűcs
University of Debrecen

Proceedings of the LFG'18 Conference

University of Vienna

Miriam Butt, Tracy Holloway King (Editors)

2018

CSLI Publications

pages 343–363

http://csli-publications.stanford.edu/LFG/2018

Keywords: operator fronting, prolepsis, control

Szűcs, Péter. 2018. Operator Fronting in Hungarian. In Butt, Miriam, & King, Tracy Holloway (Eds.), *Proceedings of the LFG'18 Conference, University of Vienna*, 343–363. Stanford, CA: CSLI Publications.

Abstract

This paper investigates the Hungarian Operator Fronting construction, where an element from a subordinate sentence appears in the matrix clause. Two subtypes are distinguished: a standard long-distance dependency and a proleptic version. In the latter, the fronted element is analyzed as a thematic object of the main verb. This thematic object is linked to an embedded grammatical function via obligatory anaphoric binding. The configuration can be integrated into a wider perspective of control from an LFG perspective.

1. Introduction[1]

Operator Fronting (OF) in Hungarian is a construction whereby some dependent of a subordinate clause surfaces in the matrix clause. The term itself is aimed to be a theory- and analysis-neutral label and is not to be viewed as a commitment to a particular view of the phenomenon. In the previous literature, the structure has also been called "Focus Raising" (Kenesei 1992, Lipták 1998, Gervain 2009), "Operator Raising" (Gervain 2002) and "Long Operator Movement" (É. Kiss 2002). Here "operator" refers to the fact that the element under question usually bears some discourse function in the main clause while "fronting" captures the noncanonical positioning of this element.

OF is illustrated in (1). (1a) is a standard Hungarian sentence with a subordinate clause, while (1b) is the actual OF example. *János* ('John') is the subject of the embedded clause in both sentences but in (1b) it occurs outside its original place, in the matrix clause preverbal position. Note that in this latter case the fronted constituent bears optional accusative case-marking. Accusative case-marking indicates that the fronted element is OBJ in the main clause. This is related to the possible occurrence of a demonstrative in the standard sentence (1a). These details are going to be important for the forthcoming discussion of the phenomenon.

(1) a. *(Az-t) mondtad, hogy <u>János</u> jön a partira.*
that-ACC said.2SG that(c)[2] John comes the party.onto
'You said that John is coming to the party.'

b. *<u>János(-t)</u> mondtad, hogy jön a partira.*
John(-ACC) said.2SG that(c) comes the party.onto
'(Of) John you said that he is coming to the party.'

[1] The Project no. 111918 (*New approaches in the description of the grammar of Hungarian pronominals)* has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the K funding scheme.

[2] The "c" stands for "complementizer". This is to avoid any confusion with the demonstrative in such sentences. If not indicated otherwise, nominative case and present tense assumed in the glosses.

The focus of my research is to answer the following three questions regarding OF:

i. What is the relationship between the fronted element and the matrix predicate?
ii. What is the relationship between the fronted element and the embedded grammatical function?
iii. How does the construction fit into syntactic theory in general, particularly from the perspective of LFG?

The structure of the paper is as follows. Section 2 will give an empirical overview of OF, with special attention to variation in the phenomenon. Here we will see that there are two separate underlying processes and the one resulting in an accusative marked fronted element in (1b) is the theoretically really interesting one. In section 3 I will outline the possible analyses of OF, contrasting an earlier analysis (Coppock (2003)) with the one that I propose. The crucial difference is that I will argue that in the accusative version of (1b), the fronted element is a thematic OBJ of the main predicate. I will support this claim with evidence from several directions. These sections should answer question i) and ii) above. Section 4 will set out to answer question iii), and argues that OF may be related to control-constructions in a broader sense and fits well into the LFG-typology of such phenomena.

2. Overview of Operator Fronting

It was Gervain (2002) who realized that OF is not a unitary phenomenon but should be divided into two configurations: using Chomskyan terms, a "movement"-like dependency and a "base-generated" one. That is, if the fronted element retains its original case (nominative in 1b), it originates in the embedded clause and is "raised"/ "moved" into the main clause (much like *wh*-dependencies or topicalization), while "case-switch" (to accusative in 1b) indicates that the it fully belongs to the main clause and is related to the embedded GF through coreference. In the discussion below, I will refer to the first type as "LDD-OF" (long-distance dependency), while the second type is going to be "proleptic-OF" (anticipating the analysis put forth in section 3).

Additionally, Gervain (2002) posits that there is a dialectal divide between native speakers regarding the acceptability of the two versions, but subsequent research (Jánosi 2014) cast doubt on this. In this paper I will treat OF as a dialectally uniform phenomenon.

It is common in the two versions of OF that the distance between the fronted element and its embedded correlate may be large.

(2) *János(-t) mondtad, hogy hallottad, hogy jön.*
John(-ACC) said.2SG that(c) heard.2SG that(c) comes
'(Of) John you said that you heard that he is coming to the party.'

Another common property is that the fronted constituent may bear any grammatical function in the embedded clause, as (3) shows with an OBL. This is usually not discussed in the literature for the case of proleptic-OF. In such cases (i.e. when in proleptic-OF, the embedded GF is not a SUBJ), a resumptive pronoun is obligatory (3b). No such pronoun is possible in (3a).

(3) a. *Londonba mondtad, hogy (*oda) mész.*
London.to said.2SG that(c) there goes.2SG

b. *London-t mondtad, hogy oda mész.*
London-ACC said.2SG that(c) there goes.2SG
'To London you said that you are going (there).'

The split between LDD- and proleptic-OF may be observed in a wide variety of syntactic contexts. Here I survey a sample of these contexts, for further contrasts, see Gervain (2002, 2009). The data in (4)-(7) also relies on these works.

