Proceedings of the 21st International Conference on Head-Driven Phrase Structure Grammar

University at Buffalo

Stefan Müller (Editor)

2014

Stanford, CA: CSLI Publications

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

Contents

Editor's note **4**

I Contributions to the Main Conference **5**

Abdulrahman A. Alqurashi, Robert D. Borsley: The comparative correlative construction in Modern Standard Arabic **6**

Doug Arnold, Robert D. Borsley: On the analysis of English exhaustive conditionals **27**

Philippa Cook: Between complex predicates and regular phrases: German collocational clusters **48**

Berthold Crysmann, Chris H. Reintges: The polyfunctionality of Coptic Egyptian relative complementisers **63**

Petter Haugereid: VP idioms in Norwegian: A subconstructional approach **83**

David Inman, Ruth Morrison: Negation in Nanti: Syntactic evidence for head and dependent negators **103**

Jean-Pierre Koenig, Karin Michelson: Deconstructing SYNtax **114**

Juwon Lee: Two types of serial verb constructions in Korean: Subject-sharing and index-sharing **135**

Frank Van Eynde, Liesbeth Augustinus: Complement raising, extraction and adposition stranding in Dutch **156**

Olga Zamaraeva, Emily M. Bender: Focus case outside of Austronesian: An analysis of Kolyma Yukaghir **176**

II Contributions to the Workshop **197**

George Aaron Broadwell: Language description and the lexicon: Verbs of wearing in two Oaxacan languages **198**

Pegah Faghiri, Pollet Samvelian, Barbara Hemforth: Accessibility and word order: The case of ditransitive constructions in Persian **217**

Michael Hahn: Predication and NP structure in an omnipredicative language: The case of Khoekhoe **238**

Shrita Hassamal, Anne. Abeillé: Degree adverbs in Mauritian **259**

Dong-yi Lin: Obligatory control and event structure in Kavalan **280**

Dejan Matić, Irina Nikolaeva: Focus feature percolation: Evidence from Tundra Nenets and Tundra Yukaghir **299**

Editor's note

The 21th International Conference on Head-Driven Phrase Structure Grammar (2014) was held at the University at Buffalo, The State University of New York.

The conference featured 2 invited talks, 12 papers, and 2 posters selected by the program committee (Anne Abeillé (chair), Farrel Ackerman, Emily M. Bender, Olivier Bonami, Francis Bond, Robert Borsley, Claire Bowern, George A. Broadwell, Rui P. Chaves, Berthold Crysmann, Elisabet Engdahl, Dan Flickinger, Jeff Good, Fabiola Henri, Jong-Bok Kim, Jean-Pierre Koenig, Valia Kordoni, Robert D. Levine, Robert Malouf, Nurit Melnik, Philip Miller, Stefan Müller, Tsuneko Nakazawa, Joanna Nykiel, Gerald Penn, Adam Przepiórkowski, Frank Richter, Louisa Sadler, Manfred Sailer, Pollet Samvellian, Frank Van Eynde, Robert D. Van Valin Jr., Gert Webelhuth, Stephen Wechsler, Shûichi Yatabe, Eun-Jung Yoo).

A workshop on *Understudied Languages and Syntactic Theory* was attached to the conference. The workshop had three invited speakers and 6 regular papers. The workshop program was put together by Anne Abeillé, Farrel Ackerman, Emily M. Bender, Olivier Bonami, Francis Bond, Robert Borsley, Claire Bowern George A. Broadwell, Rui P. Chaves (chair), Berthold Crysmann, Elisabet Engdahl, Dan Flickinger, Jeff Good, Fabiola Henri, Jong-Bok Kim, Jean-Pierre Koenig, Valia Kordoni, Robert D. Levine, Robert Malouf, Nurit Melnik, Philip Miller, Stefan Müller, Tsuneko Nakazawa, Joanna Nykiel, Gerald Penn, Adam Przepiórkowski, Frank Richter, Louisa Sadler, Manfred Sailer, Pollet Samvellian, Frank Van Eynde, Robert D. Van Valin Jr., Gert Webelhuth, Stephen Wechsler, Shûichi Yatabe, and Eun-Jung Yoo.

We want to thank the respective program committees for putting this nice program together.

Thanks go to Rui P. Chaves and Jean-Pierre Koenig, who were in charge of local arrangements, and their assistants Anastasia Stepanova, Sanghee Lee, and Aron Marvel.

As in the past years the contributions to the conference proceedings are based on the five page abstract that was reviewed by the respective program committees, but there is no additional reviewing of the longer contribution to the proceedings. To ensure easy access and fast publication we have chosen an electronic format.

The proceedings include all the papers except those by Farell Ackerman, Tsuneko Nakazawa, Rui P. Chaves and Jeruen E. Dery, Ray Jackendoff, Cristin Kalinowski and Jeff Good, and Matthew Dryer.

Part I
Contributions to the Main Conference

The comparative correlative construction in Modern Standard Arabic

Abdulrahman A. Alqurashi
King Abdulaziz University

Robert D. Borsley
University of Essex

Proceedings of the 21st International Conference on
Head-Driven Phrase Structure Grammar

University at Buffalo

Stefan Müller (Editor)

2014

Stanford, CA: CSLI Publications

pages 6–26

Alqurashi, Abdulrahman A. & Robert D. Borsley. 2014. The comparative correlative construction in Modern Standard Arabic. In Stefan Müller (ed.), *Proceedings of the 21st International Conference on Head-Driven Phrase Structure Grammar, University at Buffalo*, 6–26. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2014.1.
CC-BY 4.0

Abstract

Much discussion of the comparative correlative construction exemplified by *The more I read, the more I understand* has been concerned with how much cross–linguistic variation there is in this area. Culicover and Jackendoff (1999) suggest that there is considerable variation, but Den Dikken (2005) suggests with data from a variety of languages that the variation is quite limited. Modern Standard Arabic has a comparative correlative construction which is quite different from Engish and the other languages that Den Dikken considers, suggesting that there is more variation in this domain than he assumes. However, it is not difficult to provide an analysis of the construction and other related constructions within the HPSG framework.

1. Introduction

Since Culicover and Jackendoff (1999), the comparative correlative (CC) construction, exemplified by (1), has been an important focus of syntactic research.

(1) The more I read, the more I understand.

A central issue has been how much cross-linguistic variation there is in this area. This is important because the more variation we find, the greater is the challenge for the Chomskyan view that grammatical systems are the result of setting a relatively small number of parameters. Culicover and Jackendoff suggest that languages vary significantly and that they are 'forced to "cobble together" some kind of mechanism to express' the CC meaning (1999: 569). In a reply to Culicover and Jackendoff, Den Dikken (2005) shows that a number of languages have constructions which are broadly similar to the English construction. Among the examples he cites are the following:

(2) a. Naskol'ko luchshe mashina, nastol'ko ona.
by-how-much better car-NOM by-that-much it-F.NOM
dorozhe. (Russian)
more.expensive
'The better the car, the more expensive it is.'

---

† This paper draws in various ways on the first author's MA dissertation, Alqurashi (2008). We are grateful to the reviewers and audience at HPSG21 in Buffalo for their helpful comments and discussion and to Ewa Jaworska for editorial assistance. We alone are responsible for what appears here.

b. Minél többet olvasol, annál többet
what-ADESS more-ACC you.read that-ADESS more-ACC
megértesz. (Hungarian)
VM-you.understand
'The more you read, the more you understand.'

c. Xedii targan max, (bol) tödii
how-much fat meat TOPIC that-much
amttai. (Khalkha Mongolian)
delicious
'The fatter a piece of meat is, the more delicious it is.'

He argues that the CC constructions of various languages have essentially the same structure and are rather like the Hindi relative-correlative construction, exemplified by (3).

(3) jo larRkii khaRii hai vo lambii hai. (Hindi)
REL girl standing is DEM tall is
'The girl that is standing is tall.'

He proposes that both the relative-correlative construction and the CC construction consist of a relative clause – essentially a free relatve – adjoined to a following main clause.

Abeillé and Borsley (2008) note that broadly similar constructions may differ in important ways. They develop this point through a consideration of the French CC construction, where they show that the first clause does not resemble a free relative in any significant way and that for some speakers it is not even a subordinate clause because the two clauses are on a par, as in a coordinate structure.[1]

In this paper, we will show that Modern Standard Arabic (MSA) has a CC construction, which provides further evidence that there is more cross-linguistic variation in this domain than Den Dikken assumes. We will show, however, that it is not difficult to develop a detailed analysis within the Head-driven Phrase Structure Grammar (HPSG) framework, building on the ideas of Borsley (2004, 2011).

Before we can proceeed, we must ask what counts as a CC construction. It is only if we have an answer to this question that we can discuss the viability of Den Dikken's position. A CC construction is not just any construction which can express the CC meaning. In English, the CC meaning can be expressed by the *if-then* and *as-so* constructions. Thus, the following

1 Both clauses of the English CC construction are rather like what Huddleston and Pullum (2002: 761-5, 985-91) call the exhaustive conditional construction, exemplified by (i):

(i) however much I read

This looks like a free relative. However, Huddleston and Pullum argue that exhaustive conditionals are in fact interrogatives.

have more or less the same meaning as (1):

(4) a. If you read more, then you understand more.
b. As you read more, so you understand more.

These constructions, however, can also express other meanings, as the following illustrate:

(5) a. If you read this, then you will understand.
b. As you read this, so you will understand.

What we need, then, is not just a construction which can express the CC meaning but a construction which can only express this meaning. It is entirely possible that some languages do not have such a construction. We will argue, however, that MSA has a CC construction, but one which is very different form the type that Den Dikken focuses on.

The paper is organized as follows. In section 2, we show that MSA has a CC constructon which is quite different from those that Den Dikken discusses but is essentially a specialized version of a fairly ordinary combination of an adjunct cause and a main clause. Then in section 3, we show that MSA has a number of other special constructions which also have related examples in which an adjunct clause combines with an ordinary main clause. In section 4, we develop a fairly detailed analysis of the data within HPSG. Finally, in section 5, we conclude the paper.

2. The MSA construction

Like English, MSA can express the CC meaning with constructions which can also express other meanings. However, as we will see, it also has a construction which can only express the CC meaning. Hence it has a CC construction.

As one might expect, MSA can express the CC meaning with *ʔin* 'if', as in the following:

(6) [ʔin taqraʔ ʔakθar] [taʃhm ʔakθar]
if read.IMPF.2.M.SG more understand IMPF.2.M.SG more
'If you read more, you understand more.'

It can also express the CC meaning with other conditional particles such as *kullamã* 'whenever'.

(7) [kullamã qaraʔta ʔakθar] [tafham
whenever read-PERF.2.M.SG more understand.IMPF.2.M.SG
ʔakθar]
more
'Whenever you read more, you understand more.'

Not surprisingly, both *ʔin* and *kullamã* can also express very different meanings, as the following illustrate:

(8) [ʔin taqraʔ haðaa l-kitab]
if read. IMPF.2.M.SG this DEF-book-ACC
[fa-sa-tafhm ʔal-maqsood]
will-understand IMPF.2.M.SG DEF-idea
'If you read this book, you will understand.'
(9) [kullamã qaraʔta haðaa l-kitab]
whenever read.PERF.2.M.SG this DEF-book-ACC
[tafham ʔal-maqsood]
understand.IMPF.2.M.SG DEF-idea
'Whenever you read this book, you understand the idea.'

ʔin can introduce an imperfective clause, as in (6) and (8) above, or a perfective clause, as in (10):

(10) [ʔin qraʔta ʔakθar] [fahimta ʔakθar]
if read.PERF.2.M.SG more understand PERF.2.M.SG more
'If you read more, you understood more.'

It also allows both a verb-initial clause, as in (6) and (8), and a subject-initial clause, as in (11):

(11) [ʔin Zaid-un yaqraʔ ʔakθar]
if Zaid-NOM read.IMPF.3.M.SG more
[yafhm ʔakθar]
understand IMPF.3.M.SG more
'If Zaid reads more, he understands more.'

In contrast, *kullamã* only introduces clauses which are verb-initial and perfective, hence the ungrammaticality of the following:

(12) *[kullamã taqraʔ ʔakθar]
whenever read-IMPF.2.M.SG more
[tafham ʔakθar]
understand.IMPF.2.M.SG more
'Whenever you read more, you understand more.'

(13) *[*kullamã* Zaid-un yaqraʔ ʔakθar]
whenever Zaid-NOM read.IMPF.3.M.SG more
[yafhm ʔakθar]
understand IMPF.3.M.SG more
'Whenever Zaid reads more, he understands more.'

The main clause which it modifies may be verb-initial or subject-initial and may be perfective or imperfective, as we will show below.

If MSA only had the kinds of example that we have highlighted above, we could conclude that it does not have a CC construction. However, instead of (7), the following is possible:

(14) [kullamã qaraʔta ʔakθar] [kullamã
whenever read.PERF.2.M.SG more whenever
fahimta ʔakθar]
understand.PERF.2.M.SG more
'Whenever you read more, you understood more.'
'The more you read, the more you understood.'

Here, *kullamã* appears not only in the first clause but in the second clause as well. We might translate this in the same way as (7), but it seems equally appropriate to translate it with a CC sentence. It is not possible to have *kullamã* in the second clause with other sorts of meanings. Thus, (15) is not possible as an alternative to (9).

(15) *[kullamã qaraʔta haðaa l-kitab]
whenever read.PERF.2.M.SG this DEF-book-ACC
[kullamã fahimta ʔal-maqsood]
whenever understand.PERF.2.M.SG DEF-idea
'Whenever you read this book, you understood the idea.'

Hence, the double *kullamã* construction can only express the CC meaning. Therefore, it is a CC construction. Unlike the English construction and the other constructions discussed by Den Dikken (2005), it does not have a fronted comparative constituent in either clause. Thus, it is very different from these constructions.

The single *kullamã* construction seems to be a fairly ordinary combination of an adjunct cause and a main clause. As we might expect, the clauses may appear in either order. Thus, (16) is an alternative to (7).

(16) [tafham ʔakθar] [kullamã qaraʔta
understand.IMPF.2.M.SG more whenever read-PERF.2.M.SG
ʔakθar]
more
'You understand more, whenever you read more.'

As we might also expect, the main clause is not required to be imperfective or to be verb-initial, as the following show:

(17) [kullamã qaraʔta ʔakθar] [fahimta
whenever read.PERF.2.M.SG more understand.PERF.2.M.SG
ʔakθar]
more
'Whenever you read more, you understood more.'
(18) [kullamaa qaraʔa Zaid-un ʔakθar] [Amr-un
whenever read.PERF.3.M.SG Zaid-NOM more Amr-NOM
yafhmu ʔakθar]
understand.IMPF.3.M.SG more
'Whenever Zaid reads more, Amr understands more.'

We turn now to the double *kullamã* construction, or the CC-construction, as we will call it from now on. There are a number of points to note. Firstly, neither clause of the construction allows an imperfective verb. Thus, both of the following are ungrammatical:

(19) a. *[kullamã qaraʔta ʔakθar] [kullamã
whenever read.PERF.2.M.SG more whenever
tafham ʔakθar]
understand.IMPF.2.M.SG more
b. *[kullamã taqaraʔ ʔakθar] [kullamã
whenever read.IMPF.2.M.SG more whenever
fahimta ʔakθar]
understand.PERF.2.M.SG more

Secondly, neither clause can appear without the other:

(20) a. *kullamã qaraʔta ʔakθar.
whenever read.PERF.2.M.SG more
b. *kullamã fahimta ʔakθar.
whenever understand.PERF.2.M.SG more

Thirdly, the two clauses have a fixed order. Thus, (21) differs in meaning from (12):

(21) [kullamã fahimta ʔakθar] [kullamã
whenever understand. PERF.2.M.SG more whenever
qara'ta ʔakθar]
read-PERF.2.M.SG more
'The more you understand, the more you read.'

Given that the two clauses have the same form, this is not really surprising.

A final point to note is that while the two clauses of this construction must have a comparative interpretation, they need not contain a comparative word. Thus, as well as examples like (7), we have examples like the following:

(22) [kullamã zaada ħajmu-hu] [kullamã
whenever increase.PERF.3SGM size-its whenever
zaada siʕru-hu]
increase PERF.3SGM price-its
'The more its size increases, the more its price increases.'

This is rather like the main clause in what McCawley (1988) calls the reversed CC construction. The following is a typical example:

(23) I understand more, the more I read.

Here, the second clause, which we assume is an adjunct, looks just like the two clauses of the English CC-construction. The first clause, which we assume is a main clause, has an in-situ comparative word. However, as McCawley notes, it is also possible to have main clauses with no comparative word but with a comparative interpretation. (24) illustrates:

(24) My knowledge increases, the more I read.

We assume that the two clauses of the MSA CC-construction are subject to the same constraint as the main clause of this construction

It seems, then, that MSA has a number of ways of expressing the CC meaning. Some involve constructions which can also express other meanings, but one involving two clauses introduced by *kullamã* can only express the CC meaning. This, then, is a CC construction and one that is very different from the constructions that are the focus of Den Dikken (2005).

3. Other constructions

The MSA CC construction is a specialized construction, but, as we have seen, it is related to a fairly ordinary main clause + adjunct clause structure. This is quite like the situation in English, where the reversed CC

construction, exemplified by (23) and (24) above, is a fairly ordinary main clause + adjunct clause structure (Borsley 2004, 2011). In this section, we will show that the MSA CC construction is one of a number of specialized constructions, each of which is related to an ordinary main clause + adjunct clause structure. Again this is rather like English. Following Borsley (2004, 2011), we will call the specialized constructions correlative clauses.

In English a correlative clause which is rather like the CC construction is the *if-then* construction, illustrated in (4a) above. MSA has two constructions which resemble the *if-then* construction. These are what we will call the *ʔiðaa-fa* construction, exemplified by (25), and the *law-la* construction, exemplified by (26).

(25) [ʔiðaa qaraʔta ʔakθar]
if read-PERF.2.M.SG more
[fa-sa-tafhamu ʔakθar]
then-will-understand.IMPF.2.M.SG more
'If you read more, then you will understand more.'

(26) [law qaraʔta ʔakθar] [la-fahimta
if read-PERF.2.M.SG more then-understand.PERF.2.M.SG
ʔakθar]
more
'If you read more, then you will understand more.'

MSA has at least two further correlative clauses. The first, which we will call the *bimaa-ʔiðann* construction, is exemplified by (27).

(27) [bimaa ʔannka taqraʔu ʔakθar] [ʔiðann
as/since COMP.2.M.SG read-IMPF.2.M.SG more so
sa-tafhamu ʔakθar]
ill-understand.IMPF.2.M.SG more
'As/since you read more, so you will understand more.'

This is rather like the English *as-so* construction, illustrated in (4b). Note that *bimaa* is followed by another complamentizer. We assume this means that it takes a CP complement. The second, which we will call the *biqadri-maa-biqadri-maa* construction, is exemplified by (28).

(28) [biqadri-maa taqraʔ] [biqadri-maa
as-much-as read-IMPF.2.M.SG as-much-as
tafham]
understand.IMPF.2.M.SG
'As much as you read, so much you understand.'

In all four constructions, neither clause can appear without the other:

(29) a. *ʔiðaa qaraʔta ʔakθar.
if read-PERF.2.M.SG more
b. *fa-sa-tafhamu ʔakθar
then-will-understand.IMPF.2.M.SG more
(30) a. *law qaraʔta ʔakθar.
if read-PERF.2.M.SG more
b. *la-fahimta ʔakθar.
then-understand.PERF.2.M.SG more
(31) a. *bimaa ʔannka taqraʔu ʔakθar.
as/since COMP.2.M.SG read-IMPF.2.M.SG more
b. *ʔiðann sa-tafhamu ʔakθar.
so will-understand.IMPF.2.M.SG more
(32) a. *biqadri-maa taqraʔ.
as-much-as read-IMPF.2.M.SG
b. *biqadri-maa tafham.
as-much-as understand.IMPF.2.M.SG

In all four, the order of the clauses is fixed. Thus, (33)–(35) are ungrammatical, and (36) differs in meaning from (28).

(33) *[fa-sa-tafhamu ʔakθar] [ʔiðaa
then-will-understand.IMPF.2.M.SG more if
qaraʔta ʔakθar]
read-PERF.2.M.SG more
(34) *[la-fahimta ʔakθar] [law
then-understand.PERF.2.M.SG more if
qaraʔta ʔakθar]
read-PERF.2.M.SG more
(35) *[[ʔiðann sa-tafhamu ʔakθar] [bimaa
so will-understand.IMPF.2.M.SG more as/since
ʔannaka taqraʔu ʔakθar]
COMP.2.M.SG read-IMPF.2.M.SG more
(36) [biqadri-maa tafham] [biqadri-maa
as-much-as understand.IMPF.2.M.SG as-much-as
taqraʔ]
read-IMPF.2.M.SG
'As much as you understand, so much you read.'

Like the CC construction, all four constructions have related examples where an adjunct clause with some distinctive form modifies an ordinary main clause:

(37) [ʔiðaa qaraʔta ʔakθar] [sa-tafhamu
if read-PERF.2.M.SG more will-understand.IMPF.2.M.SG
ʔakθar]
more
'If you read more, you will understand more.'

(38) [law qaraʔta ʔakθar] [tafhamu
if read-PERF.2.M.SG more understand.IMPF.2.M.SG
ʔakθar]
more
'If you read more, you will understand more.'

(39) [bimaa ʔannaka taqraʔu ʔakθar]
as/since COMP 2.M.SG read-IMPF.2.M.SG more
[sa-tafhamu ʔakθar]
will-understand.IMPF.2.M.SG more
'As/since you read more, you will understand more.'

(40) [biqadri-maa taqraʔ] [tafhamu]
as much as read-IMPF.2.M.SG understand.IMPF.2.M.SG
'As much as you read, you understand.'

With these examples the two clauses can appear in either order:

(41) [sa-tafhamu ʔakθar] [ʔiðaa qaraʔta
will-understand.IMPF.2.M.SG more if read-PERF.2.M.SG
ʔakθar]
more
'You will understand more if you read more books.'

(42) [tafhamu ʔakθar] [law qaraʔta ʔakθar].
understand.IMPF.2.M.SG more if read-PERF.2.M.SG more
'You understand more if you read more.'

(43) [sa-tafhamu ʔakθar] [bimaa ʔannaka
will-understand.IMPF.2.M.SG more as/since COMP 2.M.SG
taqraʔu ʔakθar]
read-IMPF.2.M.SG more
'You will understand more as/since you read more.'

(44) [tafhamu] [biqadri-maa taqraʔ].
understand.IMPF.2.M.SG as-much-as read-IMPF.2.M.SG
'You understand as much as you read.'

It seems, then, that the CC construction is one of a number of special constructions, which we call correlative clauses. In each case, the component clauses have a distinctive form, appear in a fixed order, and neither can appear without the other. Also in each case, we have related examples in which an adjunct clause combines with an ordinary main clause. Thus, we have the following situation:

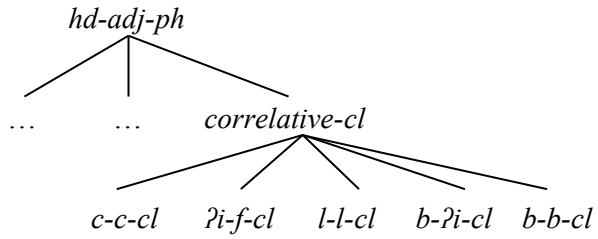
Correlative clause	**Main clause + adjunct clause**
CC construction	Main clause + *kullamã*-clause
ʔiðaa-fa construction	Main clause + *ʔiðaa*-clause
bimaa-ʔiðann construction	Main clause + *bimaa*-clause
biqadri-maa-biqadri-maa construction	Main clause + *biqadri ma*-clause

4. Analyses

We will now develop a fairly detailed analysis of the data within HPSG, adopting essentially the version of HPSG outlined in Ginzburg and Sag (2000).

Following Borsley (2004, 2011), we assume that correlative clauses are special head–adjunct–phrases, where the head has a special feature specification reflected in its distinctive form, as a result of which it cannot appear without the adjunct. We assume the following system of types:

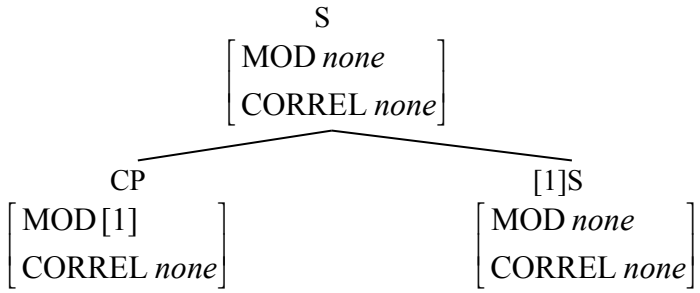
(45)



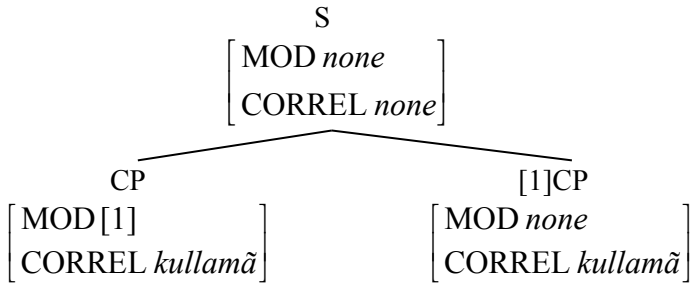
We also assume that *kullamã* and the other clause-initial elements in correlative clauses are complementizers and that they are identified by a feature CORREL(ATIVE). All other words will be [CORREL *none*], including *kullamã* in the single *kullamã* construction.

Given these assumptions, ordinary combinations of an adjunct clause and a main clause involve a CP modifying an S, as in (46), and correlative clauses involve a CP modifying a CP, and structures like (47).

(46)



(47)



The single *kullamã* construction can be analyzed in essentially the same way as other combinations of an adjunct clause and a main clause. The CC construction is a more complex matter, but we will show that it is not too difficult to provide an analysis within HPSG. We will also outline analyses for the other correlative clauses.

For *kullamã* in the single *kullamã* construction, we propose the following syntactic and semantic properties (where we use [INV +] to identify verb-initial clauses and indicate the meaning informally with 'whenever'):[2]

(48)

$$\begin{bmatrix} \text{CAT} \begin{bmatrix} \text{HEAD} \begin{bmatrix} \textit{comp} \\ \text{CORREL } \textit{none} \\ \text{MOD S} \end{bmatrix} \\ \text{SUBJ} \langle\rangle \\ \text{COMPS} \langle \text{S[ASPECT } \textit{perf}, \text{INV} + \text{]} \rangle \end{bmatrix} \\ \text{CONT 'whenever'} \end{bmatrix}$$

For head-adjunct-phrases, we assume the fairly standard constraint in (49).

(49) *hd-adj-ph* $\rightarrow$ $\begin{bmatrix} \text{DTRS} \langle [1][\text{SS}[2]], [\text{HEAD}[\text{MOD}[2]]] \rangle \\ \text{HD - DTR}[1] \end{bmatrix}$

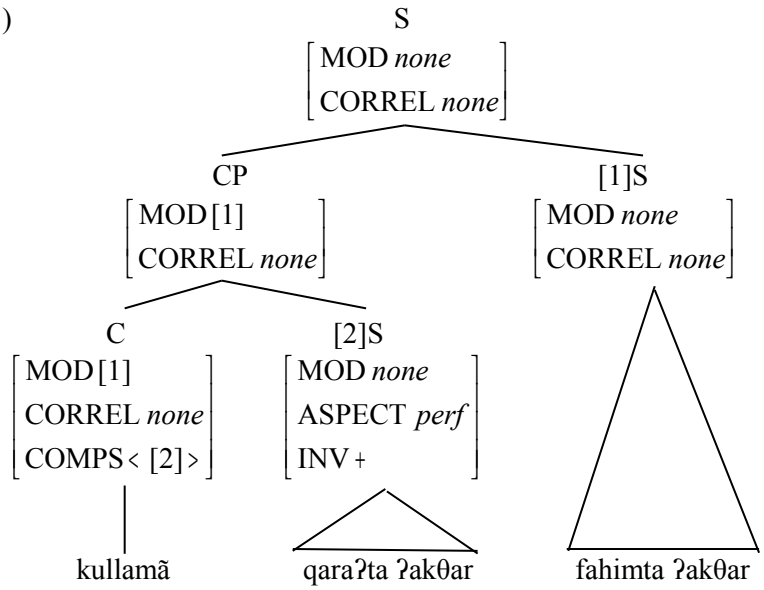
We also assume Ginzburg and Sag's Generalized Head Feature Principle (GHFP), which we can formulate as follows:

(50) *hd-ph* $\rightarrow$ $\begin{bmatrix} \text{SYNSEM } /[1] \\ \text{HD - DTR } [\text{SYNSEM } /[1]] \end{bmatrix}$

2 All complementizers will be [HEAD *comp*] and [SUBJ <>], so this information doesn't need to be included in the description of any specific complementizer.

This is a default statement, as indicated by the slash notation. It requires a headed phrase and its head–daughter to have the same syntactic and semantic properties unless some other constraint requires a difference. In the case of ordinary head–adjunct–phrases, it ensures that the phrase has the same category as its head. Given this machinery, (7) will have an analysis which we can represent as follows:

(51)



The other main clause + adjunct clause structures will have similar analyses. They just need appropriate syntactic and semantic properties. For *ʔin* we can propose the following:

(52)

$$\begin{bmatrix} \text{CAT} \begin{bmatrix} \text{HEAD} \begin{bmatrix} \textit{comp} \\ \text{CORREL } \textit{none} \\ \text{MOD S} \end{bmatrix} \\ \text{SUBJ} \langle\rangle \\ \text{COMPS} \langle \text{S} \rangle \end{bmatrix} \\ \text{CONT 'if'} \end{bmatrix}$$

This is like (48) except that it has a different CONTENT value and no restrictions are placed on the type of S that can appear as its complement. It will give a structure much like (51) for (6). The examples in (37)–(40) will

have similar structures.

We turn now to the rather more challenging CC construction. We will first introduce the necessary constraints and then provide syntactic and semantic properties for the two instances of *kullamã*. For correlative clauses, we assume the following constraints :

(53) *correlative-cl* $\rightarrow$ $\begin{bmatrix} \text{HEAD} \begin{bmatrix} v \\ \text{MOD } none \end{bmatrix} \\ \text{CORREL } none \end{bmatrix}$

(54) *correlative-cl* $\rightarrow$ $\begin{bmatrix} \text{PHON } [1] \oplus [2] \\ \text{DTRS} < [\text{PHON } [2]], [\text{PHON } [1]> \end{bmatrix}$

The first overrrides the GHFP and requires correlative clauses to be verbal, to be [MOD *none*], and to be [CORREL *none*]. (It may be that the last of these stipulations is unnecessary since it is probable that all head–adjunct–phrases are [CORREL *none*].) The second requires the first member of the daughters list, which given (49) is the head, to be second in the phonology. It accounts for the fact that all correlative clauses have a fixed order. For c-c-clauses, we propose the following constraint:

(55) *c-c-cl* $\rightarrow$ [DTRS <[CORREL *kullamã*], [CORREL *kullamã*]>]

This ensures that the two daughters in a c-c-clause are [CORREL *kullamã*].

We now need syntactic and semantic properties for the two instances of *kullamã* that appear in the CC construction. Unlike the *kullamã* of the single *kullamã* construction, both must be [CORREL *kullamã*]. They also need to ensure that their complement has an implicit comparison interpretation. They will differ, however, in two ways. In the adjunct clause, *kullamã* must be [MOD CP], whereas in the main clause it must be [MOD *none*]. We will also assume that *kullamã* in the adjunct clause has the same 'whenever' interpretation as *kullamã* in the single *kullamã* construction, whereas *kullamã* in the main clause is meaningless, having the same interpretation as its complement. This will ensure that the CC construction has essentially the same interpretation as the single *kullamã* construction. It seems, then, that we need the following syntactic and semantic properties, where we represent the fact that the complement must be comparative with the informal CONT value 'comparative':

(56)

$$\begin{bmatrix} \text{CAT} \begin{bmatrix} \text{HEAD} \begin{bmatrix} \textit{comp} \\ \text{CORREL } \textit{kullamã} \\ \text{MOD CP} \end{bmatrix} \\ \text{SUBJ} \langle\rangle \\ \text{COMPS} \langle \text{S[ASPECT } \textit{perf}, \text{INV} +, \text{CONT 'comparative'}] \rangle \end{bmatrix} \\ \text{CONT 'whenever'} \end{bmatrix}$$

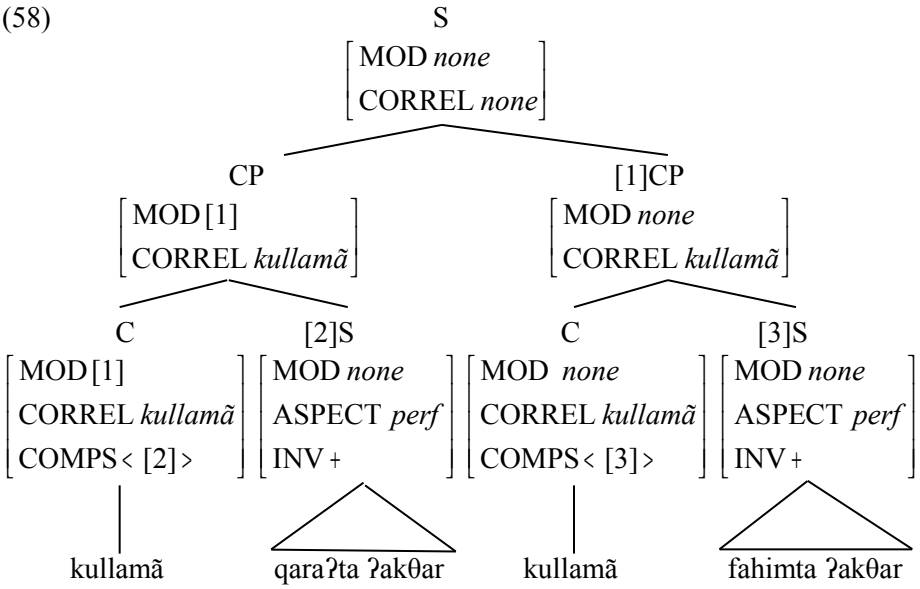
(57)

$$\begin{bmatrix} \text{CAT} \begin{bmatrix} \text{HEAD} \begin{bmatrix} \textit{comp} \\ \text{CORREL } \textit{kullamã} \\ \text{MOD } \textit{none} \end{bmatrix} \\ \text{SUBJ} \langle\rangle \\ \text{COMPS} \langle \text{S[ASPECT } \textit{perf}, \text{INV} +, \text{CONT [1]'comparative'}] \rangle \end{bmatrix} \\ \text{CONT [1]} \end{bmatrix}$$

Both are [CORREL *kullamã*], and both select a complement which is perfective and verb-initial and has an implicit comparative interpretation. They differ in that the first is [MOD CP] whereas the second is [MOD *none*], and the first has the same CONTENT value as *kullamã* in the single *kullamã* construction whereas the second has the same CONTENT value as its complement and hence is meaningless.