LDD-OF but not proleptic-OF shows island-sensitivity:

(4) a. **János mondtad, hogy hallottad a hírt, hogy jön.*
John said.3SG that(c) heard.2SG the news that(c) comes.

b. *János-t mondtad, hogy hallottad a hírt, hogy jön.*
John-ACC said.3SG that(c) heard.2SG the news that(c) comes.
'(Of) John you said that you heard the news that he is coming.'

The interpretation of a reciprocal pronoun may reconstruct into the embedded clause, allowing for the backward binding of the experiencer argument (see Pesetsky 1987) only in the case of LDD-OF.

(5) *Egymás szülei(-*t) mondtad, hogy elszomorították a fiúkat.*
each.other parents(-ACC) said.2SG that(c) saddened.3PL the boys.
'(Of) Each other's parents you said that they saddened the boys.'

The data in (4) and (5) follows directly from an approach where LDD-OF involves a direct link to the embedded position (through functional identification), while in proleptic-OF the fronted constituent fully belongs to the main clause and the link is less direct instead (as we will see, it is an anaphoric binding relationship).

An important difference is that in the case of quantified fronted elements, the embedded verb may show either singular or plural agreement in the case of proleptic-OF (6b). The possibility of the plural is surprising since these phrases trigger singular agreement in unembedded contexts (7).

(6) a. *Az összes fiú mondtad, hogy jön/ *jönnek.*
the every boy said.2SG that(c) come.3SG come.3PL

b. *Az összes fiú-t mondtad, hogy jön/ jönnek.*
the every boy-ACC said.2SG that(c) come.3SG come.3PL
'(Of) every boy you said that they are coming.'

(7) *Az összes fiú jön / *jönnek.*
the every boy come.3SG come.3PL
'Every boy is coming.'

This also makes sense if LDD-OF is a strict syntactic dependency ("movement"), while proleptic-OF exploits a coreference relationship.

Finally, the fronted element must be preverbal in the case of LDD-OF, while it can be postverbal in proleptic-OF (although the latter is somewhat stylistically marked, the contrast in (8) is clear).

(8) *Mondtad János*(-t), hogy jön.*
said.2SG John(-ACC) that(c) comes
'You said of John that he is coming.'

We may get an explanation for this also under the assumption to that the fronted constituent is "extracted" in LDD-OF, and such extractions can only target the preverbal area in Hungarian, while the accusative element is a main clause object, which may be either pre- or postverbal.

The properties and the analytical ideas outlined above will be further elaborated in the next section, where I show a possible analysis of OF.

3. Analyzing of Operator Fronting

As I have already suggested, it is clear that OF may be licensed in two ways: it could be a strict syntactic dependency (LDD-OF) or a coreference relation (proleptic-OF). In the following sections, I outline how these can be captured in the framework of LFG.

3.1 Analyzing LDD-OF

For the purposes of the analysis, I assume a basic phrase structure of Hungarian along the lines of Laczkó (2014), outlined in Figure (1). The sentence is headed by an exocentric S node, which dominates an iterative topic- and quantifier-field, followed by a unique Spec-VP (which hosts focus and various verbal modifiers) and a flat postverbal area. This is of course vastly simplified, a fuller picture would include annotations about information-structure, plus an array of checking equations for proper implementation, see Laczkó (2014).

To analyze LDD-OF, I posit that the phrase-structure rules of Hungarian specify that certain preverbal positions are optionally functionally identified with an embedded GF.

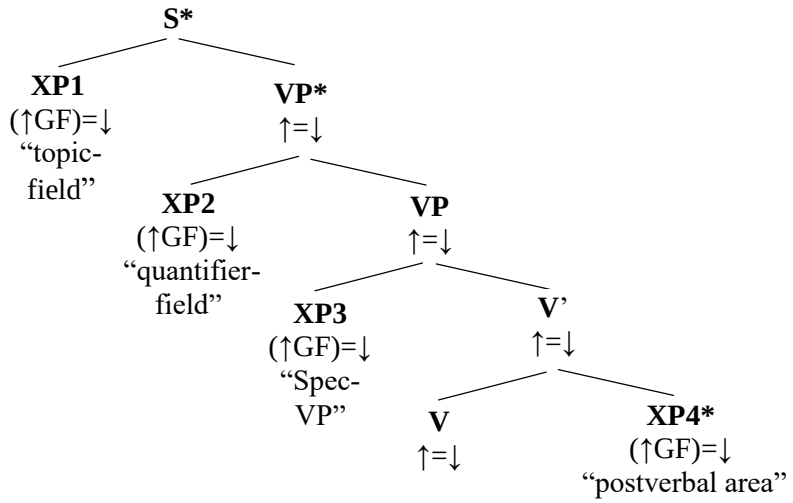


Figure 1.
The basic structure of Hungarian

Operator Fronted elements can occupy either the XP1 or the XP3 position, so these positions have to be supplied with the following annotation:

(9) a. (↑LDD-OF-PATH) = ↓

b. LDD-OF-PATH ≡ AF^{+} GF
(→TENSE)
(→BRIDGE ≠ –)

(9) indicates that the fronted element may be identified with the GF of any grammatical function embedded in a tensed argument function (AF: SUBJ, OBJ, OBL_{θ}. OBJ_{θ}). These tensed argument functions are the subordinate *that*-clauses (the interaction with the associated demonstrative pronouns will be discussed in section 3.2.1 below). The (→BRIDGE ≠ –) constraint is aimed to exclude identification over a nonbridge-verb, as these block this version of Operator Fronting, as is expected in regular long-distance dependencies.

So for instance in (1b) it is ensured by (9) that *János* ('John') is identified as the SUBJ of the subordinate clause. The clause itself is analyzed as the OBJ of the main verb.

Example (10) illustrates a case where the clause is a SUBJ and the fronted element is the OBJ of this SUBJ.

(10) *Jánost tilos, hogy meghívd.*
John.ACC forbidden that(c) invite.SBJV.2SG
'You inviting John is forbidden.'

Also, it is possible that some *that*-clauses in Hungarian are functionally OBLs (see e.g. Szűcs (this volume))[3]. For these, the path would involve this argument function.

While this analysis may notationally be different from those already proposed by others (Lipták 1998, Gervain 2002, Coppock 2003), its basic spirit is the same. This is not so with the analysis of the case-switched, proleptic-OF. In the following sections, I will concentrate on that, showing how my proposed analysis differs from the previous ones, particularly Coppock's (2003) LFG analysis.

3.1 Coppock (2003)

For the version of (1b) where the embedded subject *János* ('John') bears accusative case in the main clause, Coppock (2003) posits that it functions as an athematic OBJ of the main predicate, much like what we see in the so-called "raising to object"-constructions, see (11) and (12) below.

(11) *János-t* *mondtad, hogy* *jön.*
John-ACC said.2SG that(c) comes
'(Of) John you said that he is coming.'

(12) *I believe John to be happy.*

However, unlike "raising to object", the identification of the athematic object in proleptic-OF is not functional, but anaphoric in nature. That is, instead of having strict syntactic identity, the relationship between *Jánost* and the embedded subject is only co-reference. The embedded subject might be pronounced in this scenario, though doing so would be usually dispreferred, Hungarian being a pro-drop language (but see 3b). As indicated in (3), with nonsubject correlates, the pronoun naturally surfaces. Thus for Coppock (2003), (11) has a simplified f-structure like Figure 2. (Information-structure is disregarded.)

$$\begin{bmatrix} \text{PRED} & \textit{mond} \text{ <(SUBJ)(COMP)>(OBJ)} \\ \text{SUBJ} & [\ \text{PRED}\ \textit{pro}\ ] \\ \text{OBJ} & [\ \text{PRED}\ \textit{János}_i\ ] \\ \text{COMP} & \begin{bmatrix} \text{PRED} & \textit{jön} \text{ <(SUBJ)>} \\ \text{SUBJ} & \text{PRED}\ \textit{pro}_i \end{bmatrix} \end{bmatrix}$$

Figure 2.
F-structure for (10) in Coppock's (2003) account.

[3] However, for the purposed of the present paper, I remain conservative about the COMP-debate, I assume that this function is available in the inventory of LFG. Nothing crucial depends on this, and if the references paper is along the right track, some details might be recast in a COMP-less approach as well.

A consequence of the anaphoric link is the agreement variation shown in (6b). This may be explained with reference to an ambiguity between INDEX and pragmatic agreement, following the terminology of Wechsler & Zlatić (2003). In their theory, agreement may be of several types: CONCORD (the syntactic side, usually participating in NP/DP-internal agreement), INDEX (based on the semantic content of the nominal, manifest in e.g. subject-verb agreement, anaphora) and pragmatic agreement (based on conceptual and discourse factors).[4] As we have seen in (7), the default INDEX value for quantified noun-phrases in Hungarian is singular. However, we have here an anaphoric binding relationship between the antecedent *Jánost* ('John') and the embedded *pro*. As Wechsler & Zlatić (2003:84) explain, "this invites the possibility of pragmatic/semantic agreement. Hence the range of options for personal pronoun-antecedent agreement are INDEX agreement and pragmatic/semantic agreement". This is exactly what we see in (6b). Despite the singular index, these quantified nominals are conceptually plural, enabling the possible plural agreement.

In addition, Coppock (2003) claims that the singular-plural dichotomy correlates with the distributive-collective interpretation of the embedded predicate. While this is not true in an absolute sense (Gervain 2002:81 observes plural agreement is still possible with exclusively distributive embedded predicates, see (13), there is a preference for plural agreement to be interpreted as collective, as in (13). (The singular version, as far as I can tell, does not display a preference.)

(13) *Két fiút mondtál, hogy levegőt vettek*
two boys.ACC said.2SG that(c) air.ACC took.3PL
'(Of) two boys you said that they took a big breath.'

(14) *Két fiút mondtál, hogy vittek egy bőröndöt.*
two boys.ACC said.2SG that(c) carried.3PL a suitcase.ACC
'Of two boys you said that they were carrying a suitcase.'
→ collective reading is preferred

The main problem with this account is that it violates LFG's Semantic Coherence condition, as the semantically meaningful element *János* ('John') is not thematically linked to any predicate. Coppock (2003) realizes this and proposes that the Semantic Coherence condition is only an Optimality Theory-style[5] constraint which may be violated in given circumstances. While this is a possible theory, one may wonder if it is really optimal to downgrade a basic LFG principle to save the analysis of a specific construction. An alternative analysis, without the violation, is certainly preferable. I turn to this in the next section.

[4] Also, morphology determines declension.
[5] See e.g. Bresnan (2000).

3.2 The proposed alternative: prolepsis

Semantic Coherence is violated in Figure 2 because there is a meaningful (PRED-bearing) element that is an athematic argument, yet is only anaphorically linked to a semantically selected grammatical function. To avoid this situation one may resort to three paths: a) deny that the element in question is semantically contentful; b) posit that the link is functional; c) argue that the element is a thematic argument of the main predicate. Option a) is trivially ruled out since *János* ('John') (and any other fronted element under discussion) is clearly not an expletive. The facts about overt pronouns and agreement variation are clearly opposed to option b). This leaves us with option c): these fronted elements are the *thematic* objects of the main predicates (I will refer to these as "proleptic objects"). In this approach, the f-structure of (11) looks like Figure 3.

$$\begin{bmatrix} \text{PRED} & \textit{mond} \text{ <(SUBJ)(OBJ)(COMP)>} \\ \text{SUBJ} & [\ \text{PRED}\ \ \textit{pro}\ ] \\ \text{OBJ} & [\ \text{PRED}\ \ \textit{János}_i\ ] \\ \text{COMP} & \begin{bmatrix} \text{PRED} & \textit{jön} \text{ <(SUBJ)>} \\ \text{SUBJ} & \text{PRED}\ \ \textit{pro}_i \end{bmatrix} \end{bmatrix}$$

Figure 3.
The alternative ("proleptic") account.