With the constraints and lexical properties set out above, we have the following structure for the CC construction in (14):

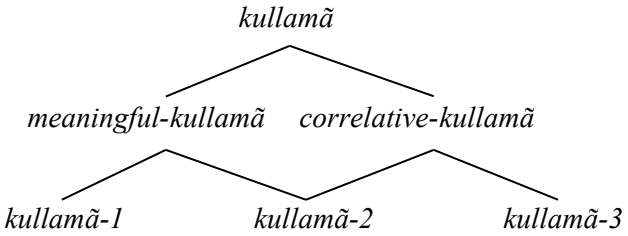
(58)



We now have analyses for both the single *kullamã* construction and the CC-construction. But we need to say more about *kullamã*. We have three separate sets of properties, (48), (56), and (57). They differ in important ways, but they also show some important similarities. All three are complementizers selecting a clausal complement which is perfective and verb-initial. The descriptions in (48), (56) have the same CONTENT value, and (56) and (57) have the same value for CORREL and require their complement to have an implicit comparison interpretation. We can capture these similarities with a system of lexical types.

We propose the following system, where *kullamã-1* is *kullamã* in the single *kullamã* construction, *kullamã-2* is first *kullamã* in a c-c-clause, and *kullamã-2* is second *kullamã* in a c-c-clause:

(59)



These are subject to the following constraints:

(60) *kullamã* → $\begin{bmatrix} \text{PHON } \textit{kullamã} \\ \text{SS|LOC|CAT} \begin{bmatrix} \text{HEAD } \textit{comp} \\ \text{SUBJ} < > \\ \text{COMPS} < \text{S[ASPECT } \textit{perf}, \text{INV} +] > \end{bmatrix} \end{bmatrix}$

(61) *meaningful-kullamã* → $[\text{SS|LOC}[\text{CONT 'whenever'}]]$

(62) *correlative-kullamã* →

$$\left[\text{SS|LOC}\left[\text{CAT}\begin{bmatrix}\text{HEAD}[\text{CORREL } \textit{kullamã}] \\ \text{COMPS} < [\text{CONT}[1]\text{'imp - comp'}] >\end{bmatrix}\right]\right]$$

(63) *kullamã-1* → $\left[\text{SS|LOC}\left[\text{CAT}\left[\text{HEAD}\begin{bmatrix}\text{CORREL } \textit{none} \\ \text{MOD S}\end{bmatrix} >\right]\right]\right]$

(64) *kullamã-2* → $[\text{SS|LOC}[\text{CAT}[\text{HEAD}[\text{MOD CP}] >]]]$

(65) *kullamã-3* → $\left[\text{SS|LOC}\begin{bmatrix}\text{CAT}\begin{bmatrix}\text{HEAD}[\text{MOD } \textit{none}] \\ \text{COMPS} < [\text{CONT}[1]] >\end{bmatrix} \\ \text{CONT}[1]\end{bmatrix}\right]$

The description in (48) combines the properties in (60), (61) and (63). The description in (56) combines those in (60), (61), (62) and (64). Finally, the description in (57) combines the properties in (60), (62) and (65). With this system of types and constraints, we capture the similarities among the three elements.

We turn now to the other correlative clauses highlighted in section 3. It is not difficult to extend the approach developed above to accommodate them. First we need the following constraints on the relevant phrase types to ensure that the right complementizers appear:

(66) *ʔi-f-cl* → [DTRS <[CORREL *fa*], [CORREL *ʔiðaa*]>]

(67) *l-l-cl* → [DTRS <[CORREL *la*], [CORREL *law*]>]

(68) *b-ʔi-cl* → [DTRS <[CORREL *ʔiðann*], [CORREL *bimaa*]>]

(69) *b-b-cl* → [DTRS <[CORREL *biqadri-maa*],
[CORREL *biqadri-maa*]>]

Then we need lexical descriptions for the complementizers. In the case of *bimaa* and *ʔiðann*, we propose the following (ignoring semantics):[3]

(70)

$$\begin{bmatrix} \text{PHON } \textit{bimaa} \\ \text{SYNSEM|LOCAL} \begin{bmatrix} \text{CAT} \begin{bmatrix} \text{HEAD} \begin{bmatrix} \textit{comp} \\ \text{CORREL } \textit{bimaa} \\ \text{MOD CP} \end{bmatrix} \\ \text{SUBJ} \langle\rangle \\ \text{COMPS} \langle \text{CP[FORM } \textit{ʔanna}] \rangle \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

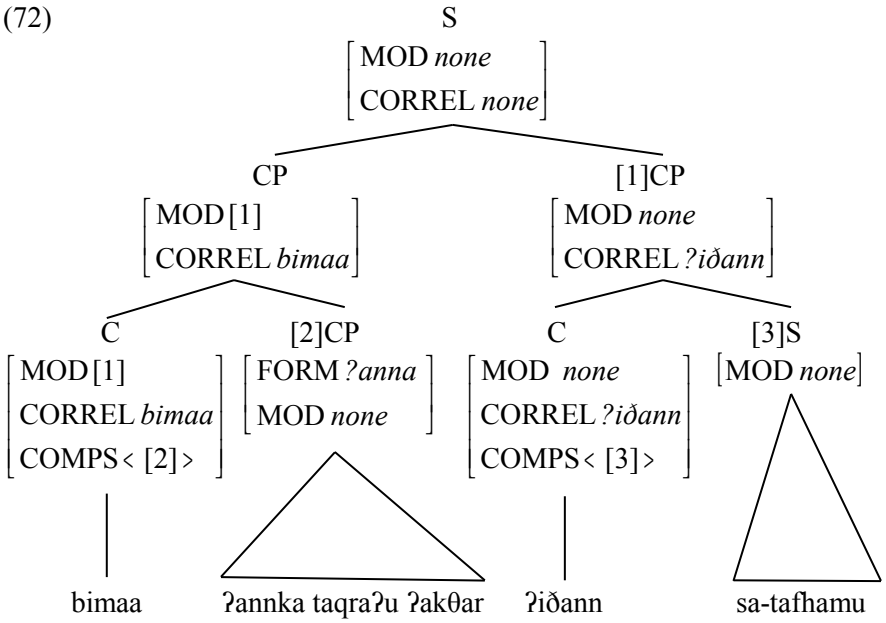
(71)

$$\begin{bmatrix} \text{PHON } \textit{ʔiðann} \\ \text{SYNSEM|LOCAL} \begin{bmatrix} \text{CAT} \begin{bmatrix} \text{HEAD} \begin{bmatrix} \textit{comp} \\ \text{CORREL } \textit{ʔiðann} \\ \text{MOD } \textit{none} \end{bmatrix} \\ \text{SUBJ} \langle\rangle \\ \text{COMPS} \langle \text{S} \rangle \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

These give the following structure for the *bimaa-ʔiðann* construction in (27):

3 *ʔiðann* allows both a verb-initial and a subject-initial complement but requires its complement to be future tense. We ignore this in (71).

(72)



Apart from the fact that *bimaa* takes a CP complement, this is very similar to the representation for the CC construction in (58). The other correlative clauses will have similar structures.

5. Conclusion

We have argued in this paper that MSA has a CC construction which is very different from the English CC construction and the other CC constructions discussed in Den Dikken (2005). We have shown that both the rather unusual CC construction of MSA and the various related constructions are unproblematic for HPSG and we have outlined detailed analyses, drawing on the approach to such constructions developed in Borsley (2004, 2011).

There is one final point that we should make here. Although we have emphasized that the MSA CC construction is quite different from those which Den Dikken focused on, we do not want to suggest that anything goes in this area. We have shown that the MSA CC construction is quite similar to a number of other MSA correlative clauses. In English and other languages, the CC construction seems to be a rather specialized correlative clause. In MSA, it is rather ordinary example of such a clause. However, the fact remains that we have evidence here that there is more variation in this area than Den Dikken (2005) assumed, and hence an important challenge for the view that grammatical systems are the result of setting a relatively small number of parameters.

REFERENCES

Abeillé, A. and R. D. Borsley (2008), Comparative correlatives and parameters, *Lingua* 118, 1139–1157.

Alqurashi, A. A. (2008), Comparative correlatives in Arabic, MA dissertation, University of Essex.

Borsley, R. D. (2004), An approach to English comparative correlatives, in S. Müller (ed.), *Proceedings of the 11th International Conference on Head-Driven Phrase Structure Grammar*, Stanford: CSLI Publications, 70-92.

Borsley, R. D. (2011), Constructions, functional heads and comparative correlatives, in O. Bonami and P. Cabredo Hofherr (eds.), *Empirical Issues in Syntax and Semantics* 8, 7–26.

Culicover, P. W. and R. Jackendoff (1999), The view from the periphery: The English comparative correlative, *Linguistic Inquiry* 30, 543–571.

Den Dikken, M. (2005), Comparative correlatives comparatively, *Linguistic Inquiry* 36, 497–532.

Ginzburg, J. and I. A. Sag (2000), *Interrogative Investigations: The Form, Meaning and Use of English Interrogatives*, Stanford: CSLI Publications.

Huddleston, R. and G. K. Pullum (2002), *The Cambridge Grammar of the English Language*, Cambridge: Cambridge University Press.

McCawley, J. D. (1988), The comparative conditional construction in English, German and Chinese, *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society*, 176–187.

On the analysis of English exhaustive conditionals

Doug Arnold
University of Essex

Robert D. Borsley
University of Essex

Proceedings of the 21st International Conference on
Head-Driven Phrase Structure Grammar

University at Buffalo

Stefan Müller (Editor)

2014

Stanford, CA: CSLI Publications

pages 27–47

Arnold, Doug & Robert D. Borsley. 2014. On the analysis of English exhaustive conditionals. In Stefan Müller (ed.), *Proceedings of the 21st International Conference on Head-Driven Phrase Structure Grammar, University at Buffalo*, 27–47. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2014.2. CC-BY 4.0

Abstract

So-called 'Exhaustive Conditionals' (ECs, also known as 'Unconditionals') have been an important focus of recent research. We develop an HPSG analysis of governed ECs (e.g. 'no matter how intelligent the students are...'), sketch an approach to ungoverned ECs (e.g. 'however intelligent the students are...'), and evaluate three possible analyses of reduced ECs (e.g. 'no matter how intelligent the students...', 'however intelligent the students...').

1 Introduction

Free relatives such as the emphasised part of (1) have had considerable attention within syntactic theory, including HPSG (e.g. Müller, 1999; Kubota, 2003). The superficially similar construction in (2) has had rather less.

(1) They will do *whatever you do*.
(2) They will do that *whatever you do*.

The free relative in (1) is an argument whereas the construction in (2) is an adjunct. Some have supposed that this is the only difference – that the construction in (2) is just an adjunct free relative (e.g. Abeillé and Borsley, 2008). But it is clear that we have a rather different construction in (2). The free relative in (1) can be paraphrased with *any* but not with *no matter*:

(3) They will do *anything you do*.
(4) *They will do *no matter what you do*.

The opposite is true with the construction in (2):

(5) *They will do that *anything you do*.
(6) They will do that *no matter what you do*.

Following Huddleston and Pullum (2002: 761-5, 985-91) (henceforth H&P), we refer to the construction in (2) and its paraphrase with *no matter* in (6) as *exhaustive conditionals* (henceforth ECs). They have also been called 'unconditionals' (Zaefferer, 1990; Rawlins, 2013, 2008).[1]

A further type of EC is exemplified by (7) and (8) – we will call these '*or* ECs':

(7) They will do that *(no matter) whether it's essential or not*.

[†]We are grateful to many colleagues for helpful discussion, notably several anonymous referees for, and participants at, HPSG21 in Buffalo. Remaining flaws are purely our fault.

[1]As noted, ECs are always adjuncts. Free relatives are often arguments, but they can also be adjuncts. This can lead to ambiguity; e.g. *They will be there whenever you are there.* can be understood as either a free relative ('They will be there all the time you are') or an EC ('They will be there no matter when you are there').

(8) Kim will have fun *(no matter) whether he goes to Wales or to Scotland*.

H&P call ECs involving '*wh-ever*' words, as in (2), 'ungoverned ECs', in contrast to 'governed ECs', like (6) and similar examples with *irrespective* and *regardless*, as in (9) and (10). As well as this terminology, we will sometimes talk about *no matter* ECs and *wh-ever* ECs.

(9) They will do that *regardless of what you do*.
(10) They will do that *irrespective of what you do*.

Both *no matter* and *or* ECs look like interrogatives, and despite the superficial resemblance to free relatives, H&P and Rawlins (2008, 2013) argue that *wh-ever* ECs are also interrogatives. H&P (p.989) note that *wh-ever* ECs are like interrogatives in allowing the *wh*-element to be modified by *the hell*:

(11) We must be attractive, $\left\{ \begin{array}{c} \text{whatever the hell} \\ \text{no matter what the hell} \end{array} \right\}$ that means.

Free relatives do not allow this:

(12) *Whoever the hell said that was wrong.

They also note that ECs, like interrogatives, allow multiple *wh*-elements:

(13) $\left\{ \begin{array}{c} \text{Whoever} \\ \text{No matter who} \end{array} \right\}$ said what to whom we must move on.

This is not possible with free relatives:

(14) *Whoever said what to whom is going to be severely dealt with.

Similarly, Rawlins (2013, 148-9) notes that the *What was X doing Y* idiom (with the interpretation of 'why') appears in interrogatives and ECs but not free relatives:

(15) Whatever they were doing reading her mail, it didn't lead to any legal problems.

(16) *She didn't worry about whatever they were doing reading her mail.

So, in (15) the idiomatic interpretation of 'regardless of the reason why they were reading her mail' is available. In contrast, (16) cannot be interpreted as 'she did not worry about why they were reading her mail'.[2]

Thus, there seems to be quite strong evidence that ECs are interrogatives, *wh*-interrogatives in the case of *wh-ever* ECs, disjunctive interrogatives in the case of *or* ECs, and most kinds of interrogatives in the case of *no matter* and other governed ECs.

As regards polar interrogatives, governed ECs are possible without obvious restrictions, e.g. (17) is a simple polar interrogative, (18) and (19) are

[2]Though of course (16) has a perfectly good non-idiomatic interpretation involving a free relative, where *whatever they were doing reading her mail* is interpreted as meaning 'the thing (whatever it was) they were doing'.

alternative polar interrogatives:

(17) We will do it, no matter whether the staff complain.
(18) I'll manage, (no matter) whether you help or you do nothing.
(19) I'll manage, (no matter) whether you help or not.

Alternative polar interrogatives can be governed or ungoverned ECs, but there is a restriction that bare polar interrogatives cannot function as ECs, so *no matter* is obligatory in (17).

Similarly, as regards constituent questions, governed ECs seem to exhibit the full range of possibilities:

(20) no matter *who/what/which problems/whose ideas* you talk about
(21) no matter *when/where/why/how cheaply* they do it.

There are no obvious restrictions on the *wh*-phrase, except that pied-piping is restricted, as it is in normal questions – the contrast between the ECs in (22) parallels that with the normal interrogatives in (23):

(22) a. no matter *what* the students are worried about
b. ?no matter *about what* the students are worried

(23) a. *What* are the students worried about?
b. ?*About what* are the students worried?

Ungoverned ECs are similar, and examples corresponding to (20) and (21) without the *no matter* and with the appropriate *wh-ever* expression are possible.[3]

However, ECs have a number of special properties compared to normal interrogatives. Most obviously, on the semantic side, they are interpreted not as questions but as a kind of conditional. This is clearest with *or* ECs, e.g. (7) is interpreted roughly as:

(24) They will do that *if it's essential and if it is not essential*.

And unlike questions, whose typical discourse function is to raise issues, the point of an EC is to explicitly *remove* an issue from discussion, to 'take it off the table'. So, for example, *They will do that no matter what you do* conveys that your potential actions are irrelevant to the issue at hand (hence the name 'unconditionals'). Unsurprisingly, they carry a presupposition that the issue to be removed would otherwise be somehow 'live', hence the bizarreness of (25a), compared to (25b):

(25) a. #This restaurant will succeed, no matter who the goalkeeper is.
b. This team will succeed, no matter who the goalkeeper is.

[3]The only exception is that there is a lexical gap with *why*: the expression *why ever* is not a normal *wh-ever* expression – it is only used as an emphatic form of *why*, expressing surprise. *Why ever did she go?* means something like *Why on earth did she go?*, and cannot be used as an EC.

The name 'exhaustive conditional' arises from the fact that they seem to be acceptable only if *all* 'live' possibilities are covered. For example (8) is only felicitous if Wales and Scotland are the only potential destinations, and the following is only acceptable on the presupposition that all outcomes involve Granny getting drunk to some degree:

(26) It'll be okay, no matter how drunk Granny gets.

Syntactically, ECs differ from other embedded interrogatives in two respects. First, ECs are required to be finite. Compare the EC in (27) with the normal embedded interrogative in (28):

(27) *They will leave, no matter what to do.
(28) I wonder what to do.

Second, they display more freedom, in allowing what H&P call 'reduction', that is, what might be interpreted as omission of the copula:

(29) a. It's hard to explain this, however good the students (are).
b. It's hard to explain this, no matter how good the students (are).

This reduction is not possible in ordinary *wh*-interrogatives, either root or embedded:[4]

(30) a. How good *(are) the students?
b. I wonder how good the students *(are).

There are a number of descriptive and theoretical challenges here, which are addressed in the remainder of the paper. Section 2 develops a basic HPSG analysis for unreduced ECs, focussing on governed cases involving *no matter*. Section 3 considers the description of reduced ECs, and considers a number of possible HPSG approaches. Our starting point is the framework of Ginzburg and Sag (2001) (G&S), in particular, the analysis of interrogatives.

2 An Analysis of Un-reduced ECs

In discussing un-reduced ECs, governed cases seem to pose fewer challenges, with ECs governed by *no matter* being the most straightforward.

The following suggests that nothing can intervene between *no* and *matter* in *no matter*:

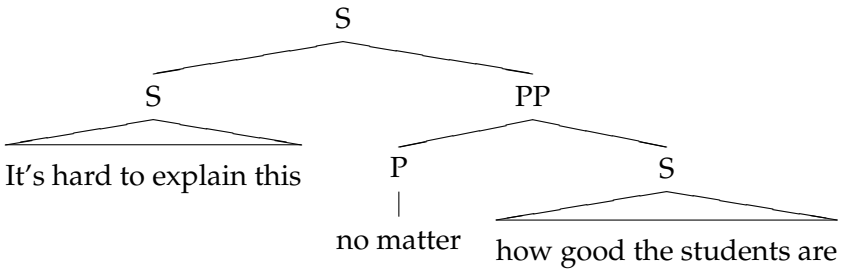
(31) *They will do that, no $\left\{ \begin{array}{c} \text{real} \\ \text{serious} \\ \text{earthly} \end{array} \right\}$ matter what you do.

[4]Interestingly, it is also possible in comparative correlatives, a point we shall return to: *The better the students (are), the more fun the classes (are).*

We conclude that *no matter* is a single lexical item, and given the preceding discussion, we assume it takes an interrogative complement and heads a conditional adjunct. Following H&P (p.761), we assume that it is a preposition. So example (32) will have a structure like (33).

(32) It's hard to explain this, no matter how good the students are.

(33)



We suggest the following lexical description for *no matter*.

(34)

$$\left[\begin{array}{ll}
\textsc{ss}\,|\,\textsc{loc} & \left[\begin{array}{ll}
\textsc{cat} & \left[\textsc{head}\left[\begin{array}{ll} \textit{prep} & \\ \textsc{mod} & \left[\textsc{head}\ \textit{verbal}\right]\!:\boxed{p} \end{array}\right]\right] \\
\textsc{cont} & \textit{exh-cond}(\boxed{q},\boxed{p})
\end{array}\right] \\
\textsc{arg-st} & \left\langle\left[\textsc{loc}\left[\begin{array}{ll}
\textsc{cat} & \left[\textsc{head}\left[\begin{array}{ll} \textit{verbal} & \\ \textsc{vform} & \textit{fin} \\ \textsc{ic} & - \\ \textsc{null} & \textit{boolean} \end{array}\right]\right] \\
\textsc{cont} & \boxed{q}\ \textit{question}
\end{array}\right]\right]\right\rangle \\
\textsc{background} & \{\ldots\}
\end{array}\right]$$

The value of ARG-ST here allows *no matter* to take as its complement any finite (VFORM *fin*) clause (HEAD *verbal*) which is interrogative (cf. the CONT value of *question*), and which is not a main clause (cf. the minus value for the independent clause (IC) feature). We will return to the NULL *boolean* feature in the discussion of reduced ECs.

This will license all the examples of governed ECs discussed above. For example, since HEAD *verbal* subsumes both verbs and complementisers, it is compatible with all forms of embedded polar questions (headed by a complementiser like *whether* or *if*), and constituent questions (which, following G&S, we assume are headed by verbs):

(35) no matter whether you go
(36) no matter whether you go or not
(37) no matter how clever the students appear to be

The minus value for IC (independent clause) not only excludes inverted

examples like (38), reprises like (39), and 'quiz show' questions like (40):

(38) *no matter how good are the students
(39) *no matter the students are how good
(40) *no matter the 1912 Olympics were held in which Scandinavian city

Reprise questions are excluded because while *The students are HOW good?* can be analysed as a question, it can only be a root question, hence [IC +].[5] Similarly, though 'quiz show' questions and similar *in situ* interrogatives can denote questions, they only do this as root clauses – on G&S's analysis an example like *And you propose that we should pay for it HOW, exactly?* only becomes a question because of a non-branching production which produces a root clause (IC+), and while it contains a non-root clause, this clause is declarative, and denotes a proposition, not a question (see G&S,p280ff). Either way it is excluded as a complement of *no matter*.

The [VFORM *fin*] restriction ensures that the complement must be finite, correctly excluding (27), repeated here:

(41) *They will leave, no matter what to do. [=(27)]

Given the value of HEAD|MOD in (34), *no matter* can modify any *verbal* expression, including Ss, CPs, and VPs. Examples of S modification can be seen above. The following show that ECs can modify VPs and CPs:[6]

(42) He will *go tomorrow* no matter what you say, and *stay away* no matter what you think.

(43) It is important *that we are early and that everyone else is on time*, no matter what happens.

The semantics of the modified expression is given as $\boxed{p}$ and the semantics of the complement of *no matter* is $\boxed{q}$. The overall semantics is given as *exh-cond(*$\boxed{q}$*,*$\boxed{p}$*)*, where we take *exh-cond(Q,P)* to be a condition that holds just in case freely choosing answers that resolve the question Q leave P holding, that is, just in case P holds for every resolution of Q.

Consider for example (44), whose semantics is given in (45c). The semantics of the antecedent is something like (45a), and the semantics of *who Cameron offends*, following G&S, is as in (45b) (this is the same as the semantics of *Who does Cameron offend?*).

(44) The Conservatives will win, no matter who Cameron offends.

[5] More generally, the combination of question semantics and the minus value for IC in (34) has the desirable effect of excluding all complements that do not contain an initial *wh*-expression – see G&S,p270ff.

[6] We are grateful to an anonymous referee for suggesting examples like (42). ECs can also attach to other kinds of phrase, as in e.g. *No true Scotsman, no matter where he lives, would tolerate this*, which are thus *prima facie* counter-examples to (34). We ignore this because ECs are just like other conditionals in this respect, e.g. *No true Scotsman, if he is honest, would tolerate this.* We assume a proper treatment of parentheticals would carry over to ECs.

(45) a. $win(TheConservatives)$
b. $\lambda\{x_{person(x)}\}offended(Cameron, x)$
c. $exh{-}cond(\lambda\{x_{person(x)}\}offended(Cameron, x), win(TheConservatives))$

Thus (44) will be true if we can freely choose among answers that resolve the question *Who does Cameron offend?* (e.g. *Abe*, *Bev*, . . . , *Zack*, . . . *Everyone*), with *The Conservatives will win* remaining true. It will be false in a situation where there is some answer (say, the answer corresponding to *Cameron offends the Queen*) whose truth is inconsistent with the Conservatives winning:

(46) a. Will the Conservatives win, no matter who Cameron offends?
b. No, if Cameron offends the Queen, the Conservatives won't win.

We have left the BACKGROUND value unspecified in (34). It should specify presuppositions to the effect that, first, $\boxed{q}$ (e.g. *Who does Cameron offend*) is a 'live' question, and second, that the possible answers to $\boxed{q}$ cover all and only the relevant possibilities (i.e. it should be exhaustive). It should also specify the intended discourse effect that $\boxed{q}$ is taken 'off the table' (e.g. does not enter, or is removed from, the set of questions under discussion).

Appropriate lexical entries for *regardless (of)* and *irrespective (of)* would be similar to (34), but raise some problems as regards the specification of the complement. In particular, coordination facts make an analysis of these items as single lexemes implausible (cf. examples like *regardless of whether you stay or of whether you go*), and notice that the syntactic and semantic requirements stated in the ARG-ST of *no matter* must be imposed on what one might plausibly take to be the complement of *of*, rather than the complement of *regardless* or *irrespective*.[7]

The obvious way to extend this approach to ungoverned ECs would be to introduce a special construction (a sub-type of *non-headed-phrase*) whose mother has the semantics of *no matter*, and a single daughter corresponding to the complement of *no matter*, along the lines of (47).

This is not satisfactory as it stands. First, and less important, it is unclear what category we should assign to the mother here – the analogy of governed ECs would suggest either preposition (like *no matter*) or adjective (like *regardless* and *irrespective*). Neither has much intuitive appeal, but nor is there a clearly motivated alternative. More important, this account will overgenerate, since it will allow ungoverned polar interrogatives like (48), and ungoverned *wh*-questions which do not contain a '*wh-ever*' form like (49), both of which are possible in governed ECs:

(48) They will win the election *(no matter) whether Cameron is replaced.

[7] One possibility here would be to treat *of* as a 'weak head' in the sense of Tseng (2002): *regardless* and *irrespective* would be like *no matter* except for being adjectives and taking a complement which is specified as [MARKING *of*].

(47)
$$\left[\begin{array}{l}\textsc{ss}\,|\,\textsc{loc}\ \left[\begin{array}{l}\textsc{cat}\ \left[\textsc{head}\ \left[\begin{array}{l}\textit{part-of-speech}\\ \textsc{mod}\ \left[\textsc{head}\ \textit{verbal}\right]:\boxed{p}\end{array}\right]\right]\\ \textsc{cont}\ \textit{exh-cond}(\boxed{q},\boxed{p})\end{array}\right]\\ \textsc{background}\ \{\ldots\}\end{array}\right]$$

$$|$$

$$\left[\textsc{ss}\,|\,\textsc{loc}\ \left[\begin{array}{l}\textsc{cat}\ \left[\textsc{head}\ \left[\begin{array}{ll}\textit{verbal}&\\ \textsc{vform}&\textit{fin}\\ \textsc{ic}&-\\ \textsc{null}&\textit{boolean}\end{array}\right]\right]\\ \textsc{cont}\ \boxed{q}\,\textit{question}\end{array}\right]\right]$$

(49) They will win the election *(no matter) who Cameron offends.

However it appears to be a reasonable starting point for an analysis.

3 An Analysis of Reduced ECs

We now turn our attention to the phenomenon of 'reduced' or 'null copula' ECs, which may be governed or ungoverned:

(50) a. This is hard to teach, no matter how good the students (are).
b. This is hard to teach, however good the students (are).

Let us call the part of an EC that denotes a question (e.g. the complement of *no matter*) the 'ECQ'. Given the analysis in Section 2, a reduced EC will just be a normal EC with a reduced ECQ daughter.[8] Pre-theoretically, an ECQ is a verbless clause with two daughter constituents: the first is a *wh*-phrase (e.g. *how good*), the second (e.g. *the students*) is interpreted as the subject of the first, and it is natural to talk informally about an 'omitted' copula. Formally, the distinction between reduced and unreduced ECQs will be encoded in the feature NULL: reduced ECQs will be [NULL +].

The kind of *wh*-phrase that appears most easily and commonly is an AP with *how*, and it is these we will focus on in developing our analysis. The basic facts are these.

[8]The difference between an EC and the ECQ it contains is easy to see with a governed EC. It is harder with an ungoverned case like *however good the students (are)* because the EC and the ECQ it contains are string identical.

First, only a copula that is the highest verb in the ECQ can be omitted, as the following illustrate:

(51) This is hard to teach, no matter how good the students may *(be).
(52) This is hard to teach, no matter how good it seems the students *(are).

Second, as a number of authors have noted, there are restrictions on the subjects of reduced ECQs (e.g. Culicover, 2013). In particular, pronominal subjects are not possible:

(53) a. This is hard to teach, however good they *(are).
b. This is hard to teach, no matter how good they *(are).

Demonstratives, proper nouns, quantificational NPs, and indefinite NPs are also excluded:

(54) no matter how good that person over there *(is)
(55) no matter how good John *(is)
(56) no matter how clever everyone *(is)
(57) no matter what time a class *(is), . . .

Moreover, *only* NPs are allowed, e.g. clauses and PPs are not possible:

(58) no matter how interesting *whether he left or not* *(might be), . . .
(59) no matter how good a place *under the bed* *(might be), . . .

In fact, it seems that only definite NPs with *the* or a possessive are possible:

(60) They are always cheerful, no matter what time $\left\{ \begin{array}{c} \text{the} \\ \text{their} \end{array} \right\}$ class.

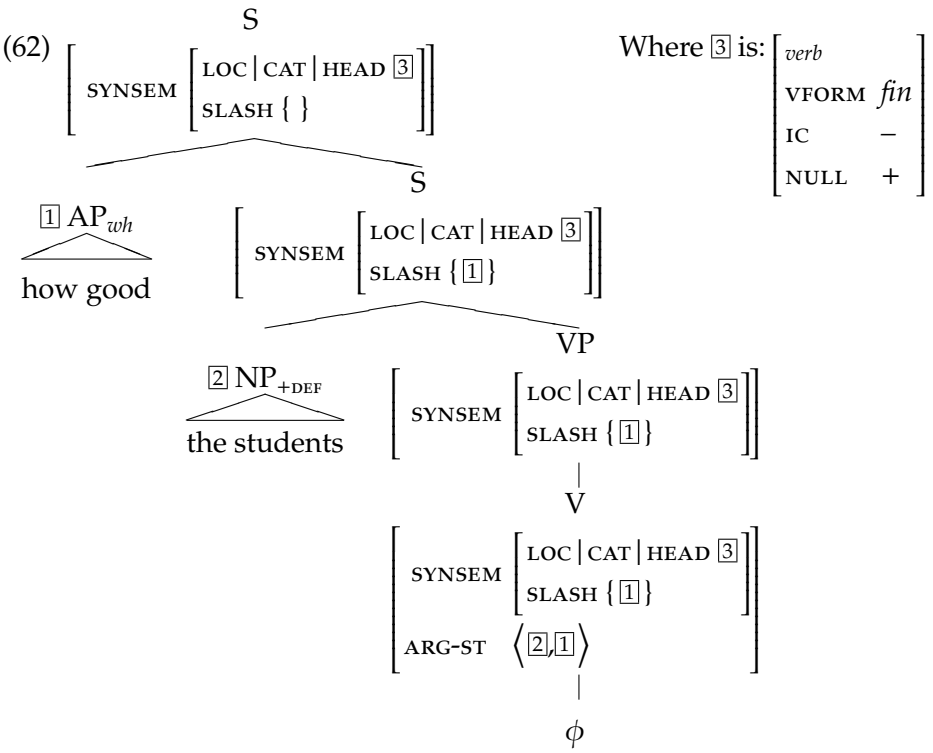
As regards the *wh*-phrase, the most obvious constraint is that it must be initial in the clause (as it must in unreduced ECQs of course):

(61) a. no matter how clever the students
b. *no matter the students how clever

We will consider three different analyses. The first involves a null-copula. The second two are constructional. According to the first and second, the *wh*-phrase is a filler. According to the third, the relation of the subject and *wh*-phrase is just that of a subject and predicate.

3.1 A Non-Constructional Filler Analysis (Empty Copula)

In unreduced ECQs the initial *wh*-phrase is plausibly analysed as a filler, as in a normal question, so it is natural to assume that it is also a filler in reduced ECQs. As noted in Borsley (2004, 2011) in relation to comparative correlatives, one way to provide an analysis of reduced phrases is to postulate a phonologically empty form of the copula which takes a gap as its complement, giving structures like (62). On this account, a reduced ECQ consists of a *wh*-filler, and a slashed S, which in turn contains a subject and



a slashed VP, containing an empty copula verb (which we write as 'ϕ', and which is we assume is lexically specified as [NULL +]).

We can rule out some examples of omission of embedded copulas, like (51) (repeated here), if we assume that ϕ has no non-finite form (ϕ is here the complement of *may*, which is required to be non-finite):

(63) *. . . no matter how good the students may ϕ. [=(51)]

To exclude other examples we would have to assume that verbs generally select [NULL –] complements, excluding complements headed by ϕ.:

(64) *. . . no matter how good it seems the students ϕ. [=(52)]

We can capture the restrictions on the subject of reduced ECQs straightforwardly, as restrictions on the subject of ϕ, and restrictions on the *wh*-phrase as restrictions on its complement.