This is configuration may be labelled as "prolepsis", which in the definition of Salzmann (2017) is a configuration whereby "a structural complement of the matrix verb is semantically related to the predicate of a finite embedded clause". I will elaborate on the place of prolepsis in syntactic theory in section 4, but first let us see what are the reasons for which one may prefer this account over Coppock (2003).

3.2.1 Arguments for a thematic OBJ

The proleptic account avoids the violation of Semantic Coherence, that is an obvious advantage. But as it stands, this is just a technical detour and the advantage melts away if no independent argumentation is put forward for the thematic nature of the OBJ argument. I set out to to provide such argumentation in this section.

One basic question that arises when considering the thematic status of the proleptic object is whether the predicates in question take thematic objects in general. Their availability would increase the plausibility of a thematic analysis. Coppock (2003:136) denies this possibility. While this indeed seems restricted for a verb like *mond* ('say') (but see the discussion below about the status of demonstratives like in sentence (1), other verbs, which may equally

participate in OF, readily allow such objects, as in (15) and (16)[6]. Similar examples may be construed with *fontolgat* ('contemplate'), *állít* ('claim'), *furcsáll* ('find strange'), *jósol* ('predict'), etc.

(15) *Kétlem János hazamenését.*
doubt.1SG John home.going.POSS.ACC
'I doubt John's going home.'

(16) *Jánost kétlem, hogy hazament.*
John.ACC doubt.1SG that(c) home.went.3SG
'(Of) John I doubt that he went home.'

Also, an accusative-marked demonstrative pronoun may occur with every subordinating verb in question, as in (1), repeated here as (17).

(17) *(Az-t) mondtad, hogy János jön a partira.*
that-ACC said.2SG that(c) John comes the party.onto
'You said that John is coming to the party.'

This takes us to the debate in Hungarian syntax about the status of such demonstratives. According to the (more or less) standard view of Kenesei (1994), these pronouns are expletives, so they count as athematic arguments in LFG terms. There is an alternative view, dating back to Tóth (2000), which regards the pronouns as contentful (see also Rákosi & Laczkó 2005). While I do not claim that the debate can be considered settled, there are good reasons to adopt this second view.[7] First, similar pronouns systematically occur not only in structural cases, but also in oblique ones, as in (18), for which an expletive-analysis is not plausible (even Lipták 1998, who otherwise endorses Kenesei's view with regards sentences like (17), refers to such oblique pronouns as "argumental referring words").

(18) *János büszke volt arra, hogy győzött.*
John proud was that.onto that(c) won.3SG
'John was proud that he had won.'

Second, unlike canonical expletives, these pronouns may be associated with discourse functions, e.g. focus in (20).

(19) **Only IT rains.*

[6] (15) also illustrates that proleptic-OF is grammatical with a nonbridge-verb. LDD-OF would be ungrammatical here.

[7] The debate is clearly related to the status and analysis of *wh*-scope marking constructions, as in (i). Though a systematic investigation has not been carried out, the present approach is straightforwardly aligned with the "indirect dependency approach", e.g. Dayal (2000).

(i) *Mit gondolsz, hogy ki jön a partira?*
what.ACC think.2SG that(c) who comes the party.onto
'What do you think, who is coming to the party?'

(20) Csak *AZ-T* *mondtad, hogy* *János* *jön* *a* *partira.*
only that-ACC said.2SG that(c) John comes the party.onto
'You said only that John is coming to the party.'

Third, these pronouns may be coordinated with the undoubtedly thematic objects of the kind shown in (21).

(21) *Kétlem* *János* *hazamenését* *és* *(azt),* *hogy*
doubt.1SG John home.going.POSS.ACC and that.ACC that(c)
nem is *szólt* *róla.*
not even told.3SG about.it
'I doubt John's going home and that he didn't even tell us about it.'

Note that the pronoun is optional. What happens is that the OBJ function of the main predicate may be fulfilled by the pronoun. In this case, the clause itself is an adjunct to it. If there is no pronoun, the clause itself bears the respective GF. This gives the prediction that extraction from the clause should only be possible in the absence of the pronoun (in its presence the clause is subject to the Adjunct Island constraint). This appears to be correct, as the following example shows. (22a) is standard LDD-OF, the complement clause being the OBJ of mond ('say'). In (22b), the OBJ is the pronoun, and clause itself is an ADJUNCT. In this latter case, the "extraction" of the fronted element fails.

(22) a. *János* *mondtad, hogy* *jön.*
John said.2SG that.ACC that(c) comes.
'(Of) John you said that he is coming.'

b. **János azt* *mondtad, hogy* *jön.*
John that.ACC said.2SG that.ACC comes.

Such a pattern actually has already been observed in a number of languages, with similar conclusions, see Bennis (1986) for Dutch and Berman (2001) for German. An example from the latter is shown in (22). *Sagen*'s ('say') OBJ argument may be realized as a clause or as a pronoun in (23a). But as can be seen in (23b), the pronoun-option makes the clause opaque for "extraction", just like in Hungarian.

(23) a. *weil* *er* *(es)* *gesagt* *hat,* *dass* *Hans* *krank* *ist*
because he it said have that(c) Hans ill is
'because he said that Hans is ill'

b **Was* *hat* *er* *es* *gesagt,* *dass* *er* *gelesen* *hat?*
what has he it said that(c) he read has
'What did he say that he read?'

Thus we can conclude that a thematic object for the object argument of these verbs is a possible scenario, be it a pronoun, a regular object or a proleptic one.