It would increase the plausibility of this analysis if it could be shown that ϕ has some similarity with other null copulas that have been independently proposed for English in the HPSG literature.[9] However, this is not possible.

Apart from the idiosyncratic restriction on the subject (only definite NPs), its complementation behaviour is quite restricted when compared to other

[9]Examples include Sag et al. (2003), Bender (2001), and Avgustinova (2006).

null copulas. For example, because the *wh*-expression in an ECQ is always fronted, the complement of ϕ is always a gap. By contrast according to Bender the complement of the AAVE null copula is never a gap – it is always *in situ*.

Notice also that ϕ must be compatible with both present and past tenses, since it is compatible both with environments which require present and environments which require past forms of the overt copula:

(65) a. They were brave, no matter how dangerous the situation was/*is.
b. They are brave, no matter how dangerous the situation *was/is.
c. They were brave, no matter how dangerous the situation ϕ.
d. They are brave, no matter how dangerous the situation ϕ.

This is also unlike the situation with AAVE null copula, which is generally assumed to have no past form (e.g. Bender (2001, p87)).

Interaction with negation is also problematic. Consider the examples in (66). In a situation in which various people have made statements about what the answer is not, a participant who feels the discussion has been excessively negative might try to move it onwards and away from these negative views by saying something like (66a) with an overt copula. Notice that the corresponding reduced example, (66b), is ungrammatical:

(66) a. No matter what the answer is not, we need to move on.
b. *No matter what the answer ϕ not, we need to move on.

This should be surprising. For example, on the widely accepted analysis of Kim and Sag (2002), negation involves the addition of an optional complement to auxiliary verbs (including so-called 'main verb' *be*). One would expect ϕ to be able to undergo the same process, licensing *not* just like an overt copula. Notice that according to Bender (2001), some speakers of AAVE specifically allow a null copula in main clauses with negation (e.g. *They say they're best friends and shit, but they not.* (Bender, 2001, 115)), suggesting that it undergoes this process.

More generally, ϕ seems to resist adverbial modification:[10]

(67) no matter how difficult the problem actually *(is)

Notice this is not because of a general constraint on adverbials appearing next to phonologically empty structure – for example it is quite possible to have an adverbial in a clause that has undergone gapping:

(68) Sam is allegedly in London, and Kim actually Δ in Rome.

Of course, these objections are not fatal. But there are more serious em-

[10]This point is different from the point about negation: while it is widely accepted that negation involves addition of an optional complement, the consensus is that preverbal adverbs like *actually* in (67) are *not* complements – see, e.g. Bouma et al. (2001).

pirical objections. In particular, on this analysis missing copula clauses consist of a slashed S, and contain a slashed VP, and one would expect it to be possible to conjoin them with similarly slashed constituents. However, this is clearly impossible with slashed VPs:[11]

(69) *. . . no matter how good the students [ϕ] or [seem to be]

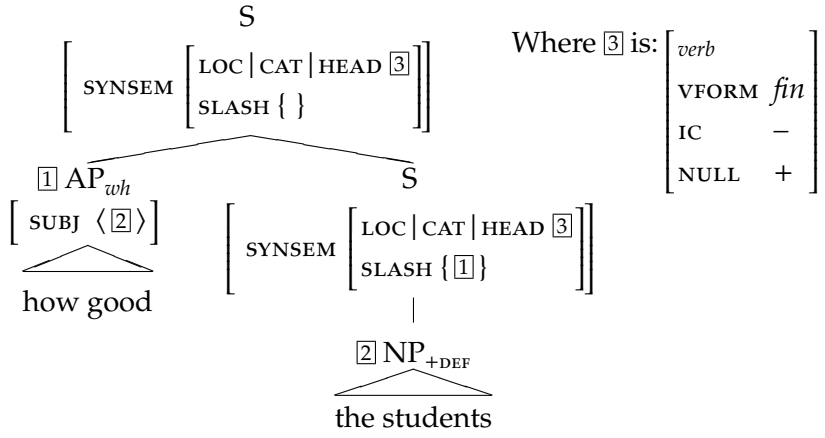
Examples involving a slashed S are equally bad: (70) cannot be interpreted as a conjunction of ECQs (the string represented in (70) can be interpreted, but not as a conjunction of ECQs – the interpretation involves a single ECQ with *the students or the lecturers* as the subject of *seem*).[12]

(70) *no matter how good [the students ϕ] or [the lecturers seem to be]

3.2 A Constructional Filler Analysis

However, it is also possible to provide a filler analysis without assuming a null form of the copula if one takes a constructional view. The non-*wh* sister can be treated as an S with a predicative expression in its SLASH value, and a single definite NP daughter which satisfies the subject requirements of the predicative expression.

(71)



[11] An anonymous referee points out that there is a general constraint that conjuncts cannot be gaps (cf. the Element Constraint, e.g. Sag et al. (2003, Ch14)), and it is conceivable that some such constraint might apply to all empty elements, including a null copula. However, it is not straightforward to invoke such a principle here – in general what excludes gaps as conjuncts in cases like **Who did you see* Δ *and Kim?* is the lack of a head to license them (gaps are licensed as an effect of argument realisation). But this constraint is not applicable here where there would be a phonologically null head, not a missing argument.

[12] Of course, a reduced clause or VP and a non-reduced clause or VP will differ in the value of HEAD | NULL, but there is in general no requirement that conjuncts agree in their *head* values. The only requirement is that all conjuncts be compatible with the environment of the coordinate structure. So for example since *know* allows interrogative complements whose HEAD values are specified as either finite or non-finite, it allows coordinations of such complements: *I don't know what to drink or whether I'm allowed to eat anything*.

Here the higher S is a normal *head-filler-phrase*, and (apart from being [NULL +]) a normal *wh-interrogative-clause*, hence, for example, the *wh*-marking on the AP. The lower S involves a new phrase type, what we might call a *missing-copula-clause*, a subtype of *non-headed-phrase*. It involves a slashed S mother with an unslashed NP daughter (since the daughter is not a head this is consistent with the head-driven view of SLASH assumed in G&S). The type *missing-copula-clause* could be constrained as follows.

(72)
$$\textit{missing-copula-clause} \rightarrow \left[\begin{array}{ll} \text{SS} \mid \text{LOC} \mid \text{CAT} & \left[\begin{array}{ll} \text{HEAD} & \left[\begin{array}{ll} \textit{verb} & \\ \text{NULL} & + \\ \text{IC} & - \end{array}\right] \\ \text{SUBJ} & \langle\,\rangle \\ \text{SLASH} & \left\{ \text{AP}\left[\text{CAT} \mid \text{SUBJ}\left\langle \boxed{1} \right\rangle\right] \right\} \end{array}\right] \\ \text{DTRS} & \left\langle \left[\text{SYNSEM}\ \boxed{1}\ \text{NP}_{+\text{DEF}}\right] \right\rangle \end{array}\right]$$

This requires a *missing-copula-clause* to be an embedded clause (*verbal*, and SUBJ $\langle\,\rangle$), with a single definite NP daughter. The clause has as its SLASH value an AP whose SUBJ value is identified with the SYNSEM value of that single NP daughter.

Empirically, this approach can account for all the data that the null-copula analysis deals with. For example, there is no possibility of examples like (51) and (52), because on this analysis the second daughter of a reduced ECQ is an S containing just a definite NP, but the relevant parts of (51) and (52) cannot be analysed in this way.

(73) *. . . no matter how good *the students may*. [=(51)]

(74) *. . . no matter how good *it seems the students*. [=(52)]

This account improves on the null-copula account in several ways (e.g. since there is no verb, there is no possibility of negation, or adverbial modfication, so examples like (66b) and (67) are excluded straightforwardly), and there is no problem with examples like (69) involving conjunction of slashed VPs – there is no VP here to be conjoined.

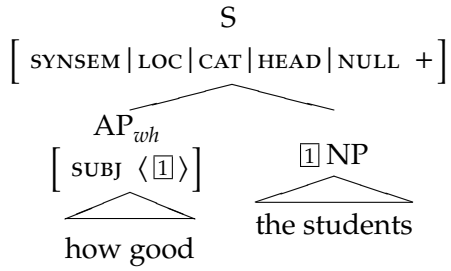
Unfortunately, however, one of the major empirical objections remains: we still have a slashed S, and so we still wrongly predict that it should be possible to conjoin the subject of the reduced clause with a slashed S, as in (70).

This suggest that we should abandon the idea that the second daughter of a reduced clause is a slashed S, and look for an alternative analysis.

3.3 A Constructional Non-Filler Analysis

One possible alternative, suggested by Culicover (2013, 121-126), is that reduced ECQs involve a predicative expression preceding its subject, as in (75). We will extend an formalise this idea.

(75)



Let us call this construction a *reduced-wh-interrogative-clause*. It will be a sub-type of *non-headed-phrase* and *wh-interrog-clause*. It consists of an S, marked NULL+, dominating a *wh*-phrase, followed by the subject of the *wh*-phrase.

This avoids the coordination problems noted above: since on this analysis *the students* is an (un-slashed) NP, we would expect conjunction with a slashed clause, or a slashed VP to be impossible, as in (69) and (70), repeated here:[13]

(76) *no matter how good [the students] or [seem to be]
(77) *no matter how good [the students] or [the lecturers seem to be]

Likewise, there is no problem with negation – since there is no auxiliary verb, there is no argument structure to which *not* can be added, hence we account for the ungrammaticality of (78), and there is nothing for *actually* to modify, accounting for the impossibility of (67):[14]

(78) *No matter what the answer not, we need to move on. [cf. (66b)]
(79) *no matter how difficult the problem actually [cf. (67)]

We can rule out examples like (51) and (52), where the 'missing copula' is not the highest verb in the ECQ, straightforwardly: on this analysis the second daughter of a reduced ECQ is just an NP, but the relevant parts of (51) and (52) cannot be analysed in this way:

(80) *no matter how good *the students may*. [=(51)]
(81) *no matter how good *it seems that the students*. [=(52)]

[13] Recall that there is nothing wrong with the *string* in (77), just it cannot be understood as the coordination of two ECQs.

[14] Reduced versions of examples with post-verbal adverbs, like *in former times* in *no matter how good the students (were) in former times*, can be analysed as having the adverbial adjoined to S in (75).

As with the other analyses restrictions on the *wh*-expression and the subject in reduced ECQs can be dealt with straightforwardly, as constructional effects. We thus have an empirically satisfactory account.

Moreover, the account is not as stipulative as one might fear, based on (75), because the structure in (75) is not as idiosyncratic as might first appear. It is, in particular, very similar to a plausible analysis of reduced comparative correlatives (e.g. Borsley, 2004, 2011; Culicover, 2013):

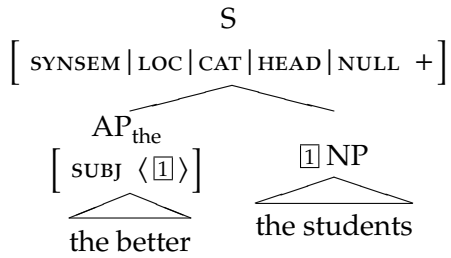
(82) The better the students (are), the more fun the class (is).

Notice there are similar restrictions on the subject NP, e.g. no pronouns or proper names:

(83) The better they *(are), the more fun the class is.

(84) The more interesting Kim *(is), the less interesting Pat *(is)

A plausible analysis for the initial part of (82) is provided in (85):

(85)


Compared to (75), the main syntactic difference is that the AP in a comparative correlative is marked with a feature that guarantees the presence of *the* (e.g. [CORREL *the*]), whereas the AP in (86) is marked +WH (more precisely, it has a WH feature whose value is a non-empty set of *parameters*).

These commonalities can be factored out, and assigned to a new construction type which we will call *reduced-phrase*, a sub-type of *non-headed-phrase*, which has sub-types *reduced-wh-interrogative-clause* and *reduced-comparative-correlative-clause*. We thus amend the HEADEDNESS dimension of the type system in G&S as in Figure 1, where our addition is highlighted.

*Reduced-phrase*s are constrained as in (86), equivalent to (87). That is, a *reduced-phrase* is a non-root *verbal* expression (e.g. S) marked [NULL +], containing a predicate and its subject NP.

(86)
$$\left[\begin{array}{l} \textit{reduced-ph} \\ \\ \text{SS} \mid \text{LOC} \left[\text{CAT} \left[\text{HEAD} \left[\begin{array}{ll} \textit{verbal} & \\ \text{NULL} & + \\ \text{IC} & - \end{array}\right]\right]\right] \\ \\ \text{DTRS} \quad \left\langle \left[\text{SYNSEM} \left[\text{SUBJ} \left\langle \boxed{1} \right\rangle\right]\right], \left[\text{SYNSEM} \; \boxed{1} \; \text{NP}\right] \right\rangle \end{array}\right]$$

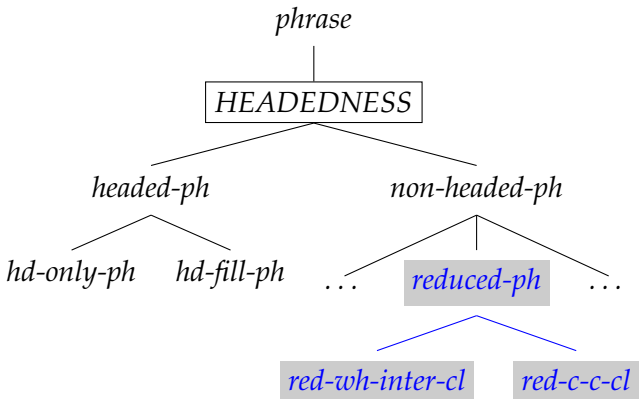


Figure 1: The HEADEDNESS dimension

(87)
$$\left[\text{SS}\,|\,\text{LOC}\,|\,\text{CAT}\left[\text{HEAD}\left[\begin{array}{ll}\textit{verbal} & \\ \text{NULL} & + \\ \text{IC} & -\end{array}\right]\right]\right]$$
$$\left[\text{CAT}\,|\,\text{SUBJ}\,\langle\boxed{1}\rangle\right] \quad \boxed{1}\text{NP}$$

Rather than stipulate the other properties of reduced ECQs (e.g. their *question* semantics), we would prefer to inherit this information from elsewhere. Since all reduced ECQs are *wh*-interrogatives, the obvious super-type for this inheritance is *wh-inter-clause*. Unfortunately, this cannot be implemented directly, since *wh-inter-clause* is a subtype of *head-filler-phrase*, itself a sub-type of *headed-phrase*, and the analysis we are developing here assumes that reduced clauses are un-headed. To accommodate this, we can amend the CLAUSALITY dimension of G&S's type hierarchy as in Figure 2, distinguishing *regular-wh-interrogative-clauses* (i.e. normal *wh*-interrogatives – what were formerly called just *wh-interrogative-clauses*) and *reduced-wh-interrogative-clauses*, which we are concerned with here.

The revised dimensions can be combined as in Figure 3 (where for readability we omit all sub-types of *inter-cl* except *wh-inter-cl*).

Notice that this leaves G&S's hierarch essentially unchanged, and allows us to derive the properties of reduced ECQs almost without stipulation. Because they are a sub-type of *reduced-phrase* they are clauses, they consist of a predicative phrase and its subject, and they are restricted to embedded contexts, and contexts that permit [NULL +] clauses. Because they are a sub-type of *wh-inter-clause* they have the semantics of questions, and contain a *wh*-expression.

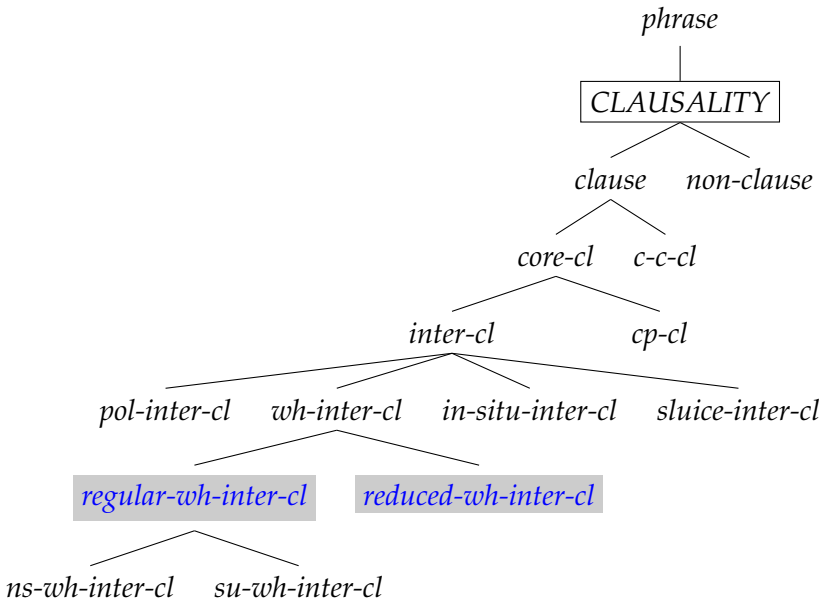


Figure 2: The CLAUSALITY dimension

Essentially the only constraint we require is one that will derive the semantics of the reduced clause from the semantics of the initial *wh*-phrase:

(88) *reduced-wh-inter-cl*:

$$\left[\text{CONT}\left[\text{PROP}\ \boxed{1}\right]\right] \quad \rightarrow \quad \left[\text{CONT}\ \boxed{1}\right], \quad \text{NP}$$

This is comparable to the G&S's Propositional Head Constraint (p229) which makes the semantics of a regular *wh*-question depend on the propositional semantics produced by its head, so that in *How good are the students?*– roughly 'the students are x-much good'. (88) will ensure we get the same semantics for a reduced ECQ (*no matter how good the students*).

4 Problems, Discussion

In the previous sections we have given a basic HPSG analysis of ECs, including reduced ECs. It consists of a lexical entry for *no matter*, and a novel construction (*reduced-wh-interrogative-clause*), a non-standard predicative construction, which has similarities with comparative correlatives, and which captures the properties of reduced ECs (for un-reduced ECs there is nothing to say – the ECQ is just a normal interrogative). This is still some way from a complete account of the phenomena, however. In this section, we summarise some of the remaining problems and open questions.

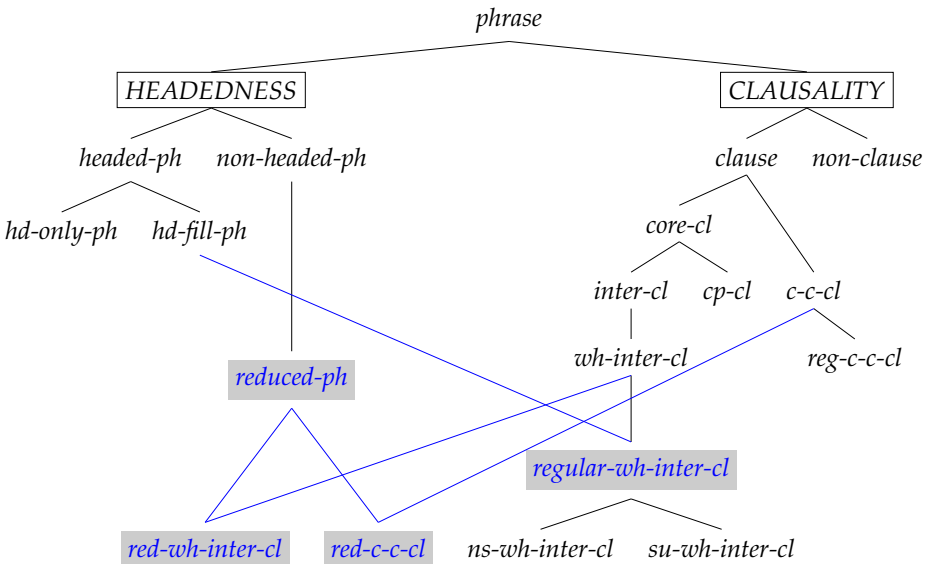


Figure 3: The revised hierarchy of *phrase* types

First, as noted at the end of Section 2, we have given only a partial account of ungoverned ECs, and of ECs governed by expressions other than *no matter*.

Second, as regards the *wh*-expression in reduced ECs, our discussion has focussed on APs involving *how*. However, other kinds of *wh*-expression are attested, the following are some examples (lightly edited from corpora):

(89) They rarely find fault with paintings, no matter *what* their subject or style.

(90) . . . not to tolerate any further human rights abuses, no matter *who* the perpetrators

(91) . . . personnel can get the information they need, no matter *where* the incident.

(92) . . . must be completed, no matter *when* the deadline.

(93) . . . should be considered, no matter *what nationality* the applicant.

(94) Massachusetts has a no-fault workers' compensation system that provides medical benefits. . . , no matter *whose fault* the accident.

For the most part, dealing with these involves a simple extension of the account we have presented. However, some of these have interesting theoretical implications. For example, our analysis involves the *wh*-phrase having a subject slot, i.e. being in some sense predicative. For APs, such as discussed in Section 3 this is clearly reasonable. For some other *wh*-expression it is less obviously correct. For example the question involved in (90) is 'identificational' (cf. a potential answer 'the perpetrators are General

X and Colonel Y'), and many approaches would assume that in such a case *who* would not have a SUBJ slot – in which case (90) would not fit any of the analyses we have looked at. While there are other approaches, including G&S (p195), which assume there is a SUBJ slot in such cases, and which would be consistent with our analysis, the issue deserves consideration.

There are also some restrictions on the kind of *wh*-expression that can occur in reduced ECs. Some, like (95), we an account for straightforwardly.

(95) *no matter which students successful

The ungrammaticality of examples like this is predicted on our account: on our account, a reduced ECQ consists of a predicate followed by a subject, but in a case like this the word order is subject-predicate – cf. the corresponding un-reduced example would be (96) (to put it another way, in this cases the *wh*-phrase cannot be analysed as having an open SUBJ slot, and the second daughter is not an NP):

(96) no matter which students are successful

However, other examples are more puzzling: (97), is ungrammatical, and an initially appealing explanation is that this is because there is something wrong with the question it involves ('What geniuses are the students?' is not a question that has any very obvious range of potential answers, which is something that is required for an EC).[15]

(97) *no matter what geniuses the students

Unfortunately, this account is hard to reconcile with the fact that the un-reduced version is acceptable:

(98) no matter what geniuses the students are

Thus, there are a number of issues that require further investigation.

References

Abeillé, Anne and Borsley, Robert D. 2008. Comparative correlatives and parameters. *Lingua* 118, 1139–1157.

Avgustinova, Tania. 2006. A Functional Typology of Copular "be": towards an HPSG Formalisation. In Stefan Müller (ed.), *The Proceedings of the 13th International Conference on Head-Driven Phrase Structure Grammar*, pages 27–38, Stanford, Ca.: CSLI Publications.

Bender, Emily M. 2001. *Syntactic Variation and Linguistic Competence: The Case of AAVE Copula Absence*. PhD thesis, Stanford University.

[15] We are grateful to an anonymous referee for bringing examples like (97) to our attention.

Borsley, Robert D. 2004. An Approach to English Comparative Correlatives. In Stefan Müller (ed.), *Proceedings of the The 11th International Conference on Head-Driven Phrase Structure Grammar*, pages 70–92, Stanford, Ca.: CSLI Publications.

Borsley, Robert D. 2011. Constructions, functional heads and comparative correlatives. In O. Bonami and P. Cabredo Hofherr (eds.), *Empirical Issues in Syntax and Semantics 8*, pages 7–26.

Bouma, Gosse, Malouf, Rob and Sag, Ivan A. 2001. Satisfying Constraints on Extraction and Adjunction. *Natural Language and Linguistic Theory* 1(19), 1–65.

Culicover, Peter W. 2013. *Grammar & Complexity: Language at the Intersection of Competence and Performance*. Oxford: Oxford University Press.

Ginzburg, Jonathan and Sag, Ivan A. 2001. *Interrogative Investigations: the form, meaning, and use of English Interrogatives*. Stanford, Ca.: CSLI Publications.

Kim, Jong-Bok and Sag, Ivan A. 2002. French and English Negation without Head-Movement. *Natural Language and Linguistic Theory* .

Kubota, Yusuke. 2003. Yet Another HPSG-Analysis for Free Relative Clauses in German. In Jong-Bok Kim and Stephen Wechsler (eds.), *The Proceedings of the 9th International Conference on Head-Driven Phrase Structure Grammar*, pages 147–167, Stanford, Ca.: CSLI Publications.

Müller, Stefan. 1999. An HPSG-Analysis for Free Relative Clauses in German. *Grammars* 2(1), 53–105.

Rawlins, Kyle. 2008. *An Investigation in the Syntax and Semantics of Conditional Structures*. PhD thesis, University of California, Santa Cruz.

Rawlins, Kyle. 2013. (Un)conditionals. *Natural Language Semantics* 21(2), 111–178.

Sag, Ivan A., Wasow, Thomas and Bender, Emily M. 2003. *Syntactic Theory: A Formal Introduction*. Stanford, Ca: CSLI Publications, second edition.

Tseng, Jesse L. 2002. Remarks on Marking. In Frank Van Eynde, Lars Hellan and Dorothee Beermann (eds.), *Proceedings of the 8th International Conference on Head-Driven Phrase Structure Grammar*, pages 267–283, Stanford, Ca.: CSLI Publications.

Zaefferer, Dietmar. 1990. Conditionals and unconditonals in universal grammar and situation semantics. In Robin Cooper, Kuniaki Mukai and John Perry (eds.), *Situation theory and its applications, Vol 1*, CSLI Lecture Notes, No. 22, pages 471–492, Stanford, Ca.: CSLI Publications.

Between complex predicates and regular phrases: German collocational clusters

Philippa Cook
Freie Universität Berlin

Proceedings of the 21st International Conference on
Head-Driven Phrase Structure Grammar

University at Buffalo

Stefan Müller (Editor)

2014

Stanford, CA: CSLI Publications

pages 48–62

Cook, Philippa. 2014. Between complex predicates and regular phrases: German collocational clusters. In Stefan Müller (ed.), *Proceedings of the 21st International Conference on Head-Driven Phrase Structure Grammar, University at Buffalo*, 48–62. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2014.3. CC-BY 4.0

Abstract

I argue for a new type of non-standard constituent in German; a *modifier-collocational-cluster*. This type of cluster combines (i) a modifier and (ii) a PP from a light-verb construction (or a Funktionsverbgefüge (FVG) as they are known in German) or a bare noun. Such strings are found in German in initial (prefield) position in certain cases of apparent multiple fronting. We are dealing with a syntax-semantics mismatch here since the modifier does not semantically modify the element with which it can first syntactically combine. I show that the modifier is a collocate of both its co-prefield element but also of the verb. I propose a schema which lexically licenses the building of such clusters and I show how we can encode information about what I refer to as collocational selection in the lexical entries of the type of lexemes involved in these multi-word strings. The analysis can be seen as lexical but does not require lexical storage of phrasal elements.

1 Introduction

I propose a new analysis of certain multi-word strings in German such as (i) *heftig in die Kritik geraten* 'to be heavily criticised', (ii) *weltweit für Aufregung sorgen* 'to cause worldwide concern' or (iii) *richtig Geld verdienen* 'to make real money', postulating units I will call (*modifier-*)*collocational chunks*. The strings in (i) and (ii) involve a (semi-compositional) support verb construction, cf. Krenn & Erbach (1994), Steinitz (1989), *in die Kritik geraten* (literally: into the criticism fall) or *für Aufregung sorgen* (literally: for excitement provide) with modification by an adverbally used adjective, *heftig* 'harsh(ly)' or *weltweit* 'wordlwide' respectively.[1] In (iii) a verb *verdienen* 'earn' selects a bare noun, and there is again modification by an adverbally used adjective *richtig*.[2] I argue that these strings are lexically encoded as multi-word expressions but we will see that this does not mean they have to stored as phrasal entries. They are, I believe, situated on a continuum inbetween genuine complex predicates at one extreme of the spectrum and canonically composed syntactic phrases at the other extreme. I take these lexical strings to

[†]Thanks to Felix Bildhauer, Stefan Müller, John Payne and Elodie Winckel as well as the audience and anonymous reviewers of HPSG 2014 (especially Farrell Ackerman, Anne Abeillé, Berthold Crysmann, Ray Jackendoff and Bob Levine) for comments and suggestions. Thanks also to the audience of the *MehrWortverbindungen* conference at the University of Basel in October 2014. This Research was funded by a DFG grant to project A6 of the SFB 632.

[1]The part of speech *adjective* can be used in German as a pre-nominal modifier or predicatively but **also** in the function of an adverb with no morphological difference. A word-form such as *richtig* is therefore ambiguous in isolation. This ambiguity is undoubtedly a central contributing factor in the licensing of what I am calling *modifier-collocation-clusters*.

[2]In the data I will be discussing, *richtig* functions as an intensifier rather than as the manner adverb 'correctly'. Since *richtig* has a dual status (manner adverb or intensifier), the string is in principle ambiguous. There are similar, but less compositional, strings for which the manner reading is much less salient than the intensifier reading, e.g. *richtig Gas geben* 'to increase effort/to really go for it', lit: really give gasoline, viz. example (1d) below. Note, *geben* is 2-place here and clearly semantically bleached.

be exemplific of several larger classes of data patterning similarly, although with small differences across subclasses (and not all involving modifiers). For reasons of space I cannot document the full array of data here but see the comments in § 5 below.

The paper is structured as follows: In section 2, data is introduced which suggests we may have to accept a non-standard type of constituent in German, licensed only in the presence of certain combinations of lexical material. In section 3, the collocational relationships, cf. Firth (1957), Sinclair (1991, 1996), Evert (2008), spanning all three subcomponents of the string are discussed and in Section 4 it is proposed that sub-parts of such strings (namely the modifier and the PP/bare noun) may combine in German via a special schema for building collocational chunks, rather than building traditionally known syntactic constituents. The proposed schema is inspired by Function Composition known from Combinatorial Categorial Grammar (CCG). Although each of the three elements in the string is individually a syntactic atom of a multi-word string, the combination as a whole should be viewed as one complex lexeme, the building of which is licensed lexically.

2 Apparent cases of multiple fronting

German main clause declaratives are subject to the verb-second constraint; i.e. precisely one constituent may occur in the initial position preceding the finite verb (in a position referred to as the *prefield*. (1a)-(1d) instantiate (a certain type of) so-called *apparent multiple fronting* construction in which the clause-initial position before the finite verb contains a string that does not fit the traditional definition of constituent. Here, we have a modifier and a bare noun or PP. Semantically the modifier in initial position modifies the whole PP/N + V string. Syntactically, though, the modifier (surprisingly) combines with the PP or the N. Not only do we have a non-isomorphism of syntax and semantics (a syntax-semantics mismatch), but also a curious constituent structure.

(1) a. [**Weltweit**] [**für Aufregung**] sorgt eine Werbekompagne von
worldwide for upset provides an advertising-campaign from
Benetton[3]
Benetton
'A Benetton advertising campaign is causing international concern'

b. [**Heftig**] [**in die Kritik**] geriet der Kostenrechnungsbericht des
heavy into the criticism fell the finance report the
Jugendamtes für 2002[4]
youth service for 2002
'The youth service's 2002 financial report got slated'

[3]http://woodz.schwarzwaelder-bote.de/alltag/lifestyle/8422-benetton-zieht-kuss-foto-von-papst-zurueck.html, checked 14.10.2014

[4]COSMAS, RHZ03/SEP.09166 Rhein-Zeitung, 12.09.2003

c. [**Richtig**] [**Geld**] wird nur im Briefgeschäft verdient[5]
right money is only in letter.business earned
'You can only make real money with letters'

d. [**Richtig**] [**Gas**] wird in der Großraum Disco "Cocos Club" ab den
right Gas will in the large-scale Disco Cocos Club from the
Sa. 16.02.2008 gegeben[6]
Sat. 16.02.2008 given
'It's going to be all-go in the large-scale Disco "Cocos Club" as of Saturday 16th February 2008'

This phenomenon has been documented by Müller (2003, 2005) who proposes an analysis in which the initial position houses a VP-constituent with an empty head (a structure that is used anyway in many approaches to German). I provide an alternative analysis for *apparent multiple fronting* data specifically of the type in (1) drawing on the concept of collocation.[7] Support for the claim that strings such as *heftig in die Kritik* are **collocational clusters** (a string akin to some kind of *chunk*/*prefab*) can be gleaned from the observation that the material in the puported cluster prefers to permute (scramble) together rather than individually, viz.

(2) a. weil **heftig in Kritik** der Bericht geriet
because heavy in criticism the report fell
'because the report got slated'

b. ? weil **heftig** der Bericht **in Kritik** geriet
because heavy the report in criticism fell
intended: 'because the report got slated'

c. ? weil **in Kritik** der Bericht **heftig** geriet
because in criticism the report heavy fell
intended: 'because the report got slated'

3 The collocational nature of the lexemes in the multiple fronting data

It has been noticed that the material in the initial string in constructions known as apparent *multiple fronting* intuitively forms a tight unit and that often at least one prefield element forms some kind of unit with the verb. We will see below that in the cases under discussion here, both prefield elements form a bond with the verb. Taking the strings from (1) above we can use a collocation association measure to ascertain whether or not this intuitive *unithood* can be verified. I employ the *Wortprofil 3.0* tool offered by the *Digitales Wörterbuch der Deutschen Sprache*

[5] taz 28./29.10.2000, p. 5, taken from Müller (2005)

[6] http://www.my-nrw.de/nachtflug.php?kat=91&id=9806, checked 14.10.2014

[7] Müller's analysis covers a much broader range of data than discussed here. It is conceivable that Müller's analysis could be retained and, for the type of data discussed here, be enhanced to include some kind of collocational analysis.