More direct evidence for the thematic nature of the object *János* ('John') in (11) is that the construction is incompatible with idiom-chunks, even when they are postverbal (as the preverbal area is associated with discourse-

functions, idiom-chunks are expected to excluded from there, regardless of the thematic status of the argument-slot).

(24) a. *Jánost elkapta a gépszíj.*
John.ACC caught.3SG the driving.belt
'The driving belt caught John.' → 'John has to work a lot.'

b. *Mondtad a gépszíj-at, hogy elkapta Jánost.*
said.2SG the driving.belt-ACC that(c) caught.3SG John.ACC
'(Of) the driving belt you said that it caught John.' (No idiomatic reading is available.)

Finally, an argument may be construed along the lines of Bresnan (1982:71-72). She observes that the complement clauses of equi-type verbs may be elided given the proper context. This is not possible with raising verbs. The reason is that the stranded object is still semantically interpretable in the first case as the main clause object (*John*) gets a thematic role from *persuade* in (25a), but not from *believe* in (25b). So (25a) is formally incomplete (but reconstructable from discourse) but semantically coherent while (25b) is incomplete and incoherent. As can be seen from (25) proleptic-OF patterns with *persuade*, suggesting that it does get a thematic role from the predicate.

(25) a. *Someone had to wash my car. I persuaded John (to wash my car).*

b. *Someone stole my car. I believed John *(to have stolen my car.)*

(26) A: *Szerintem János a legokosabb.*
in.my.opinion John the smartest.
'I think John is the smartest.'

B: *De eddig te Pétert mondtad (hogy ő a legokosabb).*
but so.far you Peter.ACC said.2SG that(c) he the smartest
'But so far you said (of) Peter (that he is the smartest).'

3.2.2 Prolepsis and argument-structure

Now that I have laid down the main ideas of the proleptic analysis, I briefly turn to some details with regards to argument-structure. The basic picture is that the proleptic construction is based on a valency-increasing operation. So the standard lexical entry of a subordinating verb like *mond* ('say') involves two subcategorized GFs, as in (27a), where the agent argument is mapped onto the SUBJ, and the propositional one onto the OBJ. In contrast, the proleptic entry in (27b) has three: the subject (agent), the proleptic object ("subject matter", see Pesetsky 1995) and the clause (proposition). This means that the two OBJs in (27) do not have the same status and the lexical entry in (27a) has undergone a morphosemantic process that rearranged the semantic participants. This may be nicely modelled in Kibort's (2007) conception of LMT, but space-limitations prevent me from demonstrating it here.

(27) a. <(SUBJ)(OBJ)>
b. <(SUBJ)(OBJ)(COMP)>

For this alternation to be possible, the added "subject matter"-argument has to be independently available in the conceptual structure of the predicate. Such elements may appear as optional delative-marked adjuncts. In other words, a morphosemantic process "argumentalizes" a nonargument participant.[8] So for instance *érez* ('feel), which only marginally licenses such a participant, is degraded in proleptic-OF too.

(28) a. *Azt mondtam/??éreztem Jánosról, hogy jön.*
that.ACC said.1SG felt.1SG John.about that(c) comes
'I said/felt of John that he is coming.'
b. *Jánost mondtam/??éreztem, hogy jön.*
John.ACC said.1SG felt.1SG that(c) comes
'Of John I said that he is coming.'

Another thing that happens with the lexical entry is that a referential identity is forced between the proleptic OBJ and some GF of the embedded clause. This is ensured with an annotation like (29).

(29) OBJ INDEX= $COMP^{+}$ GF* INDEX

The discussion so far answers the first two questions raised in the introduction: in proleptic-OF and the fronted element is a thematic argument of the main verb and it is related to the embedded grammatical function via anaphoric binding. Now it is time to turn to the third question: how does this analysis relate to a wider syntactic theory?

4. Prolepsis in syntactic theory

As already mentioned, an informal definition of prolepsis is given by Salzmann (2017): "a structural complement of the matrix verb is semantically related to the predicate of a finite embedded clause". The term itself originates in rhetoric, meaning "anticipation". Its use in linguistics goes back at least to Higgins (1981). More recently, it has been brought into the theoretical limelight by Davies (2005). It may be used as a contrast to raising constructions: despite its surface similarity, it displays the opposite behavior with respect to a number of standard tests (idiom-chunks, meaning in passive, islands, etc.). Proleptic analyses have been put forward in a number of languages. The basic picture is always the same. A main predicate is associated with three arguments: a subject, a proleptic element and a finite complement clause. Also, the proleptic element is referentially identical with some GF embedded in the complement clause.

Some examples for proleptic analyses are shown below. For English in (30) by Massam (1985), or for Madurese by Davies (2005) in (31). Other examples

[8] This likens prolepsis to applicatives, a research avenue yet to be explored.

include German (Salzmann 2017), Greek (Kotzoglou & Papangeli 2007), Japanese, Korean (Yoon 2007).

(30) *I read of Carrol that she was awfully shy.*

(31) *Siti ngera Hasan bari' melle motor.*
Siti think Hasan yesterday buy car
'Yesterday Siti thought about Hasan$_i$ that he$_i$ bought a car.'

(30) and (31) also illustrates a split between two types of proleptic constructions. In (30) the proleptic element (*of Carrol*) is an adjunct PP[9], while in (31) *Hasan* serves as a direct object of the main verb (Hungarian OF fits into this second pattern). This latter seems to be rarer cross-linguistically, which is expected if the "argumentalization" is an additional process, as suggested in the previous section.

Now if we compare the proleptic f-structure in Figure 3 with a standard object-equi sentence, the similarity is obvious.