(DWDS) corpus, cf. Didakowski & Geyken (2013), which uses the LogDice measure of Rychlý (2008). Such lexicographically-oriented approaches to collocation use a notion of headword or node and examine collocates in a relation of dependence to one another such that e.g. in a modifier-noun collocation, the noun is headword and in a verb-object collocation, the verb is headword. The association measures for the lexemes mentioned, in the stated dependency relation, are given here.

	association using LogD	frequency
heftig as modifier of *Kritik*	11.12	9882
Kritik as object of *geraten*	9.27	2453
heftig as modifier of *geraten*	5.8	174

	association using LogD	frequency
weltweit as modifier of *Aufregung*	3.51	16
Aufregung as object of *sorgen*	9.13	3774
international as modifier of *sorgen*	4.41	107

	association using LogD	frequency
richtig as modifier of *Geld*	5.07	241
Geld as object of *verdienen*	11.51	22226
richtig as modifier of *verdienen*	6.09	332

A comprehensive study of the collocational behaviour of these tuples would warrant a separate paper but, for now, the measures suffice to illustrate that the intuitively perceived bond between the components of the string is statistically verified. In a further study, the collocation of complex strings (e.g. *Geld verdienen*, *in die Kritik geraten*) with the modifier will also be measured.[8]

4 Function Composition for collocational selection

4.1 The spirit of Function Composition as a basis for the analysis

A solution to the syntax-semantics mismatch mentioned at the start of § 2 is the use of Function Composition (FC) instead of Functional Application to combine elements in syntax, cf. Jacobson (1990). Function Composition (FC) combines two functors to yield a new functor as sketched here:

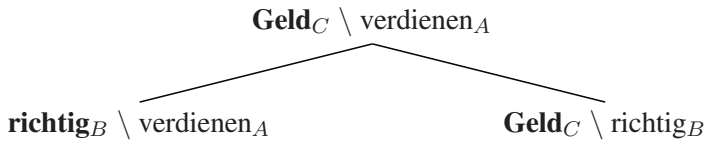
[8] Annelies Häcki Buhofer suggested to me that modification of *geraten* by *heftig* does not seem semantically likely (in contrast to cases such as e.g. *richtig verdienen*) and Kathrin Steyer suggested that the modifier should only be considered as a collocate of the whole FVG, e.g. *in Kritik geraten* 'get criticised'. This needs to be more closely examined although I note for now that in the DWDS Corpus *heftig* clearly also collocates (as a modifier) with other forms related to *geraten* such as *aneinandergeraten* 'clash with one another/come into contact with one another'. That the modifier's scope extends across the whole FVG or the whole N+V string follows from my analysis although I only actually encode the modifier as a modifier of a verb.

(3) Forward Function Composition: A/B $*$ B/C = A/C

(4) Backward Function Composition: B$\backslash$A $*$ C$\backslash$B = C$\backslash$A

Forward FC allows A/B to combine with B/C yielding A/C; a category requiring a C in order to be saturated. The *need* for a C at the initial level is postponed to the next level. Backward FC similarly postpones saturation (this time of A) to the next level. Within HPSG, Argument Inheritance draws on this type of combinatory rule, cf. Hinrichs & Nakazawa (1994) and much subsequent work on the licensing of verbal clusters. The spirit of Backward FC can be transferred to collocational cluster formation if we assume Backward FC can combine *richtig* + *Geld* (in boldface below), postponing the "requirement" for *verdienen* 'earn'. By "requirement" for *verdienen*, I am referring to the modification domain of *richtig*; the modifier is actually (informally speaking) looking for the verb to modify but combines syntactically with a different element first. FC yields a *special* instantiation of *Geld* which can syntactically combine with the modifier and yet still requires the verb, as sketched here:

(5)



Below, I will show how the spirit of this type of syntactic combination could be captured in HPSG through a combination of lexical entries and a schema that licenses the type of cluster I am arguing for. Since HPSG makes no division between lexicon and syntax in the sense that lexical entries of words and rules of syntactic combination (schemata) are stored together, cf. e.g. Müller (2013, p. 8), Jackendoff (2010, p. 19f), this means the analysis is lexical and we can think of these multi-word expressions as being lexically stored. It is conceivable that particularly frequently co-occurring material is **also** stored as a (ready-built) chunk or prefab as well, cf. the notion of *conventionalized collocation* and *prefabs* discussed by (Bybee, 2006, p. 713-4, 727). Cases in which elements of the lexical string are non-contiguously realized (e.g. in multiple fronting, partial topicalization etc.) probably then involve a schema, as we propose below, since individual atoms of the string are aligned non-adjacent to other atoms.

4.2 Lexical Entries

I will take the string *richtig Geld verdienen* 'to make heaps' as an example throughout. The lexical entry for *richtig* in its function as an intensifier is given here:

(6)
$$\begin{bmatrix} \textit{word} & \\ \textsc{phon} & \langle \textit{richtig} \rangle \\ \textsc{ss|loc} & \begin{bmatrix} \textsc{cat} & \begin{bmatrix} \textsc{head} & \begin{bmatrix} \textsc{mod} & V\boxed{\textsc{lid}\ \boxed{4}} \\ \textsc{lid} & \textit{richtig-intensifier} \end{bmatrix} \\ \textsc{subcat} & \langle\rangle \end{bmatrix} \end{bmatrix} \\ \textsc{cont} & \boxed{\textit{intensify}\ \boxed{4}} \\ \textsc{coll|lid} & \boxed{4}\textit{verdienen-idiomatic} \end{bmatrix}$$

I make use of the LID (lexical identifier) feature appropriate for the sort *head* to identify specific instantiations of words (Richter & Sailer (1999); Soehn (2004); Sag (2012); Spencer (2005)). Thus this word has the value $richtig_{intensifier}$ for the feature LID in its lexical entry. The COLL feature (which I take to be appropriate for the sort *word* and *cluster*) encodes in the lexical entry of a word (or cluster) that it collocates with (the LID value of) a particular word (cf. Sailer (2003), Richter & Sailer (1999)). I refer to this as *collocational selection*. Thus we see here that $richtig_{intensifier}$ collocates with the verb *verdienen* 'earn' (in its idiomatic instantiation). The intensifier is lexically encoded as a verb modifier (viz. the head feature MOD) and it also collocates with the verb it modifies (viz. the label 4 above). One could generalize the lexical entry so that the intensifier *richtig* always modifies the verb it collocationally selects if that turns out to be empirically correct.

I now give the lexical entry for *Geld* in the (semi-light-)verb phrase use:

(7)
$$\begin{bmatrix} \textit{word} & \\ \textsc{phon} & \langle \textit{Geld} \rangle \\ \textsc{ss|loc} & \begin{bmatrix} \textsc{cat} & \begin{bmatrix} \textsc{head} & [\textsc{lid}\ \textit{Geld-idiomatic}] \\ \textsc{subcat} & \langle\rangle \\ \textsc{spr} & \langle\rangle \end{bmatrix} \end{bmatrix} \\ \textsc{cont} & [\textsc{index}\ \textit{non-referential}] \\ \textsc{coll|lid} & \textit{richtig-intensifier} \end{bmatrix}$$

The idiomatic bare noun is lexically encoded as a collocate of the intensifier *richtig*. I am also assuming the noun is lexically specified as non-referential (this is certainly the case for nouns such as *Gas* in *Gas geben*) and cannot take a specifier (i.e. must be saturated). A separate lexical entry in which the value of COLL|LID is $verdienen_{idio}$ handles occurrences of the verb phrase *Geld verdienen* without *richtig*.[9]

The lexical entry for for the (semi-)light verb *verdienen* 'earn' is given next, below. The lexical entry would also be structured in the same way for a less-compositional (clearly) light verb such as *geben* 'give' (in e.g. *richtig Gas geben*):

[9] In the case of the bare-noun strings, we find frequent data such as the following which I think support the claims about collocation of *richtig* with *Geld* and *richtig* with *verdienen*:

(i) *er hat* ***richtig Geld*** 'he is really rich' and

(ii) *er* ***verdient richtig*** 'he earns loads'

That the existence of such frequent strings (in particular the first one) facilitates apparent multiple fronting constructions seems highly plausible but remains to be further studied.

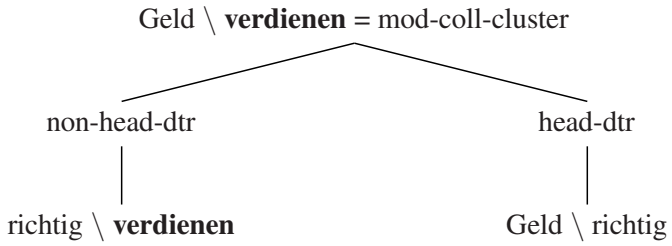
(8)
$$\begin{bmatrix} \textit{word} & \\ \text{PHON} & \langle \textit{verdienen} \rangle \\ \text{SS|LOC} & \left[\text{CAT} \left[\begin{array}{ll} \text{HEAD} & \left[\text{LID} \ \textit{verdienen-idiomatic} \right] \\ \text{SUBCAT} & \langle \text{NP-}\textit{nom}\ \boxed{5} \rangle \end{array} \right] \right] \\ \text{CONT} & \left[\text{RELS} \left\langle \left[\begin{array}{l} \textit{earn heaps} \\ \text{AGENT}\ \boxed{5} \end{array} \right] \right\rangle \right] \\ \text{COLL|LID} & \textit{Geld-idiomatic} \end{bmatrix}$$

This constitutes a new HPSG treatment of light verb phrases (or *Funktionsverbgegfüge*, FVG) in German.[10]The light verb *collocationally selects* (not subcategorizes) the (athematic) object but selects the subject NP in the normal way via SUBCAT. That this verb cannot undergo personal passive in the idiomatic use follows from the fact that there is no regular thematic argument other than the subject. Of course, impersonal passive (a subjectless construction, always requiring 3rd person singular verbal morphology, in German) is possible as we see in examples (1c-d) above). I am thus analyzing object-verb collocations involving bare nouns (e.g. *Geld verdienen* 'earn money') on a par with light verb phrases. I believe that this analysis can be extended to account for integrated objects (in the sense of Jacobs (1993, 1999) but cannot go into details here.

4.3 The *modifier-collocational-cluster* schema

In this section, I introduce the schema which licenses the *modifier-collocational-clusters* such as e.g. *richtig Geld* [lit. real money]. The composition of the *modifier-collocational-cluster*, e.g. *richtig Geld*, with the idiomatic verb *verdienen* will then be shown next. First, recall the spirit of Backward FC. I indicate here the structures that I will be assuming in the schema. In particular, I assume here that the modifier is the non-head daughter and the bare noun is the head daughter of the cluster. This essentially translates the notion of headword or node from the lexicographically-oriented approach to collocation I mentioned earlier.

(9)



The modifier *richtig* has to 'wait' until it finds the verb it modifies. What I analyze as a *modifier-collocation-cluster* is therefore not a type of *head-adjunct-structure*

[10]Recent work on the processing of light verb phrases indicates increased processing load at the verb in light verb phrases. This effect can be interpreted as providing evidence that light verb phrases are not stored as complex (phrasal) entries but, rather, require some kind of syntactic combination or, perhaps, some operation involving argument-structure merging, cf. Wittenberg & Piñango (2011), Wittenberg et al. (2014). I believe my analysis is in keeping with these findings.

(since semantic modification does not occur here). This makes sense because the tuple is, I believe, in fact more like one complex lexeme. In fact, *richtig* in its intensifier function is not a normal modifier but somewhere between modifier and argument (= a collocational modifier). The *modifier-collocational-cluster* schema given below captures this:

(10) *modifier-coll-cluster* →

$$\begin{bmatrix} \text{SS} & \begin{bmatrix} \text{LOC}|\text{CAT}|\text{HEAD} \begin{bmatrix} \text{MOD } \boxed{4} \\ \text{LID } \boxed{1} \end{bmatrix} \\ \text{COLL}|\text{LID } \boxed{4} \\ \text{CONT } \boxed{6} \end{bmatrix} \\ \text{NON-HD-DTR} & \begin{bmatrix} \text{SS}|\text{LOC}|\text{CAT} \begin{bmatrix} \text{HEAD}|\text{MOD } \boxed{4} \\ \text{LID } \boxed{2} \end{bmatrix} \\ \text{COLL}|\text{LID } \boxed{4} \end{bmatrix} \\ \text{HD-DTR} & \begin{bmatrix} \text{SS}|\text{LOC}|\text{CAT} \begin{bmatrix} \text{HEAD}|\text{LID } \boxed{1} \\ \text{SPR } \langle\rangle \end{bmatrix} \\ \text{COLL}|\text{LID } \boxed{2} \\ \text{CONT } \boxed{6} \end{bmatrix} \end{bmatrix}$$

The HD-DTR, e.g. *Geld*, collocationally selects the NON-HD-DTR, e.g. *richtig*. The NON-HD-DTR collocationally selects the verb it modifies. At the cluster level, the mother inherits the COLL|LID and MOD values from the NON-HD-DTR (= the postponement mentioned above). At the cluster level, the mother also inherits the CONT value of the HD-DTR; in keeping with the Semantics Principle. The cluster (mother) inherits the LID value from HD-DTR. In this way, the cluster can be seen as a special version of the lexeme *Geld-idiomatic*. The sub-tree for *richtig Geld* licensed by the *modifier-collocation-cluster* schema is given in Figure 1 on the following page.

The remaining question now is how the verb, in our case *verdienen* 'earn', combines with the collocational cluster *richtig Geld* 'real money'. In fact, the *modifier-collocation-cluster richtig Geld*, headed by *Geld*, collocationally selects the (idiomatic) verb *verdienen* but it also selects it via MOD. The *mod-coll-cluster* and the verb can combine via the normal *head-adjunct-schema*. The idiomatic semantics of the verb (encoded at the HD-DTR) percolate to the mother node and the (postponed) semantic modification of the verb can apply. The sub-tree for the combination of the cluster *richtig Geld* and *verdienen* is given in Figure 2 on the next page.

5 Extensions and further work

The analysis I have sketched here can, hopefully, be extended to handle a bigger range of data which behave similarly to those discussed here. In particular, there are certain lexical strings which offer an open slot which can be instantiated not just by lexically specified (collocating) material but which is, rather, open for any

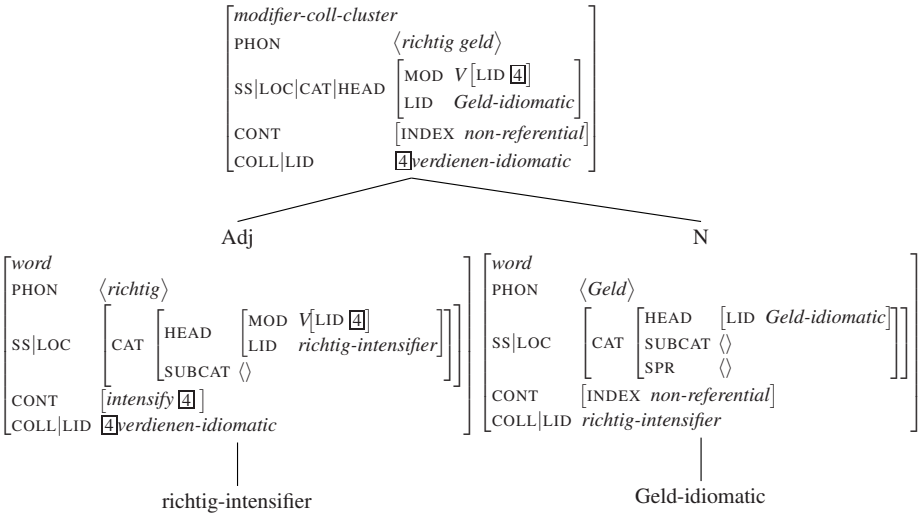


Figure 1: Sub-tree for the *modifier-collocational-cluster richtig Geld*

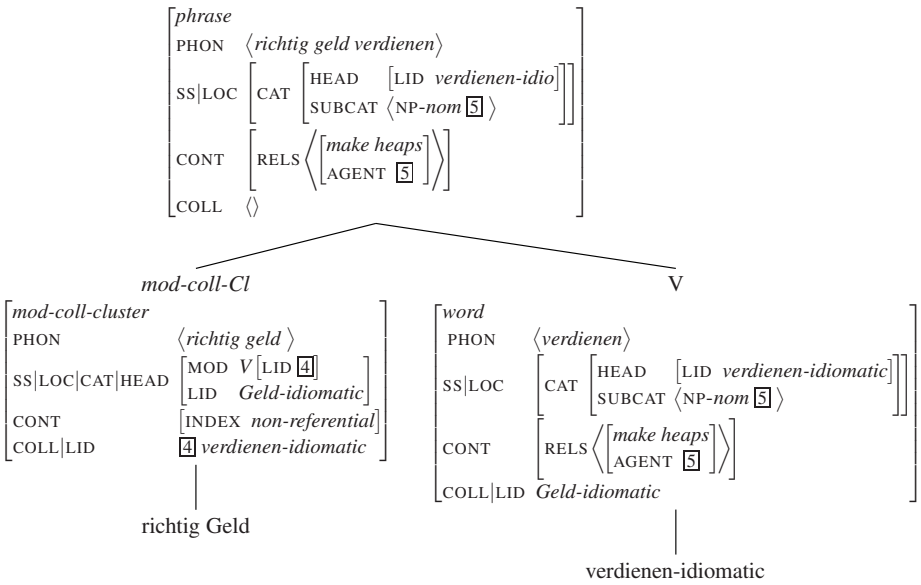


Figure 2: Sub-tree for the combination of the *modifier-collocational-cluster richtig Geld* with the verb *verdienen*

material of a particular class. For instance, the strings in the table below have all been attested with multiple fronting but range from fixed idioms, through collocating strings such as those discussed here to strings with slots for directional prepositional phrases, for instance. The strings vary in degrees of schematicity and form a continuum from full idioms to near-compositional phrases.

Licht	*ins Dunkel*	*bringen*	'bring light into the dark = shed light onto sth.'
richtig	*Gas*	*geben*	'really give Gas = increase effort'
hart	*ins Gericht*	*gehen*	'go hard into court = roast s.o.'
ihm	*zur Seite*	*stehen*	'stand by him'
am billgsten	*in XP*	*kommen*	'get to X the cheapest (way)'
trocken	*durch XP*	*kommen*	'come dry through X'
postiv/negativ	*auf XP*	*wirken*	'react positively/negatively to X'

It remains to be fully worked out how the range of data can be accommodated in the type of analysis proposed here.

A different consequence of the analysis proposed here concerns the possibility of topicalization of the collocational clusters for which I am arguing. It now seems plausible that this could be handled analogously to fronting of coherent verbal clusters, as in (11b), and could potentially offer an alternative analysis for (some) multiple fronting constructions:

(11) a. [**richtig Gas**] gibt er immer
right gas gives he always

b. [**zu schlafen versucht**] hat er
to sleep try has he
'he tried to sleep'

Just as a string *zu schlafen versucht*$_{verbal-cluster}$ can be realized in initial position, so could potentially a string *richtig Gas*$_{modifier-collocational-cluster}$. In fact, it is interesting to note that the availability of cluster formation discussed here could well be closely related to the availabilty of cluster-formation more generally in a given language (i.e. languages allowing verbal clusters may well be languages that allow other kinds of clusters too).

A futher fascinating area is the extension of the current analysis to also cover (free) datives in the prefield, as in examples such as (12) where we have a free dative together with a PP belonging to a light verb phrase in the prefield:[11]

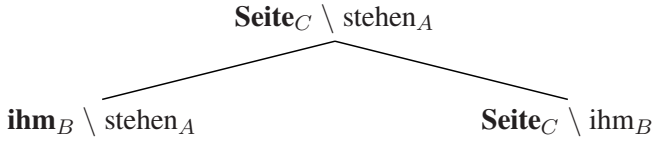
(12) [**Ihm**] [**zur Seite**] steht als stellvertretender Vorstandschef Gerd
he-DAT to.the side stands as acting ceo Gerd

[11] Examples of apparent *multiple fronting* invloving a dative an an accusative object in the prefield are extremely rare. An analysis of argument-clusters as proposed by Mouret (2006) for co-ordination structures might be relevant but it would certainly over-generate as it stands since the material that can occur in the multiple fronting data is lexically very restricted.

Tenzer[12]
Tenzer
'Gerd Tenzer is helping him out as acting CEO'

One can treat the dative as a benefactive modifier, addable to the argument-structure of any verb in German (e.g. by lexical rule). The dative is, however, also concomitantly possessor of the noun *Seite* 'side'; i.e. it is also a modifier of the type which I assume to be introducable into the argument-structure of any noun. I informally sketch here how the FC-style analysis could be extended to cover such data:

(13)



6 Conclusion

I have argued here for a new type of cluster in German; a *modifier-collocation-cluster*. Clearly, we must extend the part-of-speech hierarchy accordingly to accommodate such elements. I believe introducing this type of cluster is a justified step, though. The analysis presented here has significance for our ideas about constituency and how it interacts with usage/frequency information, cf. Bybee & Cacoullos (2009); Beckner & Bybee (2009); Bod (1998), and also for the issue of the modifier-argument distinction. We know there is a close relation between frequently co-occurring elements and standard constituents but we must also capture units beyond those standardly acknowledged up to now, I firmly believe. Collocationally selected modifiers are situated inbetween arguments and true modifiers. The availability of what I have treated as collocationally selected items seems to generalize to form a pattern, to provide a slot fillable by material of a certain grammatical class (cf. Dowty (2003). An extension of the current analysis to handle this kind of phenomenon is an exciting prospect. The analysis has, moreover, certain advantages for HPSG and specifically for the analysis of German. It interfaces usage data and a usage-based view of 'constituency' with the HPSG formalism. Further, it begins to capture the analogy between verb clusters (cluster – chunk) and the (non-standard) constituents for which I have argued in German. With some additional modification, it also offers offers the basis for a syntactic solution for handling Integration of nouns and PPs as discussed in (Jacobs, 1993, 1999).

References

Beckner, Clay & Joan Bybee. 2009. A usage-based account of constituency and reanalysis. *Language Learning* 59. 27–46.

[12]taz, 18.07.2002, p. 7, taken from Müller (2005)

Bod, Rens. 1998. *Beyond grammar. an experience-based theory of language*. Stanford: CSLI Publications.

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

Bybee, Joan L. & Rena Torres Cacoullos. 2009. The role of prefabs in grammaticization: How the particular and the general interact in language change. In Hamid Ouali Roberta L. Corrigan, Edith A. Moravcsik & Kathleen Wheatley (eds.), *Formulaic language: Volume 1. distribution and historical change*, 187 – 217. Amsterdam: John Benjamins.

Didakowski, Jörg & Alexander Geyken. 2013. From DWDS corpora to a German word profile – methodological problems and solutions. In *Network strategies, access structures and automatic extraction of lexicographical information. 2nd work report of the academic network "internet lexicography"* OPAL - Online publizierte Arbeiten zur Linguistik X/2012, 43 – 52. Mannheim: Institut für deutsche Sprache.

Dowty, Davis. 2003. The dual analysis of adjuncts and complements in categorial grammar. In Ewald Lang, Claudia Maienborn & Cathrine Fabricius-Hansen (eds.), *Modifying Adjuncts*, 1 – 22. de Gruyter.

Evert, Stefan. 2008. Corpora and collocations. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics. an international handbook. article 58*, 1212 – 1248. Berlin: Mouton de Gruyter.

Firth, J. R. 1957. A synopsis of linguistic theory 1930–55. In *Studies in linguistic analysis*, 1 – 32. Oxford: The Philological Society.

Hinrichs, Erhard W. & Tsuneko Nakazawa. 1994. Linearizing AUXs in German verbal complexes. In Nerbonne et al. (1994) 11–38.

Jackendoff, Ray. 2010. *Meaning and the Lexicon. The Parallel Architecture 1975–2010*. Oxford: Oxford University Press.

Jacobs, Joachim. 1993. Integration. In Marga Reis (ed.), *Wortstellung und Informationsstruktur*, 63–116. Tübingen: Niemeyer.

Jacobs, Joachim. 1999. Informational Autonomy. In Peter Bosch & Robert van der Sandt (eds.), *Focus. Linguistic, cognitive and computational perspectives*, 56–81. Cambridge University Press.

Jacobson, Pauline. 1990. Raising as function composition. *Linguistics and Philosophy* 13. 423–475.

Krenn, Brigitte & Gregor Erbach. 1994. Idioms and support verb constructions. In Nerbonne et al. (1994) 365–396.

Mouret, François. 2006. A phrase structure approach to argument cluster coordination. In *Proceedings of the 13th international conference on Head-Driven Phrase Structure Grammar, Bulgarian Academy of Sciences, Bulgaria*, 247–267. Stanford: CSLI Publications. http://cslipublications.stanford.edu/HPSG/2009/.

Müller, Stefan. 2003. Mehrfache Vorfeldbesetzung. *Deutsche Sprache* 31(1). 29–62.

Müller, Stefan. 2005. Zur Analyse scheinbarer V3-sätze. In Franz-Josef d'Avis (ed.), *Deutsche Syntax: Empirie und Theorie. Symposium Göteborg 13–15 Mai 2004* (Göteborger Germanistische Forschungen 46), 173–194. Acta Universitatis Gothoburgensis.

Müller, Stefan. 2013. Unifying everything. *Language* 89(4). 920–950. http://hpsg.fu-berlin.de/~stefan/Pub/unifying-everything.html.

Nerbonne, John, Klaus Netter & Carl J. Pollard (eds.). 1994. *German in Head-Driven Phrase Structure Grammar* (CSLI Lecture Notes 46). Stanford: CSLI Publications.

Richter, Frank & Manfred Sailer. 1999. LF conditions on expressions of ty2: An HPSG analysis of negative concord in Polish. In Robert D. Borsley & Adam Przepiórkowski (eds.), *Slavic in Head-Driven Phrase Structure Grammar*, 247 – 282. Stanford: CSLI.

Rychlý, Pavel. 2008. A lexicographer-friendly association score. In *Raslan 2008*, 6–9. Brno: Masarykova Univerzita.

Sag, Ivan. 2012. Sign-based construction grammar: An informal synopsis. In Hans Boas & Ivan Sag (eds.), *Sign-based construction grammar*, 69 – 202. Stanford: CSLI.

Sailer, Manfred. 2003. *Combinatorial semantics and idiomatic expressions in Head-Driven Phrase Structure Grammar*. Working Papers of the SFB 340, Vol. 161: Eberhard-Karls-Universität Tübingen Dissertation (2000).

Sinclair, John. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair, John. 1996. Beginning the study of lexis. In C. E. Bazell, J. C. Catford, M. A. K. Halliday & R. H. Robins (eds.), *In memory of J. R. Firth*, 410 – 430. London: Longman.

Soehn, Jan-Philipp. 2004. License to coll: How to bind bound words and readings to their contexts. In Stefan Müller (ed.), *Proceedings of the 11th international conference on Head-Driven Phrase Structure Grammar, Center for Computational Linguistics, Katholieke Universiteit Leuven*, 261 – 273. Stanford: CSLI.

Spencer, Andrew. 2005. Generalized paradigm function morphology – a synopsis. In Alexandra Galani & Beck Sinar (eds.), *York papers in linguistics*, vol. 2 Papers from the York-Essex Morphology Meeting, 2003, 93 – 106. University of York, UK.

Steinitz, Renate. 1989. Überlegungen zum Prädikativ. In Wolfgang Motsch (ed.), *Wortstruktur und Satzstruktur*, 211 – 234. Berlin: Zentralinstitut für Sprachwissenschaft.

Wittenberg, Eva, Martin Paczynski, Heike Wiese, Ray Jackendoff & Gina Kuperberg. 2014. The difference between "giving a rose" and "giving a kiss": Sustained neural activity to the light verb construction. *Journal of Memory and Language* 73. 31–42.

Wittenberg, Eva & Maria Piñango. 2011. Processing light verb constructions. *Mental Lexicon* 6(3). 393–413.

The polyfunctionality of Coptic Egyptian relative complementisers

Berthold Crysmann
CNRS, Laboratoire de linguistique formelle

Chris H. Reintges
CNRS, Laboratoire de linguistique formelle

Proceedings of the 21st International Conference on
Head-Driven Phrase Structure Grammar

University at Buffalo

Stefan Müller (Editor)

2014

Stanford, CA: CSLI Publications

pages 63–82

Crysmann, Berthold & Chris H. Reintges. 2014. The polyfunctionality of Coptic Egyptian relative complementisers. In Stefan Müller (ed.), *Proceedings of the 21st International Conference on Head-Driven Phrase Structure Grammar, University at Buffalo*, 63–82. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2014.4.
CC-BY 4.0

Abstract

The present study is concerned with the complex ways in which alternating relative complementisers in Coptic are employed as a morphological flagging device for unbounded dependencies in various types of relative clause constructions and wh questions. We shall argue in particular that the alternation in shape is locally conditioned by properties of the complement (TAME) and the antecedent noun (definiteness), which can be modelled via selectional features such as COMPS and MOD, plus the prosodic status of right-adjacent material (phrase vs. clitic). We shall show that all applicable conditions carry over from relatives to wh in-situ, suggesting to model the polyfunctionality of these complementisers in terms a systematic alternation between resumptive SLASH and in-situ QUE dependencies, modelled in terms of a lexical rule.

Furthermore, we shall discuss the status of unbounded dependencies and argue that the pervasiveness of resumption with relatives and ex-situ wh arguments can be attributed to the absence of *gap-synsem* on ARG-ST. We shall argue that apparent subject "gaps" in relative constructions are of a highly local nature, best to be understood in terms of subcategorisation for a finite VP complement. Finally, we shall show that the ban on argument gaps does not carry over to wh ex-situ adjuncts, providing additional motivation for maintaining a systematic distinction between these two types of extraction.

1 Typological properties of Coptic

Coptic is the vernacular of Late-Antique and Early Medieval Egypt and represents the most recent stage of Ancient Egyptian [Afroasiatic] (from around the 3rd to the 13th c. CE). The language consist of at least six regional varieties, two of which gained supra-regional importance: Sahidic (from Arabic ʔaṣ-Ṣaʕīd 'Upper Egypt') and Bohairic (from Arabic ʔal-Buhairā, a province southeast of Alexandria), the latter of which functions as the liturgical language of the Coptic Orthodox Church (for dialect variation, history, and genetic affiliation, see Layton (2000, 1–4 §§1–6) and Reintges (2004, 2–6 §0.1)). All data are taken from corpora of the classical Sahidic dialect, which is renowned for its rich literary sources.

In terms of a coarse-grained morphological typology, the language falls near the isolating pole of the analytic–synthetic dimension. The language's basic word order is Subject-Verb-Object. Tense-Aspect-Mood-Evidentiality (TAME) particles furnish a broad range of conjugation patterns, in which lexical verbs can appear. TAME markers fall into two positional classes of pre-subject and pre-verbal (=post-subject) particles. The perfect tense particle *a* in (1) precedes the subject, whereas the pre-verbal future tense particle *na* in (2) follows it.

(1) a tə=sophia ket u=ɛːï na=s
PERF DEF.F.SG=wisdom build INDEF.SG=house for=3F.SG

†We are gratefully indebted to the audience of the HPSG 2014 conference for their insightful and stimulating comments, in particular to Anne Abeillé, Doug Arnold, Bob Borsley, Matthew Dryer, Ray Jackendoff, and Philip LeSourd. In particular the discussions with Bob Borsley have been most helpful towards sharpening and clarifying the ideas presented here.

‘Wisdom has built a house for herself.’ (Proverbs 9, 1)

(2) pə=tʃɔeis na krine ən=nə=laos
DEF.M.SG=lord FUT judge PREP=DEF.PL=people
‘The Lord will judge the nations.’ (Psalm 28, 11)

The language has two negation strategies. The first strategy is to use the double negation *ən ... an*, where the negative scope marker *ən* is often omitted. The second strategy is to use a negative TAME particle in which negative polarity and a given temporal, aspectual or modal semantics are fused into a single, non-segmentable morph.

(3) a. arɛu əm pə=sɔn tɛt ən=hɛt an e=ʃatʃe
perhaps NEG DEF.M.SG=brother persuade.STAT of=heart not to=talk
nəmma=n
to=1PL
‘Perhaps the brother is not persuaded of heart to talk to us.’ (Apophthegmata Patrum, ed. Chaîne, n° 238 70, 21)

b. nə=f=na mu: an e=mpe=f nau e=pe=khristos
NEG=3.M.SG=FUT die not REL=NEG.PERF see PREP=DEF.M.SG=Christ
əm=pə=tʃɔeis
LINK=DEF.M.SG=lord
‘He will not die without having seen Christ, the Lord.’ (Luke 2, 26)

(4) awo: əmpe pə-kosmos sɔ:won=f
and NEG.PERF DEF.M.SG-world recognise=3m.sg
‘And the world did not recognise him (the Christ).’ (John 1, 10)

Coptic is a language without agreement inflection on the verb. Person, number, and gender marking on TAMEs, verbs and prepositions can be identified with enclitic subject and object pronouns, respectively, which appear in the same surface position as full NPs with which they are in complementary distribution. Moreover, pronominal arguments must always overtly be expressed; i.e. there is no pro-drop (Reintges, 2004).

2 Relative clauses

2.1 Converbal vs. canonical relative clauses

Coptic has a rich system of specialised syntax and morphology for relative constructions of various kinds. The two major relativisation strategies are represented by converbal and canonical relative clauses, which differ from each other in the range of antecedents that they can take. Converbal relative clauses typically modify indefinite and universally quantified NPs. In providing information necessary to establish the identity of the antecedent or to narrow down the set of potential referents, they can only be used as restrictive modifiers.

(5) a. ən=tə=he gar ən=u=roːme [e=fi na
in=DEF.F.SG=manner PCL LINK=INDEF.SG=man REL=3M.SG FUT
apodɛːmei]
go.abroad
'For like a man who is about to go abroad' (Matthew 25, 14)

b. roːme gar nim [e=wənta=f hah ən=nuːte]
man PCL every REL=HAVE=3M.SG many LINK=god
'For every man who has many gods' (Eudoxia, ed. Orlandi, 36, 11)

The complementary relativisation pattern features definite antecedents. In contrast to converbal relatives, canonical relative clauses have restrictive as well as non-restrictive uses. In the latter case, they are used as parenthetical assertions that provide supplementary information about a contextually given referent.

(6) a. pə=houː [ənt a=uː tʃpo=k ənhɛːtə=f]
DEF.M.SG=day REL PERF=3PL deliver.INF=2M.SG within=3M.SG
'The day on which you were born (lit. they gave birth to you)' (Koptische und Heiligen- und Martyrerlegenden, ed. Till, II 30, 13)

b. pə=houː əm-pə-hap [etere pə-tʃɔeis na ti hap
DEF.M.SG-day LINK-DEF.M.SG-law REL DEF.M.SG-lord FUT give law
ero=k]
to=2M.SG
'The day of the judgement when the Lord will judge you' (Acts of Andrew & Paul, ed. Jacques, 202,128)

2.2 Complementiser allomorphy

Besides their distributional differences, canonical and converbal relatives can also be distinguished on a morphological basis in terms of context-sensitive alternations in the shape of the relative complementiser. The language recognises five distinct relative complementisers *e, ere, et, ənt*, and *ən*, all of which show a morphosyntactic behaviour distinct from run-of-the-mill subordinate conjunctions such as *tʃe* 'that' (Reintges, 2012).

The converbal marker comes in two variants, the short form *e* and the long form *ere*. The distribution between the two allomorphs is relatively straightforward: the base form *e* is selected when the converbal marker is adjacent to an enclitic subject pronoun or a TAME marker, while the long form *ere* is selected when it is followed by a full NP subject. Given the syntactically heterogeneous character of the elements triggering the short form, we shall conclude that the distribution of *e* vs. *ere* is best understood in terms of a distinction between lexical head vs. full phrasal constituents, which is ultimately related to the presence vs. absence of a prosodic phrase boundary.

(7) a. hən u=ma [e=f ɔː ən=ʃarβa]
in INDEF.SG=place REL=3M.SG be.STAT in=scorching.heat
'In a place which (is) in (a state of) scorching heat' (Sahidic Vita of St. Pachomius, ed. Lefort 86, 24–25)

b. laau ən=ʃɛn nim [e=a=f tʃɔ=uː]
something LINK=tree every REL=PERF=3M.SG plant=3PL
'Every (single) one of the trees that he planted' (Koptische Heiligen- und Martyrerlegenden, ed. Till, II 18, 23–24)

c. hən uː=houː [e=nə=f sowən əmmɔ=f an]
in INDEF.SG=day REL=NEG=3M.SG know PREP=3M.SG not
'On a day which he does not know' (Luke 12, 46)

(8) u=hoβ [ere pə=nuːte moste əmmɔ=f]
INDEF.SG=thing REL DEF.M.SG=god hate PREP=3M.SG
'A thing which God hates' (Acts of Andrew & Paul, ed. Jacques, 202, 126–127)

In contrast to converbal relatives, canonical relative clauses display a considerable degree of complementiser allomorphy, which varies along with the TAME particle and the polarity of the embedded relative clauses. In affirmative relative clauses, alternating relative complementisers encode a rudimentary [± past] distinction, which reflects only partially the tripartite present–past–future tense system of the language. The relative complementiser *et* is selected in canonical present and future tense relatives and the allomorph *ənt* in canonical past tense relatives with the perfect tense particle *a*.

(9) a. etβe te=uː=pistis [et tʃɛk eβɔl]
because.of DEF.F.SG=3PL.POSS=faith REL accomplish.STAT PCL
'Their faith, which is accomplished' (Testament of Isaac, ed. Kuhn, 233, 19)

b. t=apophasis [et na ʃoːpe]
DEF.F.SG=verdict REL FUT happen
'The verdict that will be reached' (Shenoute, ed. Amélineau, I.2 178, 14)

c. t=irɛːnɛː əm=pa=tʃɔeis [ənt=a=f
DEF.F.SG=peace LINK=DEF.M.SG.1SG.POSS=lord REL=PERF=3M.SG
taa=s na=i]
give=3F.SG to=1SG
'The peace of My Lord that he has given to me' (Testament of Isaac, ed. Kuhn, 230, 10–11)

The binary [±past] distinction that we see with affirmative relative clauses does not carry over to the corresponding negated relatives, which are consistently marked by the complex relative complementisers *ete(re)*, regardless of the negation strategy that is employed.

(10) a. nə=hethos [ete=n=se pɛt an ənsa tə=dikaiosynɛː]
DEF.PL=gentile REL.DEF=NEG=3PL run.STAT not after DEF.F.SG=justice
'The gentiles who did not pursue justice' (Romans 9, 30)

b. ʃɛn nim [ete=nə=f na ti karpos an [e=nanuː=f]]
tree every REL.DEF=NEG=3M.SG FUT give fruit not REL=be.good=3M.SG
'Every tree, which will not give good fruit (lit. fruit which is not good)' (Luke 3, 9)

c. nai [ete=mpe hoine mate əmmɔ=uː]
DEM.PL REL.DEF=NEG.PERF some obtain PREP=3PL
'These (things) which some have not obtained' (I Timothy 1, 6)

Converbal relative clauses are characterised by a generalised resumptive pronoun strategy, in which a personal pronoun replaces the relativised subject, direct object or oblique NP constituent.

(11) a. rɔːme nim [e=f hitʃəm pə=kah]
man every REL=3M.SG on DEF.M.SG=earth
'Every man who lives on earth' (Testament of Isaac, ed. Kuhn, 233, 12)

b. laau ən=ʃɛn nim [e=a=f tʃɔ=uː]
something LINK=tree every REL=PERF=3M.SG plant=3PL
'Every (single) one of the trees that he planted' (Koptische Heiligen- und Martyrerlegenden, ed. Till, II 18, 23–24)

c. ma nim [e=uː na tʃɔuː=s ero=f]
place every REL=3PL FUT send.=3PL to=3M.SG
'Every place that they will be sent to' (Precepts of St. Pachomius, ed. Kuhn, no. 129)

The generalised resumption strategy carries over to canonical past relatives introduced by the complementiser *ənt* (Reintges, 2012).

(12) a. ne=kʲom men ne=ʃpɛːre [ənt=a=uː ʃoːpe eβol
DEF.PL=wonder with DEF.PL=miracle REL=PERF=3PL exist PCL
hi=tootə=f əm=pe=n=eiɔt Apa Matheos]
by=hand=POSS.3M.SG PREP=DEF.M.SG=POSS.1PL=father Apa Matthew
'The miracles and wonders that came about through the agency of Our Father Matthew' (Koptische Heiligen- und Martyrerlegenden, ed. Till, II 18, 14–16)

b. pə=hoβ [ənt=a pə=nuːte kjalɔ=f ero=n]
DEF.M.SG=thing REL=PERF DEF.M.SG=god entrust=3M.SG to=1PL
'The matter that God entrusted (it) to us' (Shenoute, ed. Amélineau, I.1 36, 5)

c. e=pə=ma [ənt=a=k kʲəntə=f ənhɛːtə=f]
to=DEF.M.SG=place REL=PERF=2M.SG find=3M.SG inside=3M.SG
'The place where you found it' (Acts of Andrew & Paul, ed. Jacques, 204, 145–146)

Relativisation in Coptic Egyptian involves a non-local dependency between the antecedent and the resumptive element, mediated by the relative complementiser. As illustrated in (13), relativisation out of an embedded (conjunctive) clause is indeed attested, showing that the dependency is clearly not clause-bound.

(13) nim pe pei-ke-wa [et hən te=tən-mɛte]
who COP.M.SG DEM.M.SG-other-one REL in DEF.F.SG=POSS.2PL-midst
[et<e>=mp=ei əmpəʃa [n=f ʃatʃe nəmma=i]]
REL=NEG.PERF=1SG be.worthy CONJ=3.M.SG speak with=1SG
'Who is this other one who is in your midst that I was not worthy that he speaks with me?' (Koptische Heiligen und Martyrerlegenden, ed. Till, II 30, 18-19)

Coptic recognises one construction where an apparent gap is found inside the relative clause: when introduced by the complementiser *et*, the relativised subject remains unexpressed. However, in contrast to the other relative complementisers, a subject relative marked by *et* is of a highly local nature: as shown by the data in (14) above, use of *et* is only possible, if the complementiser is immediately followed by either the lexical verb, or the post-subject future *na*.[1]

(14) a. etβe te=uː–pistis [et tʃɛk eβɔl]
because.of DEF.F.SG=3PL.POSS–faith REL.DEF accomplished PCL
'Their faith, which is accomplished' (Testament of Isaac, ed. Kuhn, 233, 19)

b. t=apophasis [et na ʃoːpe]
DEF.F.SG=verdict REL.DEF FUT happen
'The verdict that will be reached' (Shenoute, ed. Amélineau, I2 178, 14)

The complex complementiser *ete(re)* must be used in non-subject present and future tense relatives, which are characterised by the presence of a resumptive pronoun for the relativised argument.

(15) a. pə=ʃatʃe [etere pə=rəm-ɛːi na
DEF.M.SG=word REL.DEF DEF.M.SG=AGENT.NOUN-house FUT
tʃɔɔ=f]
say.INF=3M.SG
'The word that the superintendent will speak' (Precepts of St. Pachomius, ed. Kuhn, no. 122)

[1] Besides future *na* Coptic witnessed two more raising TAM markers, i.e. conditional *šan* and deontic modal *e*. However, these two markers are not attested in side relative clauses.

b. p=ɛːi [etere pei=ʃɛːre ʃɛm mɔwət ənhtə=f]
DEF.M.SG=house REL.DEF DEM.M.SG=boy little die.STAT in=3M.SG
'The house in which the young boy died' (Acts of Andrew & Paul, ed. Jacques, 206, 163–164)

Furthermore, if a pre-subject TAM auxiliary or a negative marker is present, use of a resumptive is again obligatory, together with one of the standard non-local relative complementisers *ənt* or *ete(re)*, as shown in (10) above.

Given the highly local nature of zero subjects following *et*, together with the general absence of argument gaps in the language, the Coptic data are of high significance for a general theory of resumption, ultimately providing evidence against a conception of resumption as a "last resort" operation (Shlonsky, 1992).

3 Wh questions

3.1 Wh in-situ constructions

Alternating relative complementisers are not restricted to relative clauses but may also appear in various non-relative environments, such as yes/no and wh questions, declarative focus sentences, coordinate structures, comparative constructions, predicative adjunct, temporal adverb clauses, conditionals and so on. The concern here is with Wh questions. As shown by the contrast between (16a) and (16b), clause-internal interrogative pronouns such as *nim* 'who' and *uː* 'what' only assume a genuine question interpretation, when they are construed with an initial relative complementiser; otherwise they are interpreted as specific indefinites in an ordinary declarative clause. In other words, the presence of a relative complementiser is crucially involved in specifying the interrogative force of the wh in-situ construction (Reintges et al., 2006; Reintges, 2007).

(16) a. e=i na ti uː na=k ?
REL=1SG FUT give what to=2SG.M
'What shall I give you?' (Genesis 30, 31)

b. a=i ti uː mən uː ehun e=pei=ma
PERF=1SG give what and what PCL to=DEM.M.SG=place
'I gave such and such a thing to this place.' (Shenoute, ed. Leipold, IV 105, 16)

Wh in-situ has a broad syntactic distribution, appearing in main and embedded clauses, introduced in the latter case by the finite subordinating complementiser *tʃe* 'that'.

(17) a. ənt=a uː ʃoːpe əmmɔ=k pa=tʃɔeis p=ərrɔ?
REL=PERF what become PREP=2M.SG DEF.M.SG.1SG=lord DEF.M.SG=king
'What happened to you, my Lord and King?' (Eudoxia, ed. Orlandi, 36, 24)

b. ən=ti sɔwən an [tʃe ənt=a u: ʃo:pe əmmɔ=s]
NEG=1SG know not that REL=PERF what become PREP=3F.SG
'I do not know what has happened to her.' (Hilaria, ed. Drescher, 7, 30-31)

Multiple wh in-situ questions are only marginally attested and display a pair-listing reading, in the next example, pairs of informers and informs.

(18) ənt=a nim tsaβɔ=f e=nim?
REL=PERF who teach=3M.SG about=who
'Who taught him about whom?' (Shenoute, Wessley 9, 110a: 9f)

Neither wh arguments nor wh adverbs show any resistance to wh in-situ interrogation.

(19) a. ere nim na na na=n?
REL who FUT have.mercy for=1PL
'Who will have mercy upon us?' (Shenoute, pap. Paris 13154v, a14)

b. e=i na tʃe u: na=k?
REL=1SG FUT say what to=2M.SG
'What shall I say to you?' (Apophthegmata Patrum, ed. Chaîne, n°28 5, 25)

c. awo: ənt=a=u: ei eβɔl ton?
and REL=PERF=3PL come PCL where
'From where did they come?' (Apocalypse 7, 13)

d. ənt=a=k ei e=pei=ma ən=aʃ ən=he?
REL=PERF=2SG.M come to=DEM.SG.M=place in=what of=manner
'How did you get to this place?' (Coptic Martyrdoms, ed. Budge, 206, 29)

It is also possible, although not very common, to have wh in-situ in negated questions.

(20) ete=mpe tʃo:həm hən aʃ əm=ma?
REL.DEF=NEG.PERF.2F.SG defile in what of=place
'In which place have you not become defiled?' (Besa fragment 35 116, 14-15)

Present tense and future tense wh in-situ questions are introduced by the converbal relative markers *e(re)*, while affirmative and negative past tense wh in-situ questions are marked by the relative complementisers *ənt* and *ete(re)*, respectively and pattern in this respect with canonical relative clauses. A question arises with respect to the scope of the wh in-situ constituent in embedded clauses. As shown by (17b), the in-situ wh word generally takes the embedded scope, which produces an indirect question interpretation. In this context, the relative complementiser surfaces

immediately to the left of the subordinating complementiser *tʃe*. However, there are also attested examples in which the in-situ wh constituent scope out of the embedded clause and takes matrix scope, with the resulting interpretation being that of an indirect question. When this happens, the relative complementiser occurs in the matrix clause over which the wh in-situ takes scope.

(21) ere əm=mɛɛʃe tʃoː əmmɔ=s [tʃe ang nim]?
REL DEF.PL=crowd say PREP=3F.SG that I who
'Who do the crowds say that I am?' (Luke 9, 18)

The attentive reader might have noticed that both *e(re)* and *ete(re)* are attested in wh in-situ constructions, alongside perfective *ənt*. We tentatively attribute this somewhat free variation to the absence of an antecedent noun that could restrict the use of either complementiser on the basis of a definiteness distinction.

3.2 Wh ex-situ constructions

In terms of Cheng (1991)'s typology of wh-constructions, Coptic can be classified as an optional wh fronting language, in which wh ex-situ is available as a marked alternative to the canonical Wh in-situ pattern. Relative complementisers are systematically absent in wh ex-situ questions. In contrast to wh in-situ constructions, wh ex-situ displays an argument/adjunct asymmetry, as fronted wh arguments are always construed with a resumptive pronoun, while fronted wh adjuncts are not (Reintges, 2007).

(22) a. nim a=f ent=k e=pei=ma?
who PERF=3SG.M bring=2M.SG to=DEM.SG.M=place
'Who brought you here?' (Koptische Heiligen- und Martyrerlegenden, ed. Till, I 3, 7-8)

b. eβɔl ton a=tetən ei e=pei=ma?
PCL where PERF=2PL come to=DEM.SG.M=place
'From where did you come here?' (Coptic Martyrdoms, ed. Budge, 220, 8)

The scope of wh ex-situ is contingent on the syntactic position of the wh constituent. When the wh phrase appears to the left of the subordinating complementiser *tʃe*, it takes the embedded scope and the entire construction is interpreted as an indirect question. On the other hand, if the wh phrase appears in the matrix clause, the resulting interpretation is that of a direct question.

(23) a. ən=aʃ ən=he əntɔk kə=tʃɔː əmmɔ=s [tʃe tet(ən)=na
in=what of=manner YOU.SG.M 2SG.M=say PREP=3F.SG that 2PL=FUT
ər rəmhe]?
become free.man
'How do you (sg) say that you (pl) will become free?' (John 8, 33)

b. ti=tʃənu: əmmɔ=tən [tʃe hən u: ən=ʃatʃe a=tentən mu:te
1SG=ask PREP=2PL that with what of=word PERF=2PL say
erɔ=i]
about=1SG
'I ask you with which reason do you say about me ...' (Acts 10, 29)

4 Analysis

4.1 Relative constructions

As we have seen in section 2, the relative complementisers *ənt*, *e(re)*, and *ete(re)* mark the top of an unbounded dependency, with the bottom of that dependency realised as a resumptive pronoun. Following recent work on resumption within HPSG (Taghvaipour, 2005; Crysmann, 2012; Borsley, 2010; Alotaibi and Borsley, 2013), we assume that resumption involves ordinary SL(ASH) passing, rather than a separate non-local feature RESUMP, as postulated by Vaillette (2001).

For the purposes of this paper, we shall adopt the specific proposal of Borsley (2010) and Alotaibi and Borsley (2013) who suggest that resumptive dependencies require an amendment of standard lexical SLASH AMALGAMATION (Ginzburg and Sag, 2001) to permit optional termination of a SLASH dependency by way of index-sharing between the element in SLASH with that of a pronominal on ARG-ST, to license structures as illustrated in Figure 1.

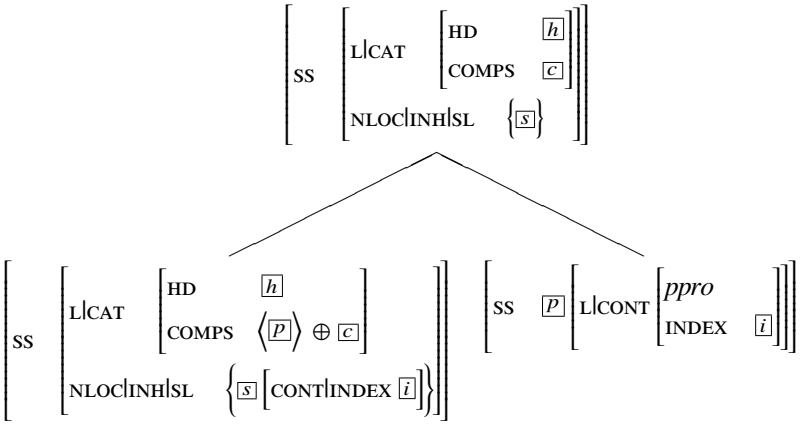


Figure 1: SLASH termination by resumption (cf. Borsley, 2010)

In order to license termination of a SLASH dependency by way of a (resumptive) pronoun, introduce this option by means of the two constraints shown in Figure 2.

Given this revised version, there are essentially three ways of satisfying this principle: either by co-indexation of the word's SLASH element with a (bound or free) pronoun (=resumption), standard termination of the SLASH dependency by a *gap-ss*, or else by a *canon-ss* that itself is slashed. As we have argued above, the grammar

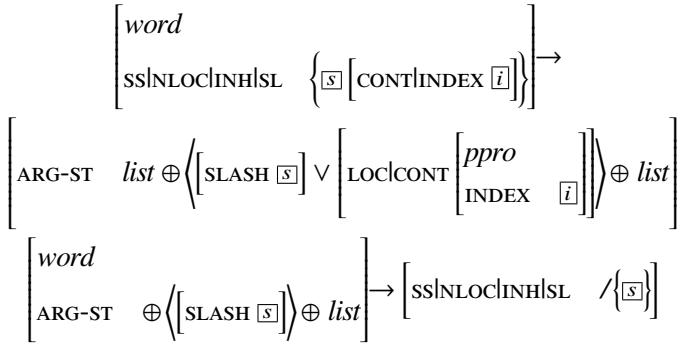


Figure 2: Revised SLASH AMALGAMATION for resumption (Alotaibi and Borsley, 2013)

of Coptic does not seem to allow for argument gaps. In order to capture this empirical generalisation about the language, all it takes is to ban termination of SLASH dependencies by way of *gap-ss*, as captured in Figure 3.

$$word \rightarrow \left[\textsc{arg-st} \quad list(canon\text{-}ss)\right]$$

Figure 3: Ban on argument gaps

Having discussed how SLASH dependencies can be terminated by means of pronominals, let us turn to the top of the unbounded dependency construction. Consider the schematic lexical entry for standard S-taking relative complementisers given in Fig. 4: apart from establishing modification of the antecedent noun via the MOD feature, these complementisers bind a SL dependency which they restrict to be an NP. In addition, they equate INDEX of the element in SLASH with that of the antecedent noun. Additional properties of individual relative complementisers, e.g. the constraint regarding definite antecedents for *ənt* and *ete(re)* can be stated by reference to the MOD value: e.g. the specific entries for *ənt/ete(re)* will require the antecedent noun to be definite, whereas those for *e(re)* will restrict it to be indefinite. Similarly, the restriction of *ənt* to past relatives can be captured by means of a constraint on its complement's INDEX.

Besides standard relatives featuring a non-local dependency with a resumptive at its foot, we observed exactly one construction with an apparent subject gap, involving the complementiser *et* . As detailed above, zero realisation was restricted to those constructions where an overt subject would otherwise surface at the left edge. Given the highly local nature of zero relativised subjects and the general absence of argument gaps in the language, we conclude that the properties of *et* are best captured in terms of local subcategorisation: as detailed in Fig. 5, *et* is subcategorised for a VP complement, i.e., a partially saturated verbal projection with an open subject valency, the INDEX of which is structure shared with the INDEX of the antecedent noun. Making the somewhat standard assumption that post-subject TAME markers are raising auxiliaries, whereas pre-subject TAME markers and negation combine with a fully saturated verbal projection, the distribution of *et* can be correlated with the different

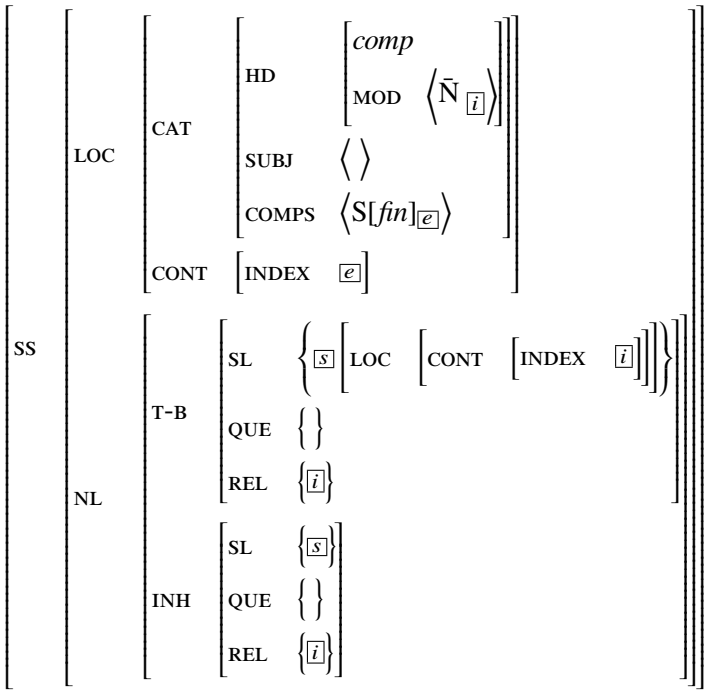


Figure 4: Relative complementisers (*ənt/e(re)/ete(re)*)

placement properties of pre-verbal TAME markers.

Having shown that apparent subject gaps in relatives are best understood as a local phenomenon, the generalisation that Coptic lacks argument gaps can be straightforwardly accounted by means of the constraint in 3 which restricts argument structure to consist entirely of canonical synsem objects.

4.2 Wh constructions

As we have seen in section 3, Coptic has (at least) two alternative constructions for wh questions: (i) wh ex-situ which is characterised by fronting of a wh phrase to the left of the clause or sentence, possibly involving pied-piping, and (ii) wh in-situ characterised by the absence of fronting and the presence of a "relative" complementiser.

4.2.1 Wh ex-situ

Similar to fronting in languages such as English (Pollard and Sag, 1994; Ginzburg and Sag, 2001), wh ex-situ phrases, as well as other fronted material, such as ex-situ focus are licensed in Coptic by a filler-head schema along the lines of Fig. 6: most crucially, this schema identifies the filler daughter's LOC information with a singleton element in the head-daughter's T(O)-B(IND)|SL(ASH).

Furthermore, the T-B|QUE value of the head daughter is constrained to be token-identical to the INH|QUE value of the filler daughter, thereby inhibiting percolation of

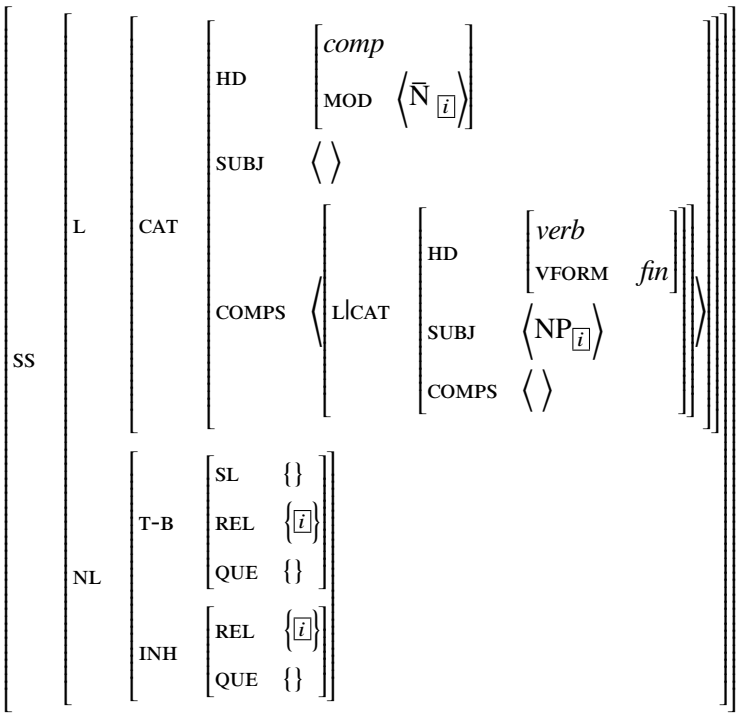


Figure 5: VP-taking complementiser *et*

a QUE dependency from an embedded ex-situ wh construction to the matrix clause. Interrogative illocutionary force can then be determined on the basis of a non-empty T-B|QUE value: if the filler contains a wh word, i.e. a word with a non-empty INH|QUE value (see Fig. 7), this value will be present on the INH|QUE of the filler daughter, by virtue of the Non-local Feature Principle (Pollard and Sag, 1994). Similarly, if no such wh word is present in the filler, the filler's INH|QUE value will be empty. Thus, as far as the filler and the determination of interrogative force are concerned, Coptic ex-situ wh constructions do not differ much from corresponding constructions in languages such as English.

Where Coptic differs from English, however, is at the bottom of the dependency: as witnessed by the data in sections 2 and 3.2, as well as the discussion in section 4.1 above, the language does not recognise any argument gaps. Besides argument fronting, which involves resumption at the bottom of the dependency, Coptic also features wh and focus fronting of modifiers, in which case there will be a gap at the extraction site.

Following the arguments presented by Levine (2003), we shall assume that adjunct extraction differs from argument extraction in being syntactic, rather than lexical in nature. Thus we shall assume that adjunct gaps are introduced by a syntactic unary rule, along the lines of Fig. 8. Given that filler-head structures equate the entire LOC value of the filler with the T-B|SL of the head daughter, a full local representation is sent down the tree, including both CAT and CONT information of the filler, thereby ac-

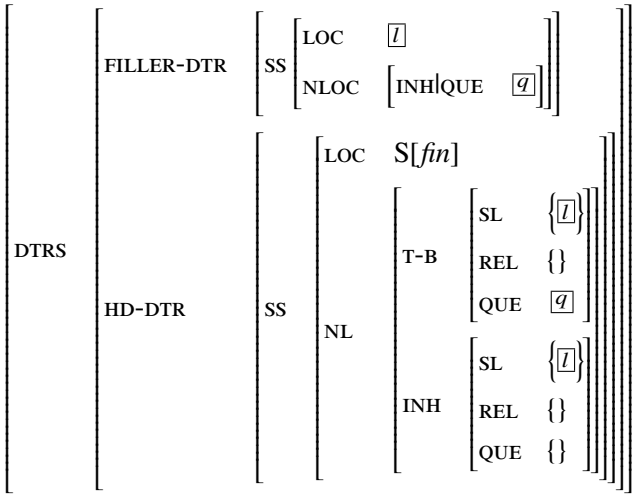


Figure 6: Filler-head schema

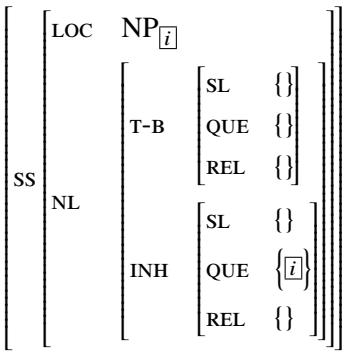


Figure 7: Lexical representation of wh words

counting for a matching effect between a modifying filler and its semantic integration at the gap site.

4.2.2 Wh in-situ

Having laid out our analyses of relative clauses and wh ex-situ constructions, we are now in a position to integrate the analysis of in-situ wh questions. To this end, we shall build on the proposal by Johnson and Lappin (1997) who exploit the non-local nature of QUE percolation for an account of in-situ wh question formation in Iraqi Arabic. Essentially, they generalise the QUE feature used for pied-piping in English wh fillers and apply it to non-local percolation from the sentence body.

The particularly compelling property of Coptic relative complementiser lies with the fact that the intricate morphosyntactic patterns regulating the choice of form generalise from relative constructions to their use in wh in-situ question formation, modulo, of course, the definiteness distinction, which we take to be neutralised by the

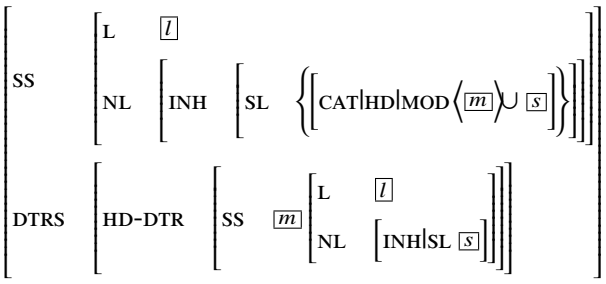


Figure 8: Adjunct extraction

absence of an antecedent noun in wh in-situ constructions. We shall therefore propose to model the polyfunctionality of these markers by means of the lexical rule depicted in Fig. 9. In essence this rule converts a relative complementiser terminating a SL dependency into a complementiser terminating a QUE dependency.

Since the output of the lexical rule, a wh complementiser, specifies a non-empty T-B|QUE value, interrogative illocutionary force will ensue, in much the same way as with overtly dislocated wh fillers. Most importantly, this illocutionary force is fixed at the level of the first complementiser or filler. Finally, conversion of a SL terminating complementiser into a QUE terminating one, already correctly rules out use of *et* in wh constructions: since the relative complementiser *et* represents a local relativisation strategy, devoid of (resumptive) SL dependency, it cannot be converted into a QUE dependency to serve in-situ wh constructions.

4.3 QUE islands

A final issue that has been brought to our attention by Bob Borsley (p.c.) concerns the locality conditions on QUE passing. Johnson and Lappin (1997) observe that island status varies according to the non-local feature involved (SL vs. QUE) and propose a parameterisation of the Non-local Feature Principle to capture these differences. Following previous observations made by Wahba (1991); Ouhalla (1994) and Simpson (1995), they note that in Iraqi Arabic, by contrast, finite clauses are QUE islands, but not SL islands, as witnessed by the data in (24) below.

(24) a. Mona shaafat meno?
Mona saw whom
'Who did Mona see.'

b. Mona raadat tijbir Su'ad tisa'ad meno?
Mona wanted to.force Su'ad to.help who
'Who did Mona want to force Su'ad to help?'

c. * Mona tsawwarat Ali ishtara sheno?
Mona thought Ali bought what

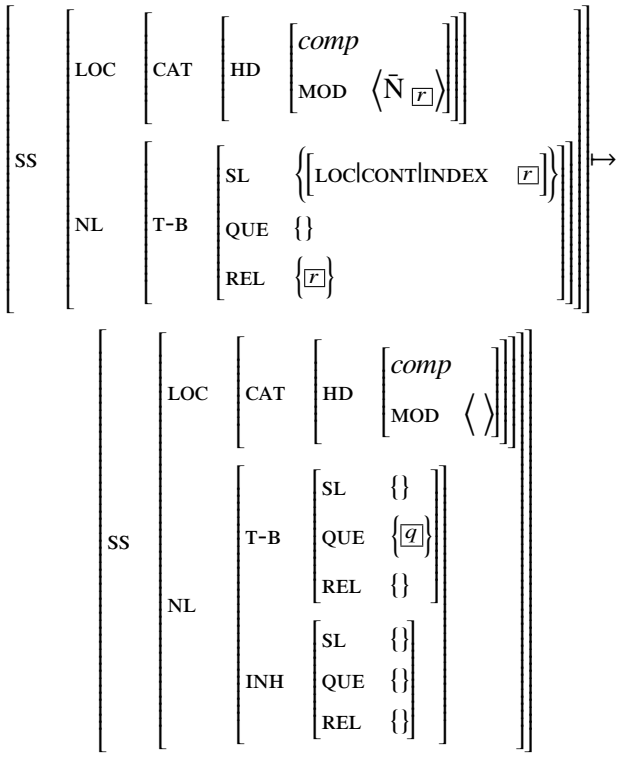


Figure 9: QUE-complementiser LR

d. Sheno tsawwarit Mona Ali ishtara?
what thought Mona Ali bought
'What did Mona think Ali bought?' (Johnson and Lappin, 1997)

While matrix scope of wh in-situ phrases contained in non-finite phrases is indeed possible, suggesting non-local QUE passing, matrix scope from wh in-situ phrases embedded in finite clauses is ruled out. Instead, fronting of the wh phrase remains the only option.

Pied-piping in English has been shown to involve unbounded QUE dependencies (Ross, 1967), Ginzburg and Sag (2001) argue that the bottom of a QUE dependency can only involve the least oblique argument. They suggest that all elements of ARG-ST except the first must be restricted to [QUE { }].