(32) *I persuaded John to come.*

$$\begin{bmatrix} \text{PRED} & \textit{persuade} \langle (\text{SUBJ})(\text{OBJ})(\text{COMP}) \rangle \\ \text{SUBJ} & [\ \text{PRED}\ \textit{pro}\ ] \\ \text{OBJ} & [\ \text{PRED}\ \textit{John}_i\ ] \\ \text{COMP} & \begin{bmatrix} \text{PRED} & \textit{come} \langle (\text{SUBJ}) \rangle \\ \text{SUBJ} & \text{PRED}\ \textit{pro}_i \end{bmatrix} \end{bmatrix}$$

Figure 3.
F-structure for (32).

In both structures, a matrix-clause thematic dependent in an anaphoric binding relationship with an embedded argument. Note that equi is also standardly contrasted to raising, essentially along the same lines that I have mentioned earlier. Thus, I suggest that prolepsis fits into syntactic theory by virtue of being a kind of control-construction[10]: a finite, anaphoric, equi-type control. The next two sections elaborate on this claim.

[9] Note however that Chomsky (1981) mentions dialectal sentences like (i), where interrogative word seems to receive accusative case from the main predicate.

(i) *Whom do you suggest should become president?*

[10] It is important to note that I use the term "control" in an LFG-sense, which encompasses both raising- and equi-type constructions (Falk 2001:117-118). In both, some sort of identity is established between a main clause- and an embedded clause-dependent. This differs to the GB/MP terminology, where "control" only refers to equi-type constructions. For some details of the raising/equi dichotomy, see section 4.2.

4.1 Prolepsis and equi

First, it must be admitted that for prolepsis to qualify as a control-construction, we must adopt a wider concept of control than is usual. Control standardly refers to constructions where the referential dependency is established with the subject of an immediately embedded, non-finite complement clause. As we have seen, these do not apply to prolepsis, as here the embedded clause is finite, the controlled function may be nonsubject (as in (3b)), and longer paths are also possible (as in (2) and (4b)). (It may be noted that the construction in Greek as described by Kotzoglou & Papangeli 2007 is restricted to immediately embedded subjects, so it conforms to standard control to a larger extent.)

However, if we abstract away from these contextual factors, and focus on the mechanism itself, the parallel is undeniable, which justifies the wider perspective of control. As we will see, there is empirical support for this position and it leads into a typology of control-constructions that is fully consistent with the architecture of LFG.

Just like in standard control, the co-reference of the matrix-dependent and the embedded argument is obligatory. Despite being finite, the subject of the embedded clause in (34) (indicated as *pro* here for convenience) must be *János* ('John'). So, the control-relationship is obligatory.

(33) *I persuaded John$_i$ pro$_{i/*j}$ to come.*

(34) *János-t$_i$ mondtad, hogy pro$_{i/*j}$ jön.*
John-ACC said.2SG that(c) comes
'(Of) John you said that he is coming.'

As summarized in Landau (2013), a core-feature of obligatory control constructions is that they license bound-variable interpretations and sloppy interpretation under ellipsis as in (35)-(38) below. Prolepsis, being an obligatory anaphoric dependency, conforms to these properties of control.

(35) *Only Bill forced himself to jump*. (Bill = Only x [x forced x to jump].)

(36) *Csak János mondta magát, hogy (ő) nyert.*
only John said.2SG himself.ACC that(c) he won.3SG
'Only John said (of) himself that he had won.' (John = Only x [x said of x that x won].)

(37) *Mary encouraged Paul$_i$ PRO$_i$ to attend the ceremony, but not David$_j$ (encourage PRO$_{*i/j}$ to attend the ceremony).*

(38) *Én Jánost$_i$ mondtam, hogy jön, te pedig Pétert$_j$*
I John.ACC said.1SG that(C) comes you but Péter.ACC
(mondtad, hogy PRO$_{*i/j}$ jön).
said.2SG that(C) comes
'(Of) John I said that he comes and you did so of Peter.'

4.2 Prolepsis and the typology of control

As already mentioned in footnote 7, the term "control" here is taken to refer to construction where a referential identity is forced between a matrix-clause and an embedded-clause dependent. Also, these constructions are rooted in the lexical items themselves, rather than being structurally enforced, as in *wh*-type structures.

Control may be classified along several axes. An essential one is the raising-equi opposition. While in the former, there is no thematic relationship between the main predicate and the controller argument, there is one in the latter. That is, as commonly held, John is not a "seemer" in (39a), but he is a "trier" in (39b).

(39) a. *John seemed to win.*
b. *John tried to win.*

Next, in LFG there are two mechanisms that can establish f-structural identities. One is anaphoric control which is semantic co-reference, as in prolepsis. The other is functional control, a strict, syntactic identity, whereby one f-structural element simultaneously satisfies two grammatical functions.

In principle, these two axes may co-vary. However, one constellation is barred by the LFG-architecture: raising with anaphoric identification. The reason is that this leads to the situation that was the original problem with Coppock's (2003) analysis of OF. The matrix argument remains semantically unintegrated: it is linked semantically neither to the main predicate, nor to the embedded one (it is simply co-referent with the controllee, which is independently licensed in the COMP).

Nevertheless, in equi, the bifurcation into a functional and an anaphoric type is a real option. While in Dalrymple (2001) equi is exclusively linked to anaphoric control, Falk (2001) argues that the picture is more complex and some equi verbs use functional control. His case in point is *try* vs. *agree*. According to Falk (2001:136-139), *try* should be analyzed as having a functionally controlled XCOMP complement, while *agree* has an anaphorically controlled COMP. Support for this is available from passivization (39) or the possibility of partial control (41).[11] In (40a), the controller (and thus the controllee) is simply missing. In contrast, discourse control of the embedded subject is possible in (40b), since it is independently licensed as an f-structural *pro* subject. (Essentially, it is a covert pronoun.)