In Coptic, however, finite clauses do not constitute islands, neither for resumption, nor for adjunct extraction, nor for wh in-situ, as we have shown above, in contrast to Iraqi Arabic. Similarly, wh in-situ does not seem to observe any restriction with respect to the obliqueness of the argument involved, being attested for subjects and objects alike.

Although more extensive corpus research on the marked pied-piping alternative in Coptic wh-formation is necessary, the data we have so far investigated currently give us very little reason to believe that the level of unboundedness, in particular the

absence of clause-boundedness observed for wh in-situ, will carry over to pied-piping in wh ex-situ constructions: while QUE dependencies originating somewhere within the complement of a relative/wh complementiser may easily cross clause boundaries, we hypothesise that QUE dependencies within fillers observe somewhat stricter conditions, presumably disallowing QUE passing across finite clause boundaries.

Assuming for the sake of the argument that the locality conditions on pied-piping in Coptic do not differ in crucial respects from those observed for English, we would need to parameterise the locality condition according to the distinction between wh pied-piping and wh-in situ, i.e. we need to be able to impose QUE island constraints relative to the origin of the dependency (filler or complement). Since the value of QUE is a set (of indices), all it takes is to impose on the elements of the set a distinction between bounded and truly unbounded elements. Technically, this can be done either by cross-classifying the hierarchy of *index* types along the bounded/unbounded distinction, or else by means of an appropriate feature. Let us settle for the type-based approach: in order to establish a distinction with respect to the origin of the dependency, all it takes is to constrain the QUE set of filler daughters to be of type *bounded-index*. Constraints on wh pied-piping will then be formulated by restricting the QUE value of relevant members of ARG-ST to be a set of *unbounded-index*. As a result, boundedness will be selectively enforced for filler daughters, i.e. in pied-piping, but not for the complement daughter of a wh complementiser, ensuring true unboundedness of wh in-situ.

5 Conclusion

We have shown in this paper that Coptic observes a blanket ban on argument gaps observable in both relative clauses and wh ex-situ constructions, arguing that the apparent exception regarding zero subjects in *et*-relatives is of a highly local nature, to be modelled in terms of subcategorisation for a VP complement. Furthermore, we have discussed the local conditioning of complementiser allomorphy that generalises from relatives to in-situ wh constructions, militating for a treatment that systematically derives the latter use from the former. More specifically, we have suggested to model the wh usage of relative complementisers by means of a lexical rule that converts a (resumptive) SLASH dependency into a QUE dependency, enabling us to capture the assignment of interrogative force uniformly across in-situ and ex-situ constructions, while at the same time accounting for complementiser allomorphy.

The Coptic data discussed here are of utmost relevance to a general theory of resumption: since gap strategies are non-existent for arguments in both relatives and ex-situ wh questions and since wh in-situ is actually always available, these data should cast some serious doubts on theories such as Shlonsky's that picture resumption as a "last resort" rather than a grammatical option in its own right. Finally, the asymmetry between argument resumption and adjunct gaps lends further support for a distinction in terms of lexical and phrasal SLASH introduction.

References

Alotaibi, Mansour and Borsley, Robert D. 2013. Gaps and Resumptive Pronouns in Modern Standard Arabic. In Stefan Müller (ed.), *Proceedings of the 20th International Conference on Head-Driven Phrase Structure Grammar, Freie Universität Berlin*, pages 6–26.

Borsley, Robert D. 2010. An HPSG Approach to Welsh Unbounded Dependencies. In Stefan Müller (ed.), *Proceedings of the 17th International Conference on Head-Driven Phrase Structure Grammar, Université Paris Diderot, Paris 7, France*, pages 80–100, Stanford, CA: CSLI Publications.

Cheng, Lisa Lai Shen. 1991. *On the Typology of Wh-Questions*. Ph. D.thesis, Massachusetts Institute of Technology.

Crysmann, Berthold. 2012. Resumption and Island-hood in Hausa. In Philippe de Groote and Mark-Jan Nederhof (eds.), *Formal Grammar. 15th and 16th International Conference on Formal Grammar, FG 2010 Copenhagen, Denmark, August 2010, FG 2011 Lubljana, Slovenia, August 2011*, volume 7395 of *Lecture Notes in Computer Science*, pages 50–65, Springer.

Ginzburg, Jonathan and Sag, Ivan A. 2001. *Interrogative Investigations: the form, meaning, and use of English Interrogatives*. Stanford, California: CSLI Publications.

Johnson, David and Lappin, Shalom. 1997. A Critique of the Minimalist Program. *Linguistics and Philosophy* 20(4), 273–333.

Layton, Bentley. 2000. *Coptic Grammar with Chrestomathy and Glossary*. Porta Linguarum Orientalium N.S., No. 20, Wiesbaden: Harrassowitz.

Levine, Robert D. 2003. Adjunct valents: cumulative scoping adverbial constructions and impossible descriptions. In Jongbok Kim and Stephen Wechsler (eds.), *The Proceedings of the 9th International Conference on Head-Driven Phrase Structure Grammar*, pages 209–232, Stanford: CSLI Publications.

Ouhalla, Jamal. 1994. Remarks on the Binding Properties of Wh-in Situ, ms., Queen Mary and Westfield College, University of London.

Pollard, Carl and Sag, Ivan. 1994. *Head–Driven Phrase Structure Grammar*. Stanford: CSLI and University of Chicago Press.

Reintges, Chris, LeSourd, Philip and Chung, Sandra. 2006. Movement, wh-agreement, and 'apparent' wh-in-situ. In Lisa Lai-Shen Cheng and Norbert Corver (eds.), *WH-Movement Moving On*, pages 165–194, Cambridge, MA: MIT Press.

Reintges, Chris H. 2004. *Coptic Egyptian (Sahidic Dialect): A Learner's Grammar*. Africanist Studybooks, No. 15, Köln: Rüdiger Köppe Verlag.

Reintges, Chris H. 2007. Variable Pronunciation Sites and Types of Wh-in-situ. In Norbert Corver and Jairo Nunes (eds.), *The Copy Theory of Movement*, Linguistik Aktuell/Linguistics Today, No. 107, pages 249–287, Amsterdam: John Benjamins.

Reintges, Chris H. 2012. The Morpho-Syntax of Alternating Complementizers in Coptic Relative Clause Constructions. *Folia Orientalia* 49, 425–448.

Ross, John R. 1967. *Constraints on Variables in Syntax*. Ph. D.thesis, MIT.

Shlonsky, Ur. 1992. Resumptive pronouns as a last resort. *Linguistic Inquiry* 23, 443–468.

Simpson, A. 1995. *Wh-Movement, Licensing, and the Locality of Feature-Checking*. Ph. D.thesis, University of London.

Taghvaipour, Mehran. 2005. *Persian Relative Clauses in Head-driven Phrase Structure Grammar*. Ph. D.thesis, University of Essex.

Vaillette, Nathan. 2001. Hebrew relative clauses in HPSG. In Dan Flickinger and Andreas Kathol (eds.), *The Proceedings of the 7th International Conference on Head-Driven Phrase Structure Grammar*, pages 305–324, Stanford: CSLI Publications.

Wahba, W. A-F. B. 1991. LF Movement in Iraqi Arabic. In C.-T.J. Huang and R. May (eds.), *Logical Structure and Linguistic Structure*, pages 253–276, Dordrecht: Kluwer.

VP idioms in Norwegian: A subconstructional approach

Petter Haugereid

Bergen University College

Proceedings of the 21st International Conference on
Head-Driven Phrase Structure Grammar

University at Buffalo

Stefan Müller (Editor)

2014

Stanford, CA: CSLI Publications

pages 83–102

Haugereid, Petter. 2014. VP idioms in Norwegian: A subconstructional approach. In Stefan Müller (ed.), *Proceedings of the 21st International Conference on Head-Driven Phrase Structure Grammar, University at Buffalo*, 83–102. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2014.5. CC-BY 4.0

Abstract

This paper presents a brief overview of idiomatic expressions in the Norwegian LFG grammar NorGram and shows how the rich lexical information of the LFG grammar can be reused in an HPSG-like grammar with a radically different approach to alternating argument frames. Rather than accounting for idioms by means of special idiom lexical entries, which is the standard approach in LFG and HPSG, a constructional approach is taken where the verbs of the idioms are left underspecified with regard to whether they are idioms or not. A hierarchy of subconstruction types is assumed, which for each piece of evidence provided by the words and rules of the sentence, narrows down the possible frames of the verb to just one.

1 Introduction

The Norwegian LFG grammar NorGram (Dyvik, 2000; Butt et al., 2002) has 56 VP idioms in the lexicon, distributed over 20 templates. Abstracting away from whether the selected object of the idiom is definite or indefinite, and what kind of argument the selected preposition has (NP, subordinate clause or infinitival clause), we are left with four main kinds of idioms.[1]

The first two kinds of idioms are semantically intransitive, hence they only take one argument, namely the subject. In the first kind of intransitive idioms the main verb selects an object, as shown in (1), and the second kind the main verb selects a PP, as shown in (2).

(1) Han **gikk konkurs**.
he went bankrupt
He went bankrupt.

(2) De **løftet i flokk**.
They lifted in flock
They worked together.

The last two kinds of idioms are semantically transitive, hence they take two arguments. They differ in that in one kind the main verb selects an object and the preposition of a PP, see (3), while in the other the main verb selects a PP and takes an object as an argument, see (4).

(3) Han **la** ikke **skjul på** sin glede.
he laid not hiding on his joy
He did not hide his joy.

[†]I would like to thank two anonymous reviewers, the INESS group in Bergen, the audience at the HPSG 2014 conference in Buffalo, and the participants at the 2014 PARSEME meeting in Frankfurt, for very useful comments and suggestions.

[1]Three idioms (*ta på kreftene* ('tax one's strength'), *sende ord* ('send a message'), and *komme på kant med* ('fall out with')), do not fall into any of the four categories, and they are left out of the present discussion.

(4) Han **bragte** temaet **på bane**.
he brought topic.the on track
He brought up the topic.

A verb that is part of a VP idiom is assigned an idiom frame in the lexicon in addition to the other frames that it appears with. For example the verb *bringe* ('bring') is listed with the following frames:

(5)
```
@(V-SUBJ-POBJrefl-OBJ bringe med)
@(V-SUBJ-PRT-OBJ bringe inn)
@(V-SUBJ-OBJ-OBJ bringe)
@(V-SUBJ-OBJ-OBLBEN bringe)
@(V-SUBJ-OBJ bringe)
@(VPIDIOM-PSELOBJ-OBJ bringe på bane)
```

A lexical entry is allowed to have more than one argument frame by using disjunctions of frames. Disjunctions are expanded into full lexical entries during parsing. This means that a lexical entry with 6 disjunctive argument frames is computationally equivalent to six lexical entries.

In this paper I will present a new way of representing information about argument frames, including the different kinds of VP idioms presented in this section. The account shifts the burden from the lexicon to a carefully designed hierarchy of subconstruction types. The transfer is achieved by means of *phrasal subconstructions* (see Haugereid & Morey (2012); Haugereid (2012)), which are construction parts that, when put together in a way that conforms with a constraint on the verb, form full constructions. The analysis is implemented in an HPSG-like grammar of Norwegian within the LKB system (Copestake, 2001).

2 Treatment of idioms in Sag et al. (2003)

In (Sag et al., 2003, 347–355), idioms are assumed to have special lexical entries for the words that constitute them. The idiom *keep tabs on* is analyzed by means of a lexical entry for *keep* (see (6)) with three items on the SUBCAT list; (i) the NP subject, (ii) an idiomatic noun *tabs*, and (iii) a constituent marked by the preposition *on*.

(6)
$$\begin{bmatrix} \textit{ptv-lxm} \\ \text{STEM}\left\langle \text{keep} \right\rangle \\ \text{ARG-ST}\left\langle \text{NP}_i\ , \left[\text{FORM tabs}\right], \begin{bmatrix}\text{FORM on}\\ \text{INDEX } j\end{bmatrix} \right\rangle \\ \text{SEM}\begin{bmatrix} \text{INDEX } s \\ \text{RESTR}\left\langle \begin{bmatrix} \text{RELN} & \textbf{observe} \\ \text{SIT} & s \\ \text{OBSERVER} & i \\ \text{OBSERVED} & j \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

As (6) shows, the relation of the idiom *keep tabs on* (*observe*) has two arguments, OBSERVER and OBSERVED, and they are linked to the subject of *keep* and the constituent marked by the preposition *on*. Both the idiomatic noun *tabs* and the selected preposition *on* are semantically empty.

Given the degree of detail required in the lexicon, one is forced to assume separate lexical entries for idiomatic verbs. From a semantic point of view, this is motivated, considering how the meaning of idioms deviates from the compositional meaning. However, there is no morphological evidence indicating that idiomatic verbs should have separate lexical entries. They share the stem with their compositional versions and have the same inflections.

In section 3 I will present an account that allows us to have a single lexical entry for verbs that alternates between argument frames, including idiomatic frames.

3 Analysis

Instead of a lexical approach to subcategorization, a fully constructional approach is taken. In an analysis of a sentence, a *START* sign is assumed at the beginning of the sentence. Each word of the sentence is attached to this sign in an incremental, left-brancing fashion (see Haugereid & Morey (2012)). A simplified structure of a sentence with three words is given in Figure 1.

The relation of the sentence is not contributed by the main verb, but rather by the *START* sign. Instead of contributing a relation, the verb is assumed to have a feature FORM, and the value of this feature is unified with the PRED value of the relation.[2]

The VFORM value of the verb is by itself not enough to determine the predicate of the event expressed. In order for it to be fully specified, the predicate needs to be unified with other pieces of information stemming from the attachment of

[2]The assumption that the relation of the sentence is introduced by the *START* sign rather than the main verb is motivated by the fact that some languages have empty copula constructions, where there is no verb to contribute the relation.

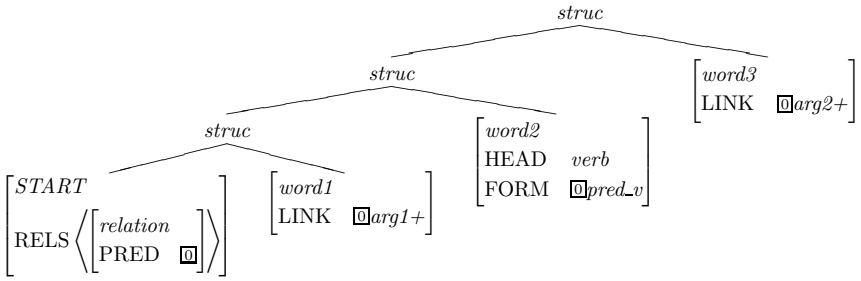


Figure 1: Leftbranching structure.

potential arguments. This is illustrated by means of the LINK features of *word1* and *word3* in Figure 1. Together, the LINK values here contribute the information that the predicate is a two-place predicate.

The motivation behind the demoted role of the verb is the fact that it is possible for verbs to alternate between different argument frames. Additionally, the approach lends itself nicely to the treatment of multiword expressions.

3.1 Lexical representation

In addition to the idiom frame shown in (4), the verb *bringe* also has a transitive and a ditransitive frame, as shown in (7).

(7) a. Han bragte maten.
he brought food.the
He brought the food.

b. Han bragte henne maten.
he brought her food.the
He brought her the food.

It also has frames that involve particles, prepositions and reflexives, as shown in (8).

(8) a. Han bragte med seg maten.
he brought with himself food.the
He brought the food.

b. Filmen bragte inn masse penger.
movie.the brought in lots-of money
The movie brought in lots of money.

c. Han bragte maten til henne.
he brought food.the to her
He brought the food to her.

Even though we have six argument frames for the verb *bringe*, I assume only one lexical entry, shown in (9). The lexical entry has information about the STEM of the lexeme, the HEAD value and the HEAD value of its (potential) arguments; C(ONSTRUCTION)-ARG1, C-ARG2, C-ARG3, and C-ARG4. These four argument features correspond to external subject, (deep) direct object, (deep) indirect object, and oblique object, respectively. Note that there is no linking of the C-ARGs to the semantics. Rather, the linking is done in what I refer to as *phrasal subconstructions*.

The lexical entry also has a feature FORM, and it is the value of this feature that determines which constructions the verb is compatible with.

(9)
$$\begin{bmatrix} \textit{bringe-v} & \\ \text{STEM} & \text{"bringe"} \\ \text{HEAD} & \textit{verb} \\ \text{VAL} & \begin{bmatrix} \text{C-ARG1} \left[\text{HEAD } \textit{noun}\right] \\ \text{C-ARG2} \left[\text{HEAD } \textit{noun}\right] \\ \text{C-ARG3} \left[\text{HEAD } \textit{noun}\right] \\ \text{C-ARG4} \left[\text{HEAD } \textit{compl-noun}\right] \end{bmatrix} \\ \text{FORM} & \boxed{1}\left[\text{PRED } \textit{bringe_v}\right] \end{bmatrix}$$

3.2 Phrasal subconstructions

One example of a phrasal subconstruction is the rule that links (external) subjects, *arg1-struct*, illustrated in (2). In this rule, the value of C-ARG1|LINK is switched from *arg1–* in the mother to *arg1+* in the first daughter. At the same time, the argument (the second daughter of the rule) is linked to the ARG1 of the KEYREL. The grammar also has subconstructions that in the same fashion link (deep) direct objects *arg2-struct*, (deep) indirect objects *arg3-struct*, and oblique objects *arg4-struct*.

The grammar has a rule *vbl-struct* which adds the verb. (See Figure 3). The verb is selected via the VBL feature of the first daughter, and the VBL value of the verb is transferred to the mother. In this way, the added verb is able to constrain the following verb, if there is one. The rule also unifies its KEYREL|PRED value with the FORM value of the verb. The verb does not contribute the full predicate, just a predicate type which, when unified with types contributed by the other subconstructions, yields the predicate of the clause.

The tree in Figure 4 shows how a transitive sentence is analysed. At the top node, the subconstruction constraints are negative. Three subconstructions apply, the *vbl-struc*, which adds the verb *bragte* ('brought'), the *arg1-struc*, which adds the subject *han* ('he'), and the *arg2-struc*, which adds the direct object *maten* ('the food'). Each subconstruction contributes a type; *vbl-struc* adds the FORM value of

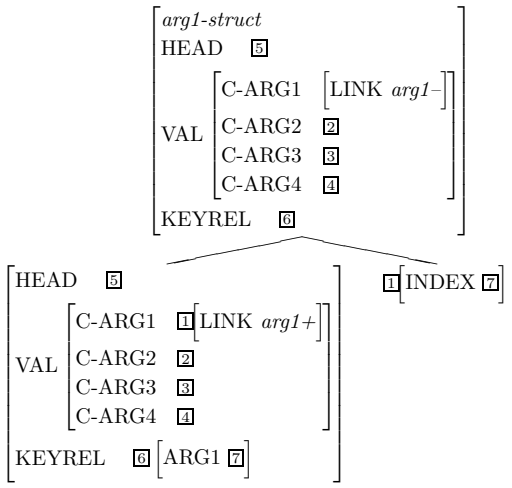


Figure 2: The *arg1-struct* rule for (external) subjects

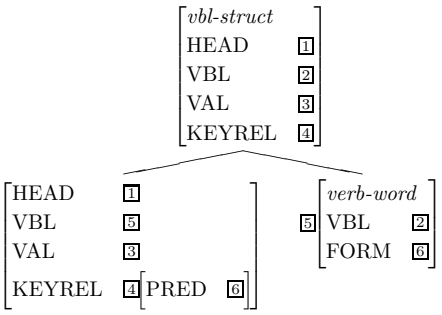


Figure 3: The *vbl-struct* rule for adding verbs

the verb, *bringe_v*, *arg1-struc* switches *arg1–* in the mother to *arg1+* in the first daughter, and *arg2-struc* switches *arg2–* to *arg2+*. As for the subconstructions that do not apply, their respective values stay negative. In this way, the *START* node reflects which subconstructions have applied, and which have not applied.

The result of unifying the subconstruction types *arg1+*, *arg2+*, *arg3–*, *arg4–*, *prt–*, and *bringe_v* in the *START* sign in Figure 4 is the predicate *bringe_12_rel*. This is shown in the type hierarchy in Figure 5, which will be discussed in Section 3.3.

3.3 Valence alternations

The valence alternations of the verb *bringe* (see (4), (7) and (8)) are accounted for by means of a hierarchy of *predicate* types. The type hierarchy in Figure 5 shows all the subconstruction types employed in order to account for the alternations of

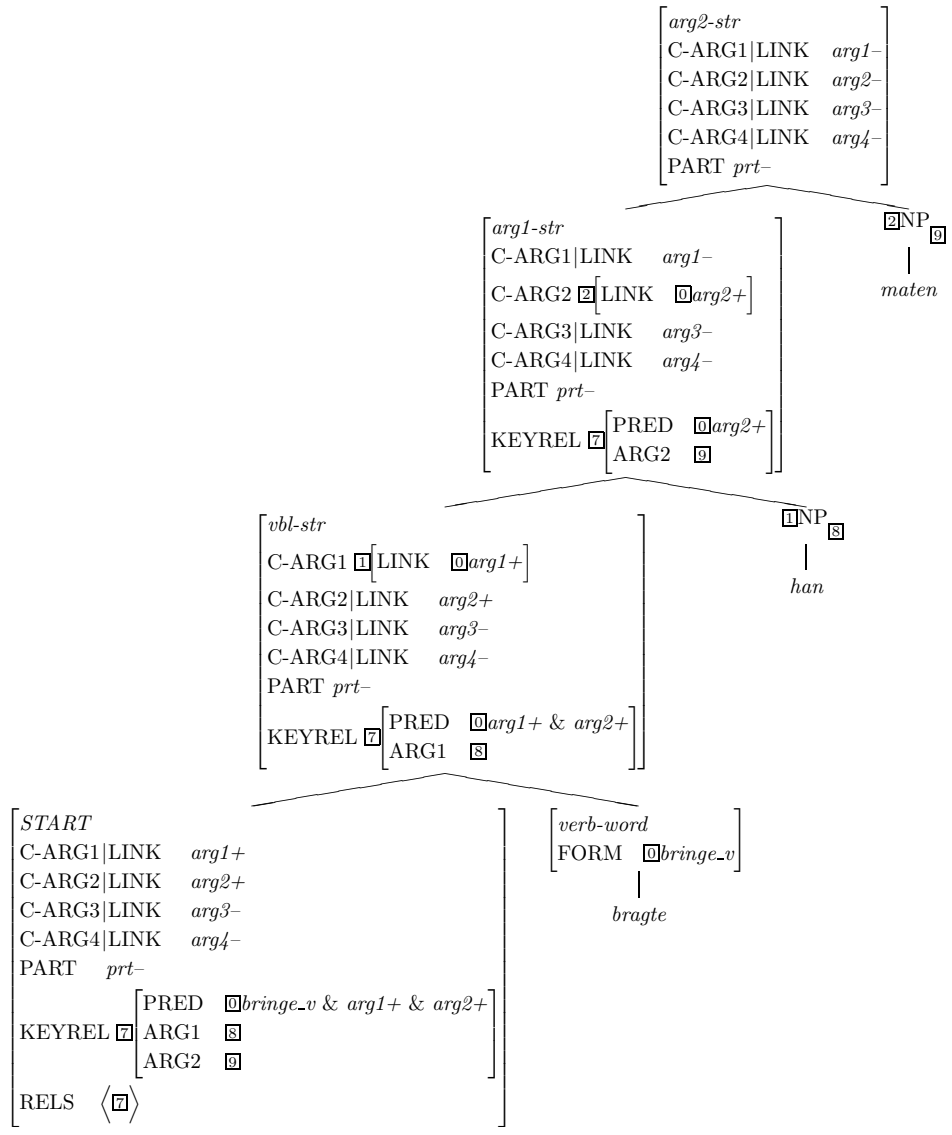


Figure 4: Analysis of the transitive sentence *Bragte han maten?* ('Did he bring the food?')

bringe.

The function of the subtypes of *link* in the hierarchy is to show whether a subconstruction has applied or not. For example, *arg1–* means that the arg1 subconstruction has not applied, while *arg1+* means that it has applied. The type *vrb+* has as immediate subtypes the FORM value of all verbs in the lexicon. (In Figure 5, only the FORM value of the verb *bringe* ('bring') is shown.) The subtypes of verb FORM values decide what frames a verb can appear in. As the hierarchy indicates,

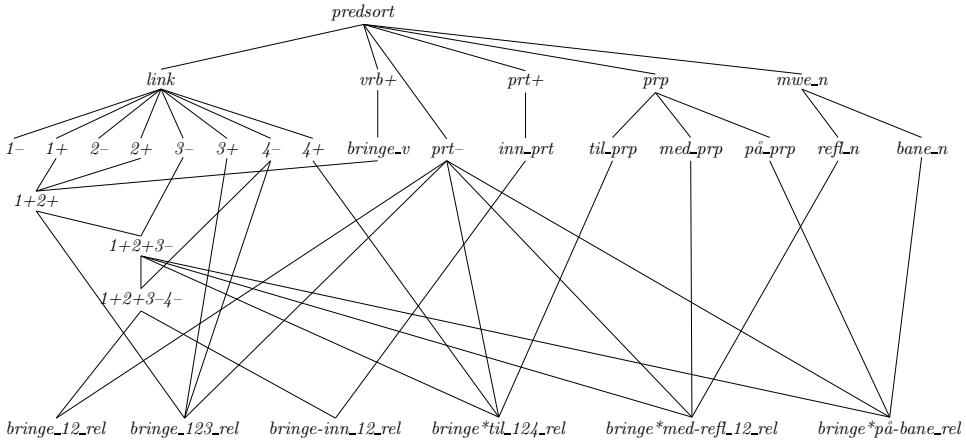


Figure 5: Type hierarchy accounting for the alternations of *bringe*.

bringe can appear in 6 frames, since it has 6 subtypes (ignoring the intermediate types). The type *prt+* has as immediate subtypes the FORM value of all the particles in the lexicon. (In Figure 5, only the FORM value of the particle *inn* ('in') is shown.) The type *prp* has as immediate subtypes the FORM value of all prepositions in the lexicon. (In Figure 5, only the FORM value of the prepositions *til* ('to'), *med* ('with'), and *på* ('on') are shown.) The type *mwe_n* has as immediate subtypes the FORM value of the reflexive (*refl_n*) and the FORM value of all the idiomatic nouns. (In Figure 5, only the FORM value of the reflexive and the idiomatic noun *bane* ('track') are shown.)

The subconstruction types are possible values of the features shown in Figure 6, and in order for a sentence to parse, these values need to unify. The features have different kinds of types as values before they are unified.

$$\begin{bmatrix} \text{VAL} & \begin{bmatrix} \text{C-ARG1}\,|\,\text{LINK} & \boxed{1} \\ \text{C-ARG2}\,|\,\text{LINK} & \boxed{1} \\ \text{C-ARG3}\,|\,\text{LINK} & \boxed{1} \\ \text{C-ARG4}\,|\,\text{LINK} & \boxed{1} \\ \text{PART} & \boxed{1} \end{bmatrix} \\ \text{KEYREL}\,|\,\text{PRED} & \boxed{1} \end{bmatrix}$$

Figure 6: Unification of subconstruction types.

The value of C-ARG1|LINK is binary; either *arg1–*, which means that no (external) subject has been realized, or *arg1+*, which means that it has been realized. In the case of *bringe*, all the frames require the *arg1+* type, which means that they are all agentive.

The feature C-ARG2|LINK can have three different kinds of values. It can be the type *arg2–*, which means that no (deep) direct object has been realized. It can have the value arg2+, which means that a (deep) direct object is realized, and that it has a semantic role (see (7a), repeated below as (10a)).[3] It is then not part of an MWE. Finally, it can have a subtype of *mwe_n* as value. In this case, the direct object is either a reflexive, as in (10b), or it constitutes a part of an idiom, as in (1), repeated below as (10c).

(10) a. Han bragte maten.
he brought food.the
He brought the food.

b. Han barberer seg.
he shaves himself
He shaves.

c. Han gikk konkurs.
he went bankrupt
He went bankrupt.

If the value is a subtype of *mwe_n*, the direct object is not assumed to have a semantic role, as regular direct objects. Instead, it is added by the *arg2-mwe-struct* rule, which, rather than linking the object to the ARG2 role of the KEYREL, unifies the FORM value of the object with the PRED value of the KEYREL. This is shown in Figure 7.

Similar to the feature ARG2|LINK, the feature ARG3|LINK can have a negative value *arg3–*, which means that no (deep) indirect object has been realized, and a positive value *arg3+*, which means that a (deep) indirect object has been realized (with its own semantic role) (see (7b), repeated below as (11a)). It can also have an indirect object that is a part of an MWE, exemplified with a reflexive in (11b). This object is not assumed to have a semantic role and is added by the rule *arg3-mwe-struct*, which is similar to the *arg2-mwe-struct* rule.

(11) a. Han bragte henne maten.
he brought her food.the
He brought her the food.

b. Han nærmer seg en løsning.
he nears himself a solution
He is closing in on a solution.

The feature ARG4|LINK can have four types of values. It can have a negative value *arg4–*, which means that no oblique argument is realized. It can have a

[3]Currently, no distinction is made between frames with NPs, CPs, or IPs as direct objects. It is possible to account for this distinction by letting *arg2+* have subtypes such as *arg2_np*, *arg2_cp* and *arg2_ip*, however this has not yet been implemented. Instead, the ARG2|HEAD value is constrained in the lexicon.

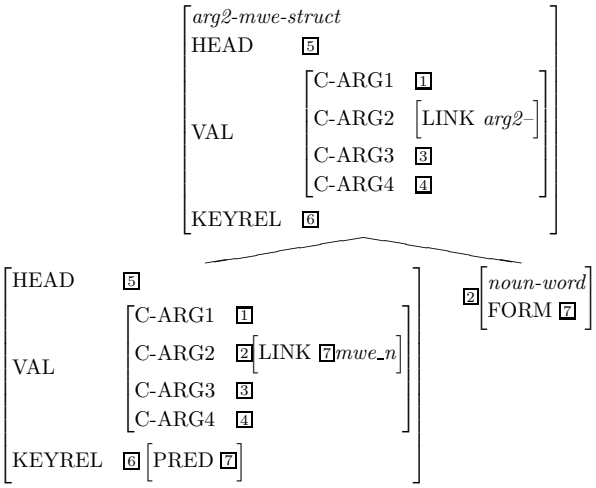


Figure 7: The *arg2-mwe-struct* rule for direct objects that are a part of an MWE.

positive value *arg4+*, which means that an oblique argument is realized, and that it has a semantic role (see (8c), repeated here as (12a)). In case the oblique argument does not have a semantic role, but constitutes a part of an MWE, the value is a subtype of *mwe_n*, for example *refl_n* in the case of reflexives (see (8a), repeated below as (12b)) or the FORM value of a oblique object that constitutes a part of an idiom (see (4), repeated below as (12c)). In the case of the idiom in (12c), the FORM value of the oblique object is *bane_n*. The FORM value of the prepositions that mark the oblique objects are the fourth type of value that the ARG4|LINK feature can have. They are unified with the *arg4+* type if the oblique object has a semantic role, or the relevant subtype of *mwe_n* if the oblique object is a part of an MWE. In (12a)–(12c), the FORM value of the prepositions marking the oblique object are *til_prp*, *med_prp*, and *på_prp*.

(12) a. Han bragte maten til henne.
he brought food.the to her
He brought the food to her.

b. Han bragte med seg maten.
he brought with himself food.the
He brought the food.

c. Han bragte temaet på bane.
he brought topic.the on track
He brought up the topic.

The subconstruction rule that adds the preposition that marks the oblique object is the *prepmark-struct* rule. (See Figure 8.) The rule unifies the FORM value of the preposition with the C-ARG4|LINK value, and switches the ARG4|MARKED

value from '–' in the (first) daughter to '+' in the mother. Once ARG4|MARKED is switched to positive, the oblique argument can be attached.

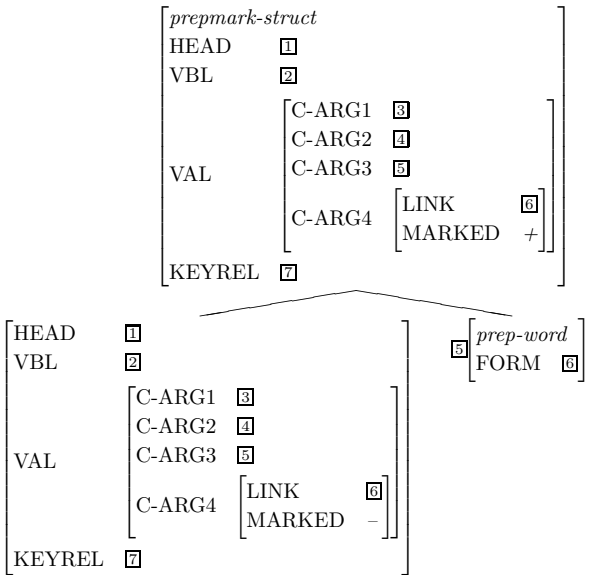


Figure 8: The *prepmark-struct* rule for prepositions marking oblique objects

If the oblique object is a part of an MWE (either a reflexive or an idiomatic noun), it is added by the *arg4-mwe-struct* rule shown in Figure 9.

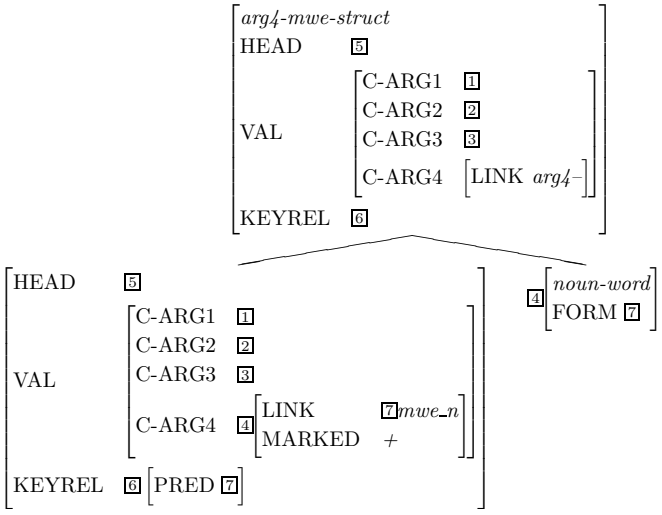


Figure 9: The *arg4-mwe-struct* rule for oblique objects that are a part of an MWE.

The feature PART has a negative value (*part–*) if the frame does not involve a

particle, and it has a subtype of *part+* if the frame involves a particle, as in (8c), repeated below as (13). The subtype will then be the FORM value of the selected particle (here *inn_prt*).

(13) Filmen bragte inn masse penger.
movie.the brought in a-lot-of money
The movie brought in a lot of money.

The feature KEYREL has as value the FORM value of the main verb, which is a subtype of *vrb+*.[4]

Figure 10 shows the subconstruction types that are unified in order to arrive at the frame type *bringe_12_rel* (*arg1+*, *arg2+*, *arg3–*, *arg4–*, *prt–*).

$$\begin{bmatrix} \text{VAL} & \begin{bmatrix} \text{C-ARG1}|\text{LINK} & \boxed{1}\textit{arg1+} \\ \text{C-ARG2}|\text{LINK} & \boxed{1}\textit{arg2+} \\ \text{C-ARG3}|\text{LINK} & \boxed{1}\textit{arg3–} \\ \text{C-ARG4}|\text{LINK} & \boxed{1}\textit{arg4–} \\ \text{PART} & \boxed{1}\textit{prt–} \end{bmatrix} \\ \text{KEYREL}|\text{PRED} & \boxed{1}\textit{bringe_v} \end{bmatrix}$$

Figure 10: Unification of subconstruction types that result in the type *bringe_12_rel*.

Similarly, Figure 11 shows the subconstruction types that are unified in order to arrive at the frame type *bringe_123_rel* (*arg1+*, *arg2+*, *arg3+*, *arg4–*, *prt–*, and *bringe_v*).

$$\begin{bmatrix} \text{VAL} & \begin{bmatrix} \text{C-ARG1}|\text{LINK} & \boxed{1}\textit{arg1+} \\ \text{C-ARG2}|\text{LINK} & \boxed{1}\textit{arg2+} \\ \text{C-ARG3}|\text{LINK} & \boxed{1}\textit{arg3+} \\ \text{C-ARG4}|\text{LINK} & \boxed{1}\textit{arg4–} \\ \text{PART} & \boxed{1}\textit{prt–} \end{bmatrix} \\ \text{KEYREL}|\text{PRED} & \boxed{1}\textit{bringe_v} \end{bmatrix}$$

Figure 11: Unification of subconstruction types resulting in the type *bringe_123_rel*.

The unifications resulting in the other frame types of *bringe* are given in Figures 12–14.

[4] In a language with empty copula constructions, one can also introduce a type *vrb–* for clauses without verbs. Norwegian, however, does not have this construction.

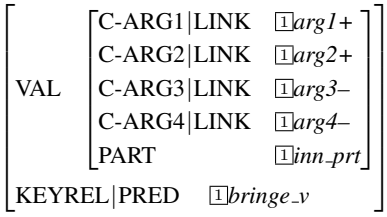


Figure 12: Unification of subconstruction types resulting in the type *bringe-inn_12_rel*.

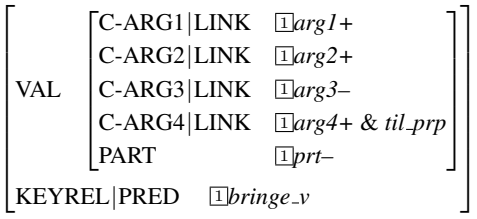


Figure 13: Unification of subconstruction types resulting in the type *bringe*til_124_rel*.

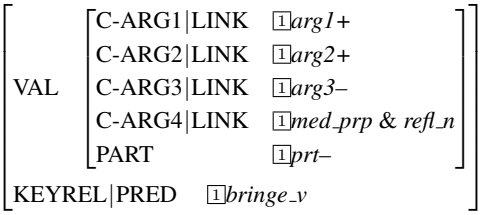


Figure 14: Unification of subconstruction types resulting in the type *bringe*med-seg_12_rel*.

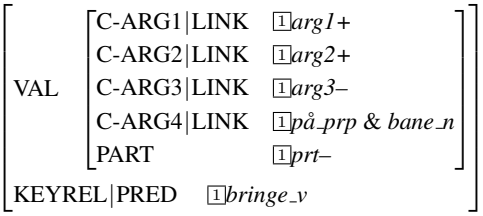


Figure 15: Unification of subconstruction types resulting in the type *bringe*på-bane_12_rel*.

	vbl-struct	*arg1-struct*	*arg2-struct*	*arg2-mwe-struct*	*prepmark-struct*	*arg4-struct*	*arg4-mwe-struct*
Intrans. with idiomatic noun	X	X		X			
Intrans. with idiomatic PP	X	X			X		X
Trans. with idiomatic noun	X	X		X	X	X	
Trans. with idiomatic PP	X	X	X		X		X

Table 1: Subconstructions involved in the different VP idiom types.

3.4 Analysis of VP idioms

The analysis of VP idioms includes the subconstruction rule for prepositions marking oblique objects *prepmark-struct* (see Figure 8) and two subconstructions rules for MWE nouns; *arg2-mwe-struct* and *arg4-mwe-struct* (see Figures 7 and 9).

An analysis of a sentence with a VP idiom (*Bragte han temaet på bane* 'Did he bring up the topic') is illustrated in Figure 16. Five subconstruction apply. The first subconstruction *vbl-struct* adds the verb *bragte* and unifies the FORM value of the verb with the KEYREL|PRED value. The second subconstruction *arg1-struc* adds the subject *han*, and links its index to KEYREL|ARG1. The third subconstruction *arg2-struc* adds the direct object *temaet*, and links its index to KEYREL|ARG2. The fourth subconstruction *prepmark-struct* adds the preposition marking the oblique object *på* and unifies the FORM value of the preposition with the KEYREL|PRED value (and the C-ARG4|LINK value of the first daughter). The fifth subconstruction adds the idiomatic noun *bane* and unifies its FORM value with the KEYREL|PRED value (and the C-ARG4|LINK value of the first daughter).

In the top node *arg4-mwe-struct*, all LINK values are constrained to be negative, and at the bottom of the tree, in the *START* node, marks from all the subconstructions that have applied can be found, and they are unified. When the subconstruction types in the *START* sign are unified, we get the type *bringe*på-bane_12_rel*.

The four kinds of idiomatic expression types introduced in Section 1 are accounted for by the combinations of subconstructions shown in Table 1

4 Implementation

The most common templates in the NorGram LFG grammar are given in Table 2. The table shows how the information encoded in these frames can be broken down

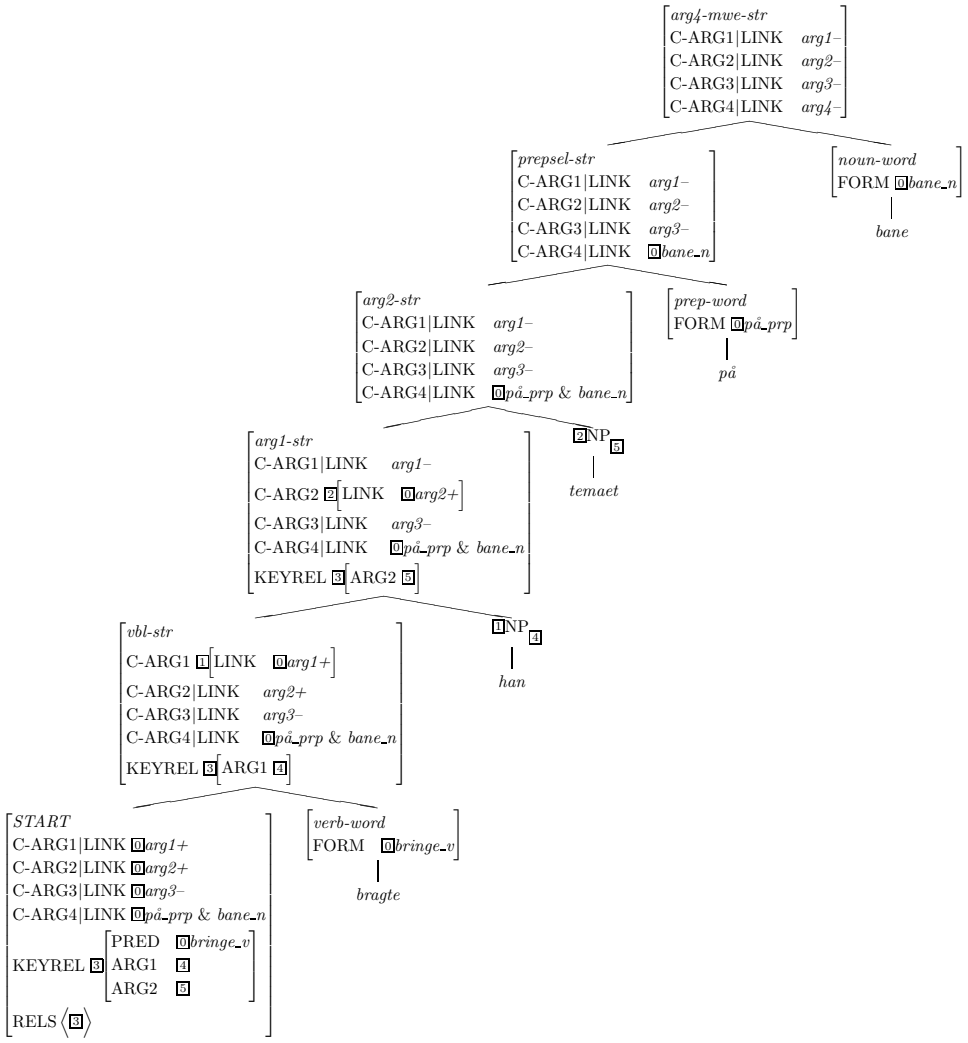


Figure 16: Linking information in the idiom *Brakte han temaet på bane?* (Did he bring up the topic?)

into subconstruction types.[5] For example, the most common template V-SUBJ-OBJ is associated with the subconstruction types *arg1+*, *arg2+*, *arg3–*, *arg4–*, and *prt–*, which are the types that come from a standard transitive sentence.

In addition to the types shown in Table 2, the FORM value of the verb, and the FORM values of prepositions and particles (if applicable), which are part of the LFG frames, are added to the subconstruction types. Given a table that maps templates to subconstruction types as shown in Table 2, the LFG frames in (5) can be translated into the following types:[6]

[5]The inquit template (Vinq-SUBJ-COMP) for sentences like *"Jeg kommer", sa han.* ("I'm coming", he said.) is not included, as inquit frames currently are not handled by the grammar.

[6]The preposition of the template 'V-SUBJ-OBJ-OBLBEN' is specified in the template to be *til*,

LFG template		Subconstruction types					
Template	Freq.	C-ARG1	C-ARG2	C-ARG3	C-ARG4		PART
V-SUBJ-OBJ	735	arg1+	arg2+	arg3–	arg4–		prt–
V-SUBJ-PRT-OBJ	572	arg1+	arg2–	arg3–	arg4–		prt+
V-SUBJ	473	arg1+	arg2–	arg3–	arg4–		prt–
V-SUBJ-POBJ	388	arg1+	arg2–	arg3–	arg4+	prp+	prt–
V-SUBJ-PRT	280	arg1+	arg2–	arg3–	arg4–		prt+
V-SUBJ-OBJrefl	201	arg1+	refl_n	arg3–	arg4–		prt–
V-SUBJ-OBJ-POBJ	111	arg1+	arg2+	arg3–	arg4+	prp+	prt+
V-SUBJ-OBJrefl-POBJ	108	arg1+	refl_n	arg3–	arg4+	prp+	prt–
V-SUBJ-COMP	101	arg1+	arg2+	arg3–	arg4–		prt–
V-SUBJ-OBJrefl-PRT	94	arg1+	refl_n	arg3–	arg4–		prt+
V-SUBJunacc	84	arg1–	arg2+	arg3–	arg4–		prt–
V-SUBJ-PRT-POBJ	66	arg1+	arg2–	arg3–	arg4+	prp+	prt+
V-SUBJ-POBJrefl-OBJ	66	arg1+	arg2+	arg3–	refl_n	prp+	prt–
V-SUBJexpl	52	arg1–	arg2–	arg3–	arg4–		prt–

Table 2: The most common frames in NorGram, and their conversion into sets of subconstruction types

```
bringe*med-refl_12_rel := bringe_v & arg1+ & arg2+ & arg3- &
                          med_prp & refl_n & prt-.
bringe-inn_12_rel := bringe_v & arg1+ & arg2+ & arg3- & arg4- &
                     prt+.
bringe_123_rel := bringe_v & arg1+ & arg2+ & arg3+ & arg4- & prt-.
bringe_124_rel := bringe_v & arg1+ & arg2+ & arg3- & arg4+ &
                  til_prp & prt-.
bringe_12_rel := bringe_v & arg1+ & arg2+ arg3- & arg4- & prt-.
bringe*på-bane_12_rel := bringe_v & arg1+ & arg2+ arg3- & bane_n &
                         på_prp & prt-.
```

The hierarchy of relation types and subconstruction types above is the same as the hierarchy in Figure 5. This shows how a type hierarchy of subconstruction types can be generated, given a conversion table. The program that generates the a type hierarchy from an LFG lexicon and a conversion table can be conceived of as a compiler.

The NorGram lexicon has 15,776 verb frames. I have tested the procedure on a slightly smaller version of the lexicon, the open source NKL lexicon with 13,069 verb frames, and loaded it into the LKB system. Loading the grammar now obviously takes more time, but the efficiency of the parser does not seem to affected by the large number of subconstruction types (almost 20,000 in all).

The MRSs resulting from parsing the four idiomatic examples in (1), (2), (3), and (4) are given in Figure 17–20.

and is not specified in the frame.

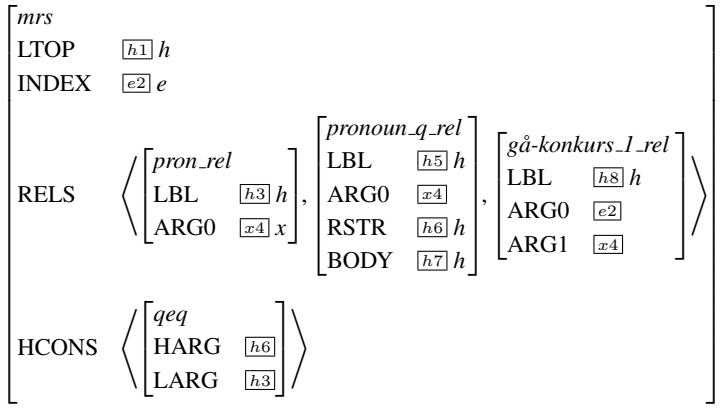


Figure 17: MRS of the sentence *Han gikk konkurs.* ('He went bankrupt')

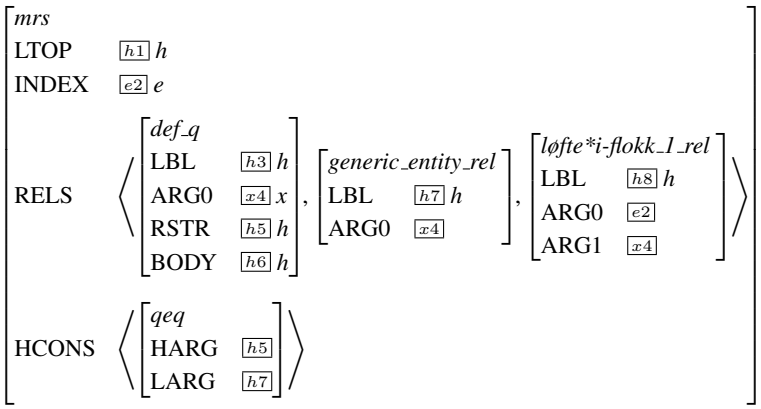


Figure 18: MRS of the sentence *De løftet i flokk.* ('They worked together.')

5 Discussion and future work

The analysis presented in this paper is not restricted to idioms, but includes several kinds of MWEs, like particle verbs, verbs with selected prepositions, reflexive verbs, and combinations of these. It can also be expanded to nouns and adjectives with selected complements.

I have dealt only with idiomatic nouns that are indefinite, although idiomatic expressions also may consist of definite idiomatic nouns, like *øynene* ('eyes.the') in *ta øynene fra* ('look away from') or even idiomatic nouns modified by an adjective, like *et godt øye* ('a good eye') in *ha et godt øye til* ('have a preference for'). Examples like these suggest that the predicates in the hierarchy of link types not only need to reflect the base form of idiomatic nouns, but also other features like definiteness and adjuncts.

The flexibility of the approach comes from the fact that it is a subconstructional approach. While lexicalist approaches need to be very specific about the argument structure of a verb, and need to use disjunctions of frames in lexical entries (LFG)

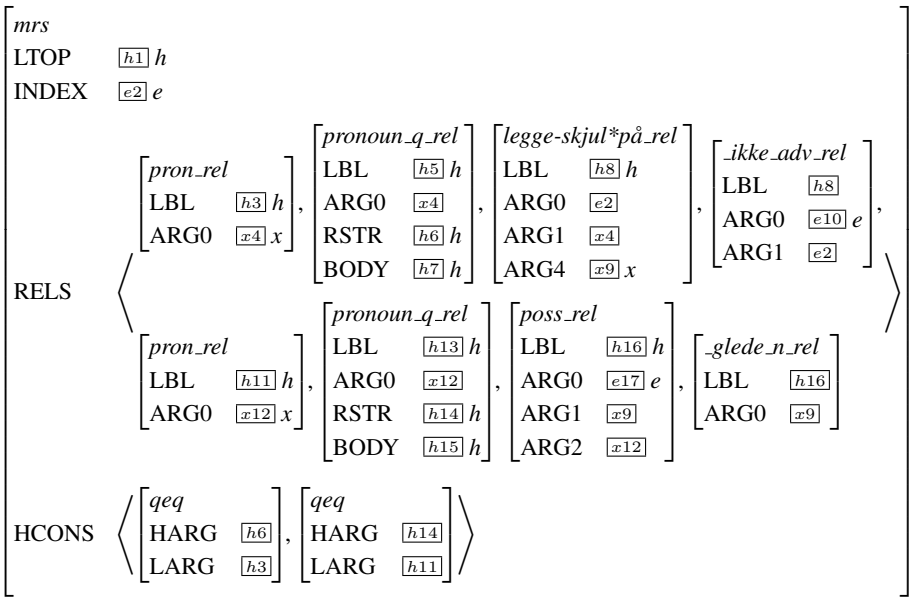


Figure 19: MRS of the sentence *Han la ikke skjul på sin glede.* ('He did not hide his joy.')

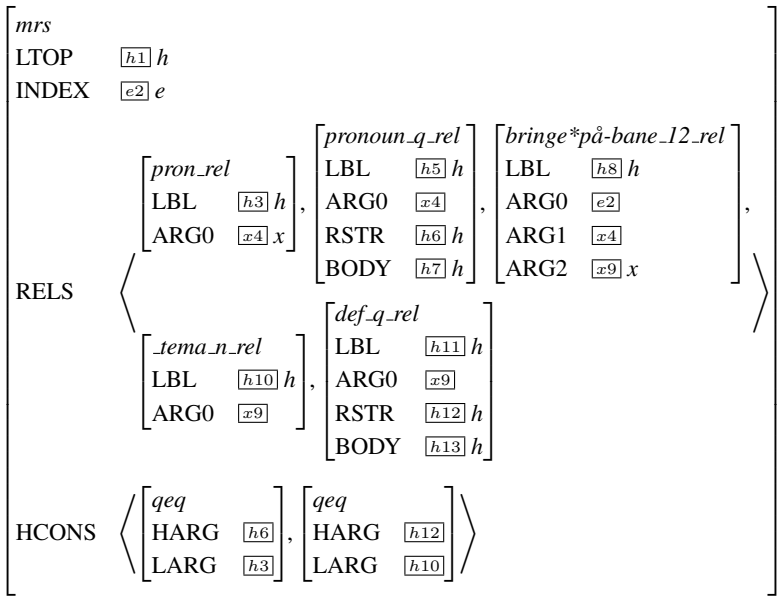


Figure 20: MRS of the sentence *Han bragte temaet på bane.* ('He brought up the topic.')

or multiple lexical entries/lexical rules (HPSG) in order to account for valence alternations, the subconstructional approach allows for precise underspecification using the hierarchy of subconstruction types. Only one lexical entry per verb is

needed. And while constructional approaches are forced to assume relatively flat syntactic structures in order to have access to the arguments of a construction, and hence risk ending up with an unmanageable amount of phrase structure rules, the subconstructional approach allows for binary structures and the number of phrase structure rules is kept relatively small (about 80). The combination of lexical underspecification and binary structures is achieved by means of the type hierarchy of subconstruction types which includes types for all verbs, prepositions, particles and idiomatic nouns in the lexicon and types for the frames they occur in. The hierarchy is designed in such a way that a verb is only allowed to combine with selected combinations of constituents. The hierarchy is huge, but finite. And it is interesting in that it reflects what kinds of subconstructions are needed in order to express all grammaticalized concepts in a grammar.

References

Butt, Miriam, Helge Dyvik, Tracy Holloway King, Hiroshi Masuichi & Christian Rohrer. 2002. The Parallel Grammar project. In John Carroll, Nelleke Oostdijk & Richard Sutcliffe (eds.), *Proceedings of the workshop on grammar engineering and evaluation at the 19th international conference on computational linguistics (coling), taipei, taiwan*, 1–7. Stroudsburg, PA, USA: Association for Computational Linguistics.

Copestake, Ann. 2001. *Implementing typed feature structure grammars* CSLI Lecture Notes. Stanford: Center for the Study of Language and Information. http://cslipublications.stanford.edu/site/1575862603.html.

Dyvik, Helge. 2000. Nødvendige noder i norsk: Grunntrekk i en leksikalsk-funksjonell beskrivelse av norsk syntaks. In Øivin Andersen, Kjersti Fløttum & Torodd Kinn (eds.), *Menneske, språk og felleskap*, Novus forlag.

Haugereid, Petter. 2012. A grammar design accommodating packed argument frame information on verbs. *International Journal of Asian Language Processing* 22(3). 87–106.

Haugereid, Petter & Mathieu Morey. 2012. A left-branching grammar design for incremental parsing. In Stefan Müller (ed.), *Proceedings of the 19th international conference on head-driven phrase structure grammar, chungnam national university daejeon*, 181–194. http://cslipublications.stanford.edu/HPSG/2012/haugereid-morey.pdf.

Sag, Ivan A., Thomas Wasow & Emily M. Bender. 2003. *Syntactic theory: A formal introduction*. Stanford: CSLI Publications 2nd edn. http://cslipublications.stanford.edu/site/1575864002.html.

Negation in Nanti: Syntactic evidence for head and dependent negators

David Inman
University of Washington

Ruth Morrison
University of Washington

Proceedings of the 21st International Conference on
Head-Driven Phrase Structure Grammar

University at Buffalo

Stefan Müller (Editor)

2014

Stanford, CA: CSLI Publications

pages 103–113

Inman, David & Ruth Morrison. 2014. Negation in Nanti: Syntactic evidence for head and dependent negators. In Stefan Müller (ed.), *Proceedings of the 21st International Conference on Head-Driven Phrase Structure Grammar, University at Buffalo*, 103–113. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2014.6.
CC-BY 4.0

Abstract

In this paper we argue that, despite a lack of morphological markers on its negators, Nanti shows syntactic evidence for two negation strategies in the main clause: head negation and modifier negation. The head negator motivates the construction of a hierarchy of forms, and the interaction of the main clause negators motivates an additional HEAD feature. We then extend the analysis to a previously unconsidered negator in the language. Finally, our analysis is implemented and tested in a grammar based on the LinGO Grammar Matrix.

1 Introduction

This paper examines and presents an analysis of negators and their interaction in Nanti [ISO 693-3 code: cox], a Kampan-branch Arawakan language spoken in Peru. We argue that Nanti uses two different negation strategies among three negators. Our work also serves as an illustration of identifying head versus dependent negators without the help of morphological distinctions. We first begin with background and motivation for the analysis, followed by the data and analysis itself, and finally typological implications.

This also is an example of hypothesis testing through grammar engineering (Bender, 2008). We implemented a small, functional grammar fragment for Nanti as part of a course taught by Emily Bender at the University of Washington (Bender, 2007), and the grammar includes the negation analysis presented here. The grammar was developed from the LinGO Grammar Matrix customization system (Bender et al., 2002, 2010), followed by manual modification and expansion by the authors. We developed a testsuite of 206 sentences, 118 grammatical and 88 ungrammatical. Of these, 33 deal with negation, and so are of immediate relevance to the current paper. Both the testsuite and the grammar are publicly available for download at `https://github.com/faiuwle/Nanti`.

2 Motivation

Lev Michael describes the negation system in Nanti as consisting of a pair of internal negators and an external negator (Michael, 2008, 2014b). The internal negators *tera* and *hara* are described as having basic semantic negation properties, as well as forcing an alternation of verbal mood. The external negator *matsi* is semantically a metalinguistic negator (Michael, 2014b). All negators take scope over clauses, and it is possible for an internal and external negator to cooccur, but only with a particular ordering. While the distribution is well-described, the reasons for it remain elusive, at least within Michael's grammar. We propose that an HPSG analysis of the negators as auxiliaries and modifiers captures these distribution patterns.

3 Data

Nanti employs the following negation strategies: the metalinguistic negator *matsi*, the descriptive negators *tera* and *hara* (with reduced clitic forms *te* and

ha), existential negation, and exhaustive negation (Michael, 2014b). We focus on the descriptive and metalinguistic negators. The data presented in this section is all taken from Michael 2014b.

Both metalinguistic *matsi* and the descriptive negators *tera* and *hara* appear to the left of the verb and its arguments (excepting any in the initial topic position), as seen in examples (1) and (2):

(1) *Matsi nopakeri maika peremisa.*

matsi no=p-ak-e=ri maika peremisa
NEG.META 1S=give-PERF-REAL.I=3MO now permission

'It is not the case that I gave him permission at that time.' [cox] (Michael, 2014b, p.194)

(2) *Tera imporohe.*

tera i=N-poroh-e
NEG.REAL 3MS=IRREAL-clear.land-IRREAL.I

'He is not clearing land.' [cox] (Michael, 2014b, p.188)

The difference between *tera* and *hara* lies in their interaction with the Nanti mood system, a binary realis/irrealis system (called reality status in the literature), which is used, among other things, to distinguish future events from non-future ones (Michael, 2014a). *Tera* is used only with notionally realis (non-future) clauses, while *hara* is used only with notionally irrealis ones (Michael, 2008). Nevertheless, *tera* requires its clauses to be irrealis-marked, and *hara* requires its to be realis-marked. Michael 2014b refers to these latter as "doubly irrealis" clauses, with the negation adding an extra element of irrealis.

(3) a. *Opoki.*

o=pok-∅-i
3NMS=come-IMPF-REAL.I

'She is coming.' [cox] (Michael, 2014b, p.190)

b. *Tera ompoke.*

tera o=N-pok-e
NEG.REAL 3NMS=IRREAL-come-IRREAL.I

'She did not come.' [cox] (Michael, 2014b, p.191)

(4) a. *Ompoke.*

o=N-pok-∅-e
3NMS=IRREAL-come-IMPF-IRREAL.I

'She will come.' [cox] (Michael, 2014b, p.191)

b. *Hara opoki.*

hara o=pok-i
NEG.IRREAL 3NMS=come-REAL.I

'She will not come.' [cox] (Michael, 2014b, p.191)

Tera and *hara* also prohibit aspect marking in Nanti, which is otherwise obligatory on verbs, either as the perfective *-ak* suffix as in (5b) or as the null imperfective suffix as in (5a).

(5) a. *Inihi.*
i=nih-∅-i
3MS=speak-IMPF-REAL.I
'He was speaking.' [cox] (Michael, 2014b, p.193)

b. *Inihake.*[1]
i=nih-ak-i
3MS=speak-PERF-REAL.I
'He spoke.' [cox] (Michael, 2014b, p.193)

(6) a. *Hara inihi.*
hara i=nih-i
NEG.IRREAL 3MS=speak-REAL.I
'He will not speak.' [cox] (Michael, 2014b, p.193)

b. **Hara inihake.*
hara i=nih-ak-i
NEG.IRREAL 3MS=speak-PERF-REAL.I
*'He will not speak.' [cox] (Michael, 2014b, p.193)

It is also possible for *tera* or *hara* to follow *matsi* to create a doubly negated clause as in (7), but it is not possible for any negator to follow *tera* or *hara*.

(7) *Matsi te pishinetemparo oka.*
matsi te pi=N-shine-eNpa=ro o-oka
NEG.META NEG.REAL 2S=IRREAL-like-IRREAL.A=3NMO 3NM-this
'It is not the case that you don't like this.' [cox] (Michael, 2014b, p.195)

Another negator, which we were not aware of during our initial analysis, is the "exhaustive" negator *mameri*, used to indicate that the state of the clause is not realized even to the smallest degree, as in (8). Like *tera*, *mameri* applies only to notionally realis clauses with irrealis marking, and does not allow the verb to take aspect marking (Michael, 2014b). Because of these commonalities, our analysis for *tera* also works for *mameri*.

(8) *Mameri inehakotero saburi, kotsiro.*
mameri i=N-nehako-e=ro saburi
NEG.EX 3MS=IRREAL-be.familiar.with-IRREAL.I=3NMO machete
kotsiro
knife
'He had no familiarity with machetes or knives at all.' [cox] (Michael, 2014b, p.198)

[1] As noted in Michael 2014b, the realis and irrealis suffixes for *-i* verbs are neutralized after perfective *-ak*.

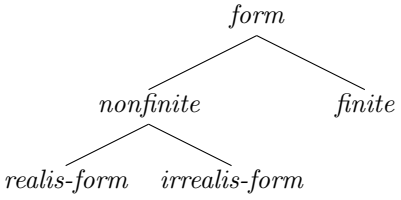
In summary, *tera* takes notionally realis clauses while *hara* takes notionally irrealis ones, and *matsi* can take either. The descriptive negators *tera* and *hara* require clauses to take on the opposite reality status marking to their notional/semantic value, and while *matsi* can be followed by a descriptive negator, the descriptive negators cannot be followed by other negators. Additionally, there is an exhaustive negator *mameri*, which behaves like *tera*.

4 Analysis

The challenge for the analysis is to capture the phenomena described above within the HPSG framework, with well-motivated rule sets and feature geometries that generate and parse grammatical examples, while failing to generate ungrammatical examples. To this end, we use the Grammar Matrix system (Bender et al., 2002, 2010) as an implementation tool to fully test our analyses.[2] The two chief phenomena to address are: the TAM restrictions for dependent clauses of the descriptive negators *tera* and *hara* (examples 3–6); and the ordering restriction that *matsi* must precede *tera* or *hara* (example (7)).

The clauses following the descriptive negators *tera* and *hara* exhibit two restrictions: they cannot take aspect marking (6b), and they exhibit mood-marking inversion (that is, their syntactic mood-marking is the opposite of their semantic mood). However, *tera* and *hara* themselves do not take morphological marking. In terms of head features (as described comprehensively in Zwicky 1985), the descriptive negators cannot be easily defined as morphosyntactic loci (since they have no morphology themselves), but are clearly governing the following sentential complement by restricting the morphological shape of its head verb. Zwicky cites this governing pattern as sufficient to analyze English auxiliaries as heads, and we agree and apply the same reasoning to Nanti. We analyze the descriptive negators *tera* and *hara* as heads, and further analyze their aspectless, mood-inverted complements as nonfinite sentences (necessarily headed by nonfinite verbs) governed by the negator. We accordingly define nonfinite FORMs for these verbs, in the following type hierarchy:

(9)



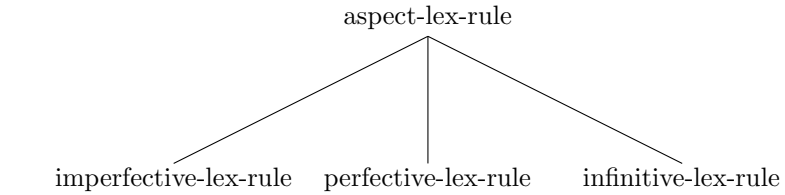
Form serves as a general type for the FORM value on HEAD[3], with daughters *finite* and *nonfinite*, and *nonfinite* leaves *realis-form* and *irrealis-form* representing nonfinite verb forms with the respective realis or irrealis morphology. This use of the FORM feature on HEAD is necessary to ensure that the analysis of the negation interacts correctly with other analyses, namely that all verbs in Nanti not negated with descriptive negators require both aspect and mood

[2]The feature geometry shown in this paper is that of the implemented grammar, which is based on the Grammar Matrix.

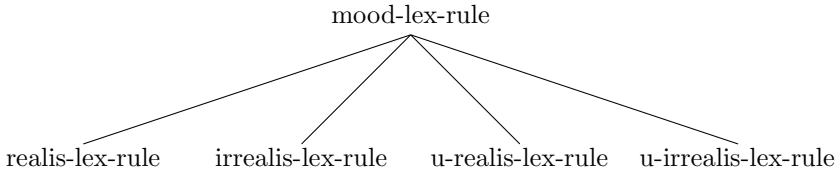
[3]While it may seem strange to have a feature which relates to mood in HEAD rather than somewhere in CONT, this is necessary to effect the "mood reversal" triggered by the descriptive negators without changing the notional (semantic) information in the MRS itself.

marking. We require [FORM *finite*] for the root node, and all [FORM *finite*] verbs are required to be marked for both aspect and mood. This is handled by a system of flags (see Goodman 2013) which require all verb lexemes to go through lexical rules corresponding with morphological positions for aspect and mood marking. There are two sets of these rules: the ones which require the verb to be [FORM *finite*] and the ones which require [FORM *nonfinite*] or one of its daughter types. The former apply both aspect and mood markings as usual (with the realis morphology matched to [E.MOOD *realis*] verbs and irrealis morphology attached to [E.MOOD *irrealis*] verbs). The latter do not apply aspect, and assign a FORM value rather than an E.MOOD value, leaving the actual application of mood for later in the unification process. Resultingly, morphological forms can be assigned with their opposite E.MOOD values, so long as the verbs are [FORM *nonfinite*]. These morphologically "mismatched" verbs can then be selected appropriately as complements of the negators *tera* and *hara*, thanks to the FORM values. The type hierarchies for the lexical rules are presented in examples (10) and (11), with infinitive-lex-rule, u-realis-lex-rule, and u-irrealis-lex-rule assigning non-finite FORM values and not assigning E.ASPECT or E.MOOD values (hence "u" for "unspecified"), letting *tera* and *hara* constrain the mood of their complements appropriately.[4]

(10)



(11)



With the non-finite complements and their associated sections of the type hierarchy worked out, we return to *tera* and *hara*. Taking the above conclusion that these are heads selecting for sentential complements, we turn to the question of what kind of head they are. One candidate is that these are auxiliary verbs, since they are taking a verbal (sentential) complement. The descriptive negators lack lexical meaning, only contributing grammatical function (negation) to the clause, and *tera* and *hara* also undergo phonological reduction to their respective clitic forms *te* and *ha*, both qualities shared with many auxiliary constructions (Anderson, 2006). We take these factors as sufficient to posit an analysis of *tera* and *hara* as defective auxiliary verbs. These auxiliaries specify the FORM values *realis-form* or *irrealis-form* on their complements as described in the above paragraph. We introduce the boolean value HEAD.AUX to distinguish these negators from other verbs, and also to prohibit auxiliaries from taking verbal morphology.[5] Finally we introduce a boolean feature HEAD.NEGATED

[4]The actual names of the lexical rules are slightly different here than in the TDL code, so as to perserve formatting.

[5]To prohibit or permit particular morphology on particular lexemes, we make use of a

to keep track of negation in the syntax. The NEGATED feature allows the syntax to distinguish between the grammatical negation *matsi te* and ungrammatical *te matsi*, by specifying that the descriptive negators must take a non-negated complement. These common properties are shared in a common supertype for *tera* and *hara*, which we have termed *neg-aux-lex* (12).

(12)
$$\begin{bmatrix} \textit{neg-aux-lex} & & \\ \text{SYNSEM...CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{verb} & \\ \text{AUX} & + \\ \text{NEGATED} & + \end{bmatrix} \\ \text{VAL.COMPS} & \left\langle \begin{bmatrix} \text{LOCAL.CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{verb} & \\ \text{FORM} & \textit{nonfinite} \\ \text{NEGATED} & - \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SPR} & \text{null} \\ \text{COMPS} & \text{null} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix} \\ \text{INFLECTED} & \textit{infl-satisfied} \end{bmatrix}$$

The individual negators *tera* and *hara* inherit from the constraints specified in (12), with the following additions defining their particular types of mood-marking inversion:

(13) a.
$$\begin{bmatrix} \textit{neg-notionally-realis-aux-lex} & \\ \text{STEM} & \left\langle \text{“tera”} \right\rangle \\ \text{SYNSEM...COMPS} & \left\langle \begin{bmatrix} \text{LOCAL} & \begin{bmatrix} \text{CONT...E.MOOD} & \textit{realis} \\ \text{CAT.HEAD.FORM} & \textit{irrealis-form} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

b.
$$\begin{bmatrix} \textit{neg-notionally-irrealis-aux-lex} & \\ \text{STEM} & \left\langle \text{“hara”} \right\rangle \\ \text{SYNSEM...COMPS} & \left\langle \begin{bmatrix} \text{LOCAL} & \begin{bmatrix} \text{CONT...E.MOOD} & \textit{irrealis} \\ \text{CAT.HEAD.FORM} & \textit{realis-form} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

Thus, when all the lexical rules and constraints from negators are applied, we have the following analysis for the verbs from example (3), excluding irrelevant parts of the feature structures, such as the PNG values associated with the subject:

---

structure called INFLECTED introduced via the Grammar Matrix. INFLECTED contains a number of flags indicating which lexical rules a lexeme has gone through, and a special type [INFLECTED *infl-satisfied*] indicates a fully-inflected form. We give the descriptive negators an INFLECTED value of *infl-satisfied* (to permit them to enter into the syntax as fully-formed words), and specify [AUX −] on all lexical rules in the verbal morphology (to prevent these negators from acquiring verbal morphology). For a fuller discussion of the role of INFLECTED features, see Goodman 2013.

(14) a.
$$\begin{bmatrix} \textit{word} & \\ \textsc{stem} & \left\langle \text{“opoki”} \right\rangle \\ \textsc{cat} & \begin{bmatrix} \textsc{head} & \begin{bmatrix} \textit{verb} & \\ \textsc{form} & \textit{finite} \end{bmatrix} \\ \textsc{val} & \begin{bmatrix} \textsc{spr} & \textsc{null} \\ \textsc{comps} & \textsc{null} \end{bmatrix} \end{bmatrix} \\ \textsc{cont...e} & \begin{bmatrix} \textsc{aspect} & \textit{imperfective} \\ \textsc{mood} & \textit{realis} \end{bmatrix} \end{bmatrix}$$

b.
$$\begin{bmatrix} \textit{word} & \\ \textsc{stem} & \left\langle \text{“ompoke”} \right\rangle \\ \textsc{cat} & \begin{bmatrix} \textsc{head} & \begin{bmatrix} \textit{verb} & \\ \textsc{form} & \textit{irrealis-form} \end{bmatrix} \\ \textsc{val} & \begin{bmatrix} \textsc{spr} & \textsc{null} \\ \textsc{comps} & \textsc{null} \end{bmatrix} \end{bmatrix} \\ \textsc{cont...e} & \begin{bmatrix} \textsc{aspect} & \textsc{null} \\ \textsc{mood} & \textit{realis} \end{bmatrix} \end{bmatrix}$$

In particular, the mood value in (14b) is the result of unification with constraints imposed from *tera*.

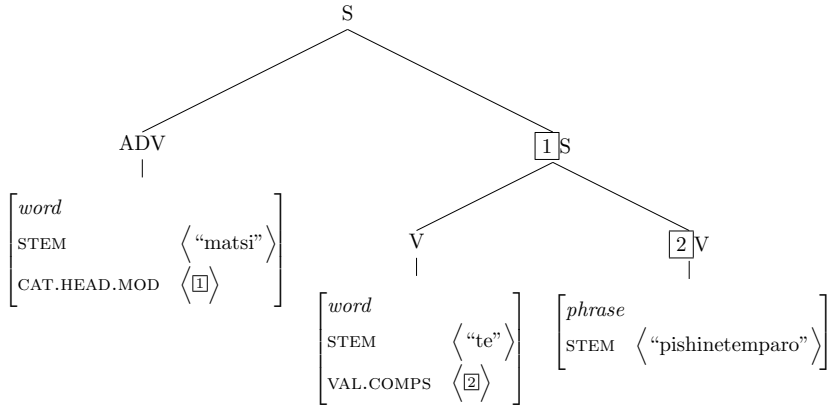
As mentioned earlier, exhaustive negator *mameri* functions in exactly the same way as the descriptive negators: *mameri* is captured with identical structure to *tera*, but with a different PRED value representing exhaustive negation.

The analysis for metalinguistic negator *matsi* is somewhat simpler. Michael 2008 describes *matsi* as being 'external' to the clause structure, and it does not interact with reality status or aspect in any way. The motivating factors for headedness in the descriptive negators are absent for *matsi*: *matsi* does not govern any following clause, determine concord features, support morphosyntactic marking, subcategorize what can occur with it, nor semantically head its phrase. In fact it fails all of Zwicky 1985's tests for headedness. In the absence of evidence for headedness, we simply analyze *matsi* as a scopal adverb that takes a saturated sentence in its MOD list and only appears to the left of the head (*i.e.*, is [POSTHEAD −]). This attaches via the usual head-modifier phrase. To allow *matsi* to interact with the descriptive negators, we further constrain its MOD value to be [NEGATED +]. Thus, descriptive negators cannot take as a complement any clause which *matsi* has modified, and *te matsi* fails to unify.

(15)
$$\begin{bmatrix} \textit{neg-adv-lex} \\ \textsc{stem} & \langle \text{“matsi”} \rangle \\ \textsc{synsem...cat} & \begin{bmatrix} \textsc{head} & \begin{bmatrix} \textit{adv} \\ \textsc{mod} & \left\langle \begin{bmatrix} \textsc{local.cat} & \begin{bmatrix} \textsc{head} & \begin{bmatrix} \textit{verb} \\ \textsc{negated} & + \end{bmatrix} \\ \textsc{val} & \begin{bmatrix} \textsc{spr} & \textsc{null} \\ \textsc{comps} & \textsc{null} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix} \\ \textsc{val} & \begin{bmatrix} \textsc{spr} & \textsc{null} \\ \textsc{comps} & \textsc{null} \end{bmatrix} \\ \textsc{posthead} & - \end{bmatrix} \end{bmatrix}$$

With these analyses, we may then produce a rough tree for the sentence *matsi te pishenetemparo*[6] in example (16).

(16)



These combined analyses allow the descriptive negators, as auxiliaries, to take a complement verb that is of one form (realis or irrealis) while semantically/notionally indicating the opposite. We are also able to successfully reject examples with both a descriptive negator and an aspect, such as **Hara inihake* (6b). The HEAD.NEGATED feature and associated constraints prevent sequences of “*te matsi*” from parsing while allowing “*matsi te*”, regardless of intervening adjuncts between the negators. Thus we have a well-motivated analysis of two negators *tera*/*hara* as syntactic auxiliaries, and one negator *matsi* as a pre-head modifier, even though neither type takes inflectional morphology, and thus there are no morphological cues to differentiate them in this case.

5 Typology

Crowgey 2013 presents a survey of predicted negation strategies from an HPSG perspective. These predictions include the simple set: negation by inflection; by auxiliary verb; by selected complement; and by free modifier; as well as the more complex bipartite set, where negation is expressed via two obligatory morphemes with one selecting for the other (not observed in Nanti). Since we

[6]This is similar but not identical to the sentence in example (7), as the final *oka* in that sentence exhibits topicalization, a feature of Nanti which is beyond the scope of this paper.

built our grammar on the Grammar Matrix, it is built on a foundation that assumes Crowgey's theoretical framework. Nevertheless, our analysis shows Nanti to be compatible with these theoretical predictions, with the descriptive negators *tera* and *hara* and the exhaustive *mameri* mapping onto an auxiliary strategy (aux-neg in Crowgey's typology), and metalinguistic *matsi* using a free modifier strategy (mod-neg). The Nanti data does bring up an interesting complication in showing a language with multiple syntactic strategies for main-clause negation. So far as we are aware, there is no reason to presuppose that languages will exclusively use one strategy for negating main clauses rather than several. Indeed, the (at this point dated) use of sentence-final pause and emphatic *not* in English can be analyzed as a mod-neg (17), in addition to the normal use of *not* as a comp-neg (18):

(17) We had fun... not.

(18) We did not have fun.

While there have been formal analyses showing distinct negation strategies for different kinds of clauses (such as Borsley and Jones 2005, which illustrated different negation strategies for finite main clauses versus non-finite subordinate clauses and imperative clauses in informal Welsh), we do not know of any that indicate multiple negation strategies simply for main clauses. However, if syntactic strategies for negation can vary with slang in the above way in English, there is no reason to assume a language cannot have more than one stable main-clause negation strategy. We have outlined the means by which we have determined the type of negation strategies presented in the data, and we believe that the interaction between different negators is a potentially fruitful area of future typological research. Within the set of field descriptions for minority languages, one item to look for is the putative difference between external-clause and internal-clause negators: this descriptive account may indicate different syntactic strategies.

Morphology can sometimes help determine if a negator is a head or a dependent, but in this case we used the interaction between the negators and the verbs they combine with to determine that *tera* and *hara* are heads. This approach can be applied cross-linguistically to distinguish negators in languages that employ multiple strategies.

6 Acknowledgements

For the description of the language we are deeply indebted to Lev Michael (2008) and his Nanti grammar.

References

Anderson, Gregory D.S. 2006. *Auxiliary Verb Constructions*, 4–8. Oxford University Press.

Bender, Emily M. 2007. Combining research and pedagogy in the development of a crosslinguistic grammar resource. In *Proceedings of the GEAF07 Workshop*, ed. Tracy Holloway King and Emily M. Bender, 26–45. Stanford, CA: CSLI.

URL http://csli-publications.stanford.edu/GEAF/2007/geaf07-toc.html.

Bender, Emily M. 2008. Grammar Engineering for Linguistic Hypothesis Testing. In *Proceedings of the Texas Linguistics Society X Conference: Computational Linguistics for Less-Studied Languages*, 16–36. Stanford, CA.

Bender, Emily M., Scott Drellishak, Antske Fokkens, Michael Wayne Goodman, Daniel P. Mills, Laurie Poulson, and Safiyyah Saleem. 2010. Grammar prototyping and testing with the lingo grammar matrix customization system. In *Proceedings of the ACL 2010 System Demonstrations*, 1–6. Uppsala, Sweden: Association for Computational Linguistics. URL http://www.aclweb.org/anthology/P10-4001.

Bender, Emily M., Dan Flickinger, and Stephan Oepen. 2002. The Grammar Matrix: An Open-Source Starter-Kit for the Rapid Development of Cross-Linguistically Consistent Broad-Coverage Precision Grammars. In *Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics*, 8–14. Taipei, Taiwan.

Borsley, Robert D., and Bob Morris Jones. 2005. *Welsh Negation and Grammatical Theory*. University of Wales Press.

Crowgey, Joshua. 2013. The Syntactic Exponence of Sentential Negation: a model for the LinGO Grammar Matrix. Master's thesis, University of Washington.

Goodman, Michael. 2013. Generation of Machine-Readable Morphological Rules from Human-Readable Input. *Working Papers in Linguistics* 30.

Michael, Lev. 2008. Nanti evidential practice: Language, knowledge, and social action in an Amazonian society. Doctoral Dissertation, University of Texas.

Michael, Lev. 2014a. The Nanti reality status system: Implications for the typological validity of the realis/irrealis contrast. *Linguistic Typology* 18.

Michael, Lev. 2014b. Negation in Arawak Languages.

Zwicky, Arnold M. 1985. Heads. *Journal of Linguistics* 21:1–29.

Deconstructing SYNtax

Jean-Pierre Koenig
University at Buffalo

Karin Michelson
University at Buffalo

Proceedings of the 21st International Conference on
Head-Driven Phrase Structure Grammar

University at Buffalo

Stefan Müller (Editor)

2014

Stanford, CA: CSLI Publications

pages 114–134

Koenig, Jean-Pierre & Karin Michelson. 2014. Deconstructing SYNtax. In Stefan Müller (ed.), *Proceedings of the 21st International Conference on Head-Driven Phrase Structure Grammar, University at Buffalo*, 114–134. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2014.7. CC-BY 4.0

Abstract

There are at least two distinct ways of conceiving of syntax: the set of rules that enable speakers and listeners to combine the meaning of expressions (compositional syntax), or the set of formal constraints on the combinations of expressions (formal syntax). The question that occupies us in this paper is whether all languages include a significant formal syntax component or whether there are languages in which most syntactic rules are exclusively compositional. Our claims are (1) that Oneida (Northern Iroquoian) has almost no formal syntax component and is very close to a language that includes only a compositional syntax component and (2) that the little formal syntax Oneida has does not require making reference to syntactic features.[1]

There are at least two distinct ways of conceiving of syntax:

Definition 1. *Syntax is the set of rules that enable speakers and listeners to combine the meaning of two or more expressions (words or phrases) (hereafter, compositional syntax)*

Definition 2. *Syntax is the set of formal constraints on the combinations of two or more expressions (words or phrases) (hereafter, formal syntax)*

Syntactic rules in most languages partake of both conceptions of syntax: They are statements about how speakers can combine the meaning of expressions while at the same time restricting the form of the expressions they license the combination of. But only the first, i.e. compositional syntax, is a conceptual necessity. Whatever syntax does, it must at a minimum ensure that when two expressions of the right semantic kind combine, they combine semantically in the right way. This is what the syntax of natural and logical languages share. It seems impossible to imagine a natural language whose syntax would not provide recipes for combining the meanings of expressions that are part of well-formed constituents. But, because most syntactic rules in most languages also include a formal component, we tend to think of syntax in the second sense (what we call *formal syntax*) as syntax proper. The question that occupies us in this paper is whether *all* languages include a significant formal syntax component or whether there are languages in which most syntactic rules are exclusively compositional and do not restrict the form of the expressions that combine. Our claim is that Oneida (Northern Iroquoian) is such a language. Most of its syntactic rules or constructions are strictly compositional, and very few include a formal component and that formal component is very restricted. More precisely, we make the following two claims.

Claim 1. *Compared to most languages, Oneida has almost no formal syntax component and is very close to a language that includes only a compositional syntax component.*

[1]The examples come from a compilation of recorded texts or stories (Michelson, Kennedy and Doxtator, to appear), and we are grateful to those who have contributed recordings; we would like especially to acknowledge Norma Kennedy and the late Mercy Doxtator for their collaboration. References to works on Iroquoian languages may be found in an annotated bibliography (Michelson 2011). Grammars of several other Iroquoian languages are presently underway (e.g. Chafe In press).

Claim 2. *The little formal syntax Oneida has does not require making reference to syntactic features: Words and phrases in Oneida do not carry any syntactic feature.*

Our paper is organized as follows. In Section 1, we discuss traditional kinds of evidence that can justify positing formal syntactic rules or constraints as well as evidence for positing syntactic features. Section 2 briefly reviews our previous work in Oneida that argues that the kind of evidence adduced for formal syntactic constraints and syntactic features is absent in Oneida. Section 3 is the core of our paper, as it discusses many of the constructions present in Oneida and demonstrates how one can do compositional syntax without formal syntax. Section 4 argues that the little bit of formal syntax you have in Oneida does not require syntactic features (at least, as long as you have constructions as first-class grammatical objects). The paper concludes with a brief discussion in Section 5.

1 What good are syntactic features?

Syntactic features are the mainstay of syntactic theories since at least the time of American structuralism. Within HPSG or SBCG three kinds of syntactic features can be distinguished:

1. *Selectional features*: ARG-STRUC, VAL, SPEC, MOD for bounded dependants and REL, EXTRA, SLASH and the like for unbounded dependants
2. *Categorial features*: CASE, VFORM, AUX,...
3. *Other features*: ROOT, LEX, ...

One of our claims is that the grammar of Oneida does not require the introduction of any of these features on words or phrases. Since the category *other features* is heterogeneous and the need for these features less cogent than that of other features, we do not discuss them any further here and focus instead on selectional and categorial features. Both selectional and categorial features percolate from *syntactic* heads to mothers of local trees (except, of course, SPEC and MOD, which are introduced to model selection of heads by non-heads). Although percolation is not definitional of either kind of features, percolation is one of the main reasons you need syntactic features in the first place, and it is one of the consequences of syntactic selection.

1.1 What do you need categorial features for?

In most languages, categorial or part-of-speech features are needed to constrain the combination of expressions beyond semantic types. In other words, if one cannot predict what combinations are grammatical or ungrammatical on the basis of the semantic types of combining expressions, categorial features are needed. Taking well-known examples from English, nouns cannot combine with (nominative

or accusative) noun phrases, only with possessive noun phrases or PPs (see(1)); prepositions cannot combine with verbs (see (2); note that we analyse gerunds as verbal categories but not verbs following Malouf 2000); some verbs subcategorize for PPs (headed by particular prepositions) (see (3)), for particular kinds of clauses, or VPs whose main verb has a particular form (see (4)); singular count nouns require determiners (see (5)). The fact that our description of these well-known facts may not be the most appropriate or that some of the constraints may follow from more general principles is not important. What is critical is that the kinds of constraints illustrated in (1) through (5) provide the traditional motivation for positing categorial distinctions in English, for it is not clear how to *derive* these constraints from the semantic types of combining expressions no matter how semantically motivated some of these constraints may be.

(1) Bill's daughter/*Bill daughter/The book of Job/*The book Job

(2) Bob dreams of getting a new car/*get a new car

(3) John laughed *at* the idea

(4) I want for him to be happy no matter what he ends up doing

(5) Milk/A student/*student/students

1.2 What is the evidence for syntactic selection?

Most syntactic thinking since at least Adjuckiewicz (1935) assumes that a good portion of the syntax of natural languages can be characterized through a relation of selection between an expression and other dependant expressions. The evidence is multi-varied, and we only mention some of the basic kinds of evidence: Some dependants of heads are "obligatory" (see (6)); the order of dependants and heads is (relatively or partially) "fixed" (see (6)); verbs undergo valence alternations (or their phrase-structural, movement-driven equivalents) (see (7)); dependants (and heads) can enter into binding relations (including WH-dependencies) (see (8)). Scare quotes around some of the terms are meant to suggest that various analyses are possible. Again, what matters is not whether the descriptions of the facts illustrated in (6)-(8) are the most appropriate ones; rather what matters is the existence of these kinds of facts since they motivate positing relations of selection between, say, heads and dependants. Note that in many cases selection goes together with an ordering of the dependants selected by heads, as the binding of the English reflexives in (8) illustrates (see Bickel 2011 for a discussion of the kinds of evidence for (ordered) grammatical relations).

(6) *(Mary) loves *(John).

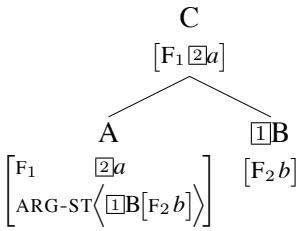
(7) John is not loved by his students.

(8) Mary loves herself.

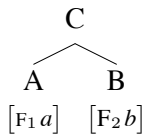
1.3 Different kinds of formal syntax

Although we have discussed models of the facts illustrated in (6) through (8) in terms of selection, the facts we discussed in Sections 1.1 and 1.2 can be modelled in two distinct ways. We can use *selectional syntactic constraints* to ensure that the presence of one daughter appropriately restricts the presence (or absence) of other daughters, their position, and their *form*. We can alternatively use *constructional syntactic constraints*, i.e. posit one-off structural patterns which may restrict the form or formal properties of daughters but do not require one daughter to select other daughters. Of course, one and the same approach to syntax can adopt both selectional and constructional constraints. In HPSG (Pollard and Sag 1994), syntax is mostly projectionist but some is constructional; in SBCG (Boas and Sag 2012), a little more syntax is constructional but selection still plays an important role. We illustrate selectional and constructional syntax in (9) and (10), respectively.

(9)



(10)



The choice of a projectionist vs. constructional syntax is not orthogonal to the issue of selection. Selecting expressions are most often heads and the mother node records whether selected expressions were realized locally. There is little doubt that syntactic selection and projectionist approaches to selection have been very successful. But, irrespective of whether one is more inclined to projectionists or constructional solutions, models of syntactic knowledge include a good bit of formal syntax and that formal syntax requires the use of syntactic features, typically for the kinds of reasons we alluded to in this section. The question is whether this is part of the design of natural languages or is merely an overwhelmingly frequent aspect of natural languages syntax. In the next section, we suggest that the latter answer is the correct one. There are languages like Oneida where syntactic features are dispensable and formal syntax is minimal: In such languages, most syntactic rules are examples of combinatorial syntax.

2 Formal syntax in Oneida

We have argued in past work that Oneida syntax is not based on syntactic selection or ordering of syntactic dependants (subjects and objects) (see Koenig and Michel-

son 2012; Koenig and Michelson 2014). We cannot recapitulate our argument in detail here. Its logic is simple: None of the evidence typically adduced in favor of syntactic selection or ordering of syntactic dependants is present in Oneida, and in the absence of such evidence the simplest model of the grammar of Oneida does not make use of these notions.

2.1 There is not much formal syntax in Oneida

In this section, we list the phenomena whose presence justifies syntactic selection or syntactic part-of-speech information and whose absence in Oneida suggests its syntax is of a different kind than that of most languages (see Koenig and Michelson 2014 for much more detail).

No syntactic selection

1. There is no requirement by words based on noun or verb stems that they have dependants, as a comparison of the Oneida discourse in (11) and its translation shows.[2]

2. There is no necessary co-indexing relation between a word's semantic arguments and expressions that further specify these arguments. (12) shows that an external referential phrase can denote a subset of the entities referenced by the pronominal prefix on the verb while (13) shows that an external referential phrase can denote a superset of the entities referenced by the pronominal prefix on the verb.

3. There is no restriction on the order of dependants headed by words based on noun or verb stems (see (14) vs. (15)). As argued for by Mithun (1987) for Mohawk, the concept of default word order has no application in Iroquoian.

4. There is no restriction on the *form* of dependants headed by words based on noun or verb stems.

5. There are no valence alternations. All (derivational) morphological operations that have a reflex in pronominal prefixes on verb stems are morphosemantic in Ackerman's (1992) sense and the morphological effect follows from the fact that they alter the meaning of the stem they apply to.

[2]Abbreviations are A agent, CAUS causative, CSL cislocative, COIN coincident, DISTR distributive, DL dualic, DP dual-plural or non-singular, DU dual, EX exclusive, FACT factual, FI feminine-indefinite, FZ feminine-zoic, FUT future, HAB habitual, JN joiner vowel, LOC locative, M masculine, NEG negative, OPT optative, P patient, PART partitive, PL plural, PNC punctual, PRES present, PROG progressive, REP repetitive, SG singular, STV stative, TRL translocative. Z/N zoic/neuter. The symbol > indicates a proto-agent acting on a proto-patient; for example, 3M.SG>1SG should be understood as 3rd person masculine singular acting on 1st person singular. Not all clitics or particles are glossed, as appropriate word glosses do not always exist. A set of robust phonological modifications occur at the end of "utterances." A pervasive utterance-final process is the devoicing of a word-final vowel or syllable, indicated by underlining.

6. There is no syntactic *ordering* (subject, object, ...) of dependants of words based on noun or verb stems. Phenomena that typically justify ordering dependants are absent in Oneida:

   (a) Principle A is irrelevant (reflexive/reciprocal marking is strictly morphological)

   (b) There is no VP ellipsis/conjunction (*pace* Baker 1996)

   (c) There is no clear evidence supporting the claim that Oneida shows weak cross-over effects (*pace* Baker 1996)

   (d) Principle C is not operative in Oneida (*pace* Baker 1996), as (16) and (17) show.[3]

(11) Né·=s wí· né·n tshiwahu·níseʔ lon-uʔwéskwani-heʔ
so it's a long time ago 3M.DP.P-enjoy-HAB

a·-hati-yʌt-a-kó·n-aʔ kʌ́·,
OPT-3M.PL.A-wood-JN-go.somewhere.to.harvest-PNC see

tahnú·=s kwí· kwahotokʌ́·u tsiʔ wa-hu-nakla·kó· tho
and=habitually just for real that FACT-3M.PL.A-move.away:PNC there

y-a-hu-náklat-eʔ tsiʔ nú· ye-hoti-yoʔtʌ́-st-aʔ.
TRL-FACT-3M.PL.A-settle-PNC where TRL-3M.DP.P-work-CAUS-HAB

'A long time ago they used to like to go cut wood, and so they would move away and they would settle over there, where they were working.' (Mercy Doxtator, Some Woodcutters Get a Visitor, recorded 1996)

(12) Nʌ kwí· wa-hy-atlihwísa-neʔ kaʔikʌ́ ló-nhah-seʔ,
so then FACT-3M.DU.A-agree-PNC this 3M.SG>3M.SG-hire-HAB

so then he (my father) and his boss agreed/planned, (Norma Kennedy, A Haunted Car, recorded 2010)

(13) tsiʔ náhteʔ wa-h-atkátho-ʔ uky-atyóha.
that what FACT-3M.SG.A-see-PNC 1DU.P-brother-in-law

'(This is my favorite story,) what my brother-in-law saw.' (Mildred Cutcut, The Hunter, recorded 1982)

(14) nʌ kwí· úska útlatsteʔ thikʌ́ Tsyó kháleʔ í· yakn-í·tlu-ʔ,
so then one time that Joe and me 1EX.DU.A-be.at.home-STV

'so then one time Joe and I were home,' ((Clifford Cornelius, A Lifetime Working, recorded 1994)

[3]Both these examples were elicited to test for Principle C. They (and similar sentences) are based on situations that occurred in the recorded texts. Our consultant considered the sentences perfectly okay. However, in certain respects they are somewhat atypical, which is perhaps not unexpected for elicited sentences.

(15) né· tsiʔ ni-ho-naʔkhwʌ́·-u thikʌ́ lakeʔnihkʌ́,
because PART-3M.SG.P-get.mad-STV that my late father

'because my father was so mad,' (Clifford Cornelius, A Lifetime Working, recorded 1994)

(16) Waʔ-utat-hlo·lí· kwí· tsiʔ yako-yo·té· aknulhá·
FACT-3FI.SG>3FI.SG-tell-PNC that 3FI.P-work:STV my mother

aʔé· Heinz Factory.
way over there Heinz Factory

'My mother told me that she was working way over at the Heinz Factory.'

(17) Wa-hak-hlo·lí· tsiʔ wa-huwá-hsle-ʔ lakeʔníha
FACT-3M.SG>1SG-tell:PNC that FACT-3>3M.SG-chase-PNC my father

nʌ ka-list-aʔké-shuʔ te-ho-tawʌlye-háti-ʔ.
when 3Z/N.SG.A-iron-LOC-DISTR DL-3M.SG.P-travel-PROG-PRES

'My father told me that she chased him when he was going along on the railway tracks.'

No *syntactic* parts of speech. Oneida has a robust notion of stem classes. Several derivational and inflectional processes allow us to distinguish between four kinds of Oneida stems: noun stems, verb stems, uninflected stems, and kin stems (see Koenig and Michelson 2010 on Oneida kin stems). But, the same is not true of Oneida words. To illustrate the difference between nominal stems and what would be putative NPs (and their N heads), we counted in the naturally produced discourses in Michelson, Kennedy, and Doxator (To appear) all referring expressions headed by words based on the four kinds of stems. Table 1 summarizes the relevant part of this corpus study: Over 60% of referring expressions (what typically would be encoded by NPs in English) are headed by words based on stems with *no* nominal morphology.

Table 1: Proportions of referring expressions according to morphology

	REs headed by words with exclusive nominal morphology	REs headed by words with some nominal morphology	REs headed by words with no nominal morphology	Total
Count	575	686	2027	3288
As % of REs	17.5%	20.8%	61.7%	100%
As % of Wds	1.9%	2.2%	6.5%	10.6%
As % of clauses				39.93%

The data in Table 1 do not provide conclusive evidence that there is no need for syntactic part-of-speech information in Oneida. In fact, there can only be negative evidence: No syntactic rule/constraint makes reference to part-of-speech, as we show in the next section. But the low percentage of referential expressions headed

by words based on noun stems (less than 40%) will, hopefully, suggest to readers how different Oneida is from many better-studied languages. The irrelevance of part-of-speech distinctions is, of course, not surprising since without syntactic selection there is little need for part-of-speech information. Similarly the absence of syntactic headedness in Oneida is not surprising since without syntactic selection syntactic headeness is hardly useful.

2.2 The little formal syntax Oneida has

The previous section showed that the typical evidence in favor of formal syntax is nowhere to be found in Oneida. This is the sense in which Oneida's syntax is, as we show in more detail in the next section, almost exclusively compositional. But, not quite. There are several syntactic constraints that are formal, i.e. that make reference to the form, ordering, or lexical identity of daughters within a constituent. We mention a few here. First, some clauses begin with the word *tsi?*, e.g. argument clauses with realis interpretations. Second, some words must co-occur with other words: *ok* must follow the word *ukha?* for an "indefinite person" meaning to be encoded. Third, question words must occur first in a clause, while argument clauses must follow the verb whose propositional argument they further specify. Crucially, constraints such as these are very restricted. They all involve particular semantic types or particular words (or classes of words) and therefore do not require the projection of syntactic features (including the projection of part-of-speech information or categorial features). Linear order constraints need only mention the semantic type of daughters or their lexical identity. So, not only does Oneida have little formal syntax, the little formal syntax it has does not require the introduction of syntactic features. In other words, there is no need for a SYN attribute in Oneida. In the next section, we show what an almost exclusively compositional syntax and one that does not include syntactic features looks like.

3 The constructions (the goods)!

A few preliminary remarks are in order. First, we leave out a couple of constructions for reasons of space. Second, the list of constructions we have identified is most probably incomplete. It has been compiled over the last few years; more recently we added to it by, together, going through a few pages of texts on a regular basis and accounting for all the constructions speakers made use of. While the list is probably incomplete, we are fairly confident that it includes the bulk of Oneida constructions and that variants of these constructions or other constructions would not significantly alter our claims. Third, if syntactic phrases are not built through syntactic selection, semantic composition must be done constructionally (including variable identification) (see Bach 1976; Klein and Sag 1985). Fourth, since there is only "existential" quantification in Oneida (no quantifiers of type $< 1, 1 >$ in the sense of Peters and Westerståhl 2006; see Koenig and Michelson 2012), we

can dispense with (generalized) quantifiers altogether, have only free variables, and get the existential force of variables from the anchoring of atomic formulas (à la Kamp 1981; Kamp and Reyle 1993), as shown in (18). Fifth, we adapt (quite liberally) a semantic underspecification approach called *Lexical Resource Semantics* (see Richter and Sailer 2004). Semantic underspecification is quite useful when all semantic combinatorics is constructional, although we do not know if it is truly needed.

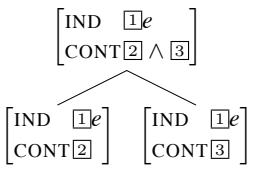
(18) $[\![P(x_1, \ldots, x_n)]\!]^M = 1$ iff
there is an anchoring g such that $< g(x_1), \ldots, g(x_n) >$ that is in the denotation of P.

The following sections present the list of Oneida constructions. Our goal is two-fold. First, illustrate what compositional syntax looks like; second, demonstrate that we can model Oneida syntax without the use of syntactic features (SYN in HPSG parlance). For space reasons, we give an example of each construction with its English translation and the meaning of the mother node, but do not discuss the example nor provide