(40) a. **It was tried to finish earlier.*
b. *It was agreed to finish earlier.*

[11] See Haug (2013) for an LFG account of partial control.

In (41) for *try* functional control forces a strict referential identity of the controller and the controllee, the semantic nature of anaphoric control allows for some latitude for *agree*.

(41) a. *John tried to go outside.* (understood subject of *go*: only John)
b. *John agreed to go outside.* (understood subject of *go*: John + possibly other people)

However, while functional control strictly implies exhaustive control, anaphoric control is not necessarily semantically loose. It may be, as in (40), but exhaustively interpreted strict anaphoric control is possible (in fact, prolepsis is such a scenario). I agree with Haug (2013), who suggests that anaphoric control should be viewed as a continuum from obligatory through "quasi-obligatory" to no-control.

CONTROL-TYPE			**Example**
Thematicity of controller	**Nature of identification**	**Finiteness**	
Equi (thematic)	Anaphoric identification	Finite complement	prolepsis
		Non-finite complement	canonical control (*"agree*-type")
	Functional identification	Finite complement	Turkish object control (?), see 4.2.2
		Non-finite complement	canonical control (*"try*-type")
Raising (non-thematic)	Anaphoric identification	Finite complement	not expected
		Non-finite complement	not expected
	Functional identification	Finite complement	Copy Raising/ Hyperraising see 4.2.1
		Non-finite complement	canonical raising

Table 1.
An LFG-typology of control.

Putting all these dimensions together (raising/equi, anaphoric/functional, finite/nonfinite embedded clause), we may arrive at the above typology of control-constructions from an LFG-perspective.[12] In the remainder of the paper, I give a brief overview of those constructions in the taxonomy that have not been mentioned so far.

4.2.1 Finite raising

Although raising, just like equi, is traditionally associated with nonfinite clauses, there are well-described cases of raising from finite clauses. Instances of this with an empty embedded position are labelled Hyperraising, while cases with an overt resumptive pronoun are called Copy Raising. The schematic structure of these are shown in (42).

(42) a. DP_i [IP_{finite} $resumptive_i$] b. DP_i [IP_{finite} ____$_i$]

These two structures are amenable to a parallel analysis, the overtness of the resumptive pronoun seems to boil down to the status of the pro-drop parameter in the given language (see Ademola-Adeoye 2010).

A Copy Raising example is shown in (43), from Igbo (Ura 1998), a Hyperraising one in (44), from Lubukusu (Carstens & Diercks 2013). Space limitations prevent me from elaborating on the details, but the cited works provide evidence from expletives, idioms and scope considerations that these are indeed raising structures.[13]

(43) *Ézè$_i$ di m [kà ọ$_i$ hũrũ Adá]*.
Eze seems to me that(c) he saw Ada
approx.: 'Eze seems to me that he saw Ada.'

(44) *Chisaang'i chilolekhana mbo chikona*
animal seem that(C) sleep
approx.:'The animals seem that they are sleeping.'

4.2.2 Finite equi, functionally identified

Ince (2006) describes a Turkish object-control structure where the embedded clause has both tense and agreement features, thus qualifying as a finite clause. Idiom-chunks are excluded from this sentence-type, so it must be equi. There is no mention of any referential flexibility in the data, so I tentatively categorize this as functional identification. (Further investigation may alter this conclusion.)

---

[12] A possible extension of the typology is to integrate the direction of the control-relation, i.e. forward vs. backward control. For discussion, see Haug (2017).

[13] The proper analysis of the related English sentences like (i) is not settled at this point, see Asudeh & Toivonen (2012) vs. Landau (2011).

(i) *Richard seems like he smokes.*

(45) *Ben Ali-yi yarın balığı yiyecek sanıyordum.*
I Ali-ACC tomorrow fish eat.fut.3SG assumed.1SG
'I thought that Ali will eat the fish tomorrow.'

5. Summary

In this paper I set out to investigate the properties and the analysis of the Hungarian Operator Fronting construction. My main questions concerned the relationship between the fronted element, the main predicate and the embedded correlate as well as the theoretical implications of the phenomenon.

It was established that two subtypes of OF should be distinguished: LDD-OF and proleptic-OF. LDD-OF is a standard long-distance dependency: the fronted element is not related to the main predicate and is linked to its embedded position via a standard functional uncertainty-equation. In proleptic-OF, the fronted constituent becomes the direct thematic object of the main predicate (contra Coppock 2003) and is linked to the embedded GF via obligatory anaphoric binding. This may be regarded as a finite, equi-type anaphoric control relationship, which can be neatly placed in the LFG-conception of control-constructions.

As one of my reviewers noted, Asudeh (2005) already anticipated that the interaction of resumption, raising, control and finiteness constitute a promising avenue for typological research in LFG.[14] This paper may be seen as some fulfillment of this anticipation.

References

Ademola-Adeoye, Feyisayo Fehintola. 2010. *A Cross-linguistic Analysis of Finite Raising Constructions*. Doctoral dissertation, University of KwaZulu-Natal.

Asudeh, Ash. 2005. Control and semantic resource sensitivity. *Journal of Linguistics* 41: 465–511.

Asudeh, Ash, Ida Toivonen. 2012. Copy raising and perception. *Natural Language and Linguistic Theory* 30(2): 321-380.

Bennis, Hans. 1986. *Gaps and dummies*. Dordrecht: Foris.

Berman, Judith. 2001. On the cooccurrence of *es* with finite clause in German: an LFG analysis. In Meurers, Walt D., Kiss Tibor (eds.), *Constraint-Based Approaches to Germanic Syntax*. Stanford: CSLI Publications. 7-30.

Bresnan, Joan (ed.). 1982. *The Mental Representation of Grammatical Relations*. Cambridge, MA: MIT Press.

[14] His theory of control is based on a resource management semantics and arrives at this conclusion: "(…) Finite control can therefore be seen as the control analogue of copy raising. The analysis has typological implications for control mechanisms and extends the relation between control and raising in a new dimension" (Asudeh 2005: 509).

Bresnan, Joan. 2000. Optimal syntax. In Dekkers, Joost, Frank van der Leeuw, Jeroen van de Weije (eds.), *Optimality Theory: Phonology, Syntax and Acquisition*. Oxford University Press. 334–385.

Carstens, Vicki, Michael Diercks. 2013. Parameterizing case and activity: hyper-raising in Bantu. In Kan, Seda, Claire Moore-Cantwell, Robert Staubs (eds.), *Proceedings of the 40th Annual Meeting of the North East Linguistic Society*. Amherst: University of Massachusetts Graduate Linguistic Student Association. 99-118.

Chomsky, Noam (1981). *Lectures on Government and Binding*. Dordrecht: Foris.

Coppock, Elizabeth. 2003. Sometimes it's hard to be coherent. In Butt, Miriam, Tracy H. King (eds.), *Proceedings of the LFG03 Conference.* Stanford: CSLI Publications. 126-143.

Dalrymple, Mary 2001. *Lexical Functional Grammar*. San Diego: Academic Press.

Davies, William D. 2005. Madurese prolepsis and its implications for a typology of raising. *Language* 81: 645–665.

Dayal, Veneeta. 2000. Scope marking: cross-linguistic variation in indirect dependency. In Lutz, Uli, Gereon Müller, Armin von Stechow (eds.), *Wh-scope marking*. John Benjamins Publishing Company. 157-193.

É. Kiss Katalin. 2002. *The Syntax of Hungarian.* Cambridge: Cambridge University Press.

Falk, Yehuda N. 2001. *Lexical-Functional Grammar*. Stanford: CSLI Publications.

Gervain Judit. 2002. *Linguistic Methodology and Microvariation in Language: The Case of Operator-raising in Hungarian*. MA Thesis, University of Szeged.

Gervain Judit. 2009. Resumption in Focus(-raising). *Lingua* 119: 687-707.

Haug, Dag. 2013. Partial control and the semantics of anaphoric control in LFG. In Butt, Miriam, Tracy H. King (eds.), *Proceedings of the LFG13 Conference*. Stanford: CSLI Publications. 274–294.

Haug, Dag. 2017. Backward control in Ancient Greek and Latin participial adjuncts. *Natural Language and Linguistic Theory* 35(1): 99–159.

Higgins, Roger. 1981. Proleptic objects and verbs of perception in Zacapoaxtla Nahuat. *Texas Linguistic Forum* 18: 69-88.

Ince, Atakan. 2006. Direct complement clauses as object control structures in Turkish. In Bainbridge, Erin, Brian Agbayani (eds.), *Proceedings of the thirty-fourth Western Conference On Linguistics*. Department of Linguistics, California State University, Fresno. 208-221.

Jánosi Adrienn. 2013. *Long split focus constructions in Hungarian with a view on speaker variation*. Doctoral dissertation, CRISSP.

Kenesei István. 1994. Subordinate clauses. In Kiefer Ferenc, É. Kiss Katalin (eds.), *The syntactic structure of Hungarian*. San Diego: Academic Press. 141-165.

Kibort, Anna. 2007. Extending the applicability of Lexical Mapping Theory. In Butt, Miriam, Tracy H. King (eds.), *Proceedings of the LFG08 Conference*. Stanford: CSLI Publications. 250–270.

Kotzoglou, George & Dimitra Papangeli. 2007. Not really ECM, not exactly control: The "quasi-ECM" construction in Greek. In Davies, William D., Stanley Dubinsky (eds.), *New Horizons in the Analysis of Control and Raising*. Dordrecht: Springer. 111–131.

Laczkó Tibor. 2014. The essentials of an LFG-analysis of Hungarian finite sentences. In Butt, Miriam, Tracy H. King (eds.), *Proceedings of the LFG14 Conference*. Stanford: CSLI Publications. 325-345.

Landau, Idan. 2011. Predication vs. aboutness in copy raising. *Natural Language and Linguistic Theory* 29: 779-813.

Lipták Anikó. 1998. A magyar fókuszemelések egy minimalista elemzése. In Büky László, Maleczki Márta (eds.), *A mai magyar nyelv leírásának újabb módszerei 3*. Szeged: University of Szeged. 93-116.

Pesetsky, David. 1987. Binding Problems with Experiencer Verbs. *Linguistic Inquiry* 18(1): 126-140.

Pesetsky, David. 1995. *Zero syntax: Experiencers and Cascades*. Cambridge, Massachusetts/London, England: The MIT Press.

Rákosi György, Laczkó Tibor. 2005. Verbal category and nominal function – Evidence from Hungarian Subject Clauses. In Butt, Miriam and Tracy H. King (eds.), *Proceedings of the LFG05 Conference*. Stanford: CSLI Publications. 353-370.

Salzmann, Martin. 2017. Prolepsis. In Everaert, Martin, Henk van Riemsdijk (eds.), *The Wiley-Blackwell Companion to Syntax, 2nd edition*. Wiley-Blackwell. 3203-3245.
http://martinsalzmann.com/files/Salzmann2017_Prolepsis_final.pdf

Tóth Ildikó. 2000. *Inflected Infinitives in Hungarian*. Tilburg: TILDIL Dissertation Series.

Wechsler, Stephen, Larisa Zlatić. 2003. *The Many Faces of Agreement: Morphology, Syntax, Semantics, and Discourse Factors in Serbo-Croatian agreement*. Stanford: CSLI Publications.

Yoon, James H-S. 2007. Raising of major arguments in Korean and Japanese. In Davies, William D., Stanley Dubinsky (eds.), *New Horizons in the Analysis of Control and Raising*. Dordrecht: Springer. 71-107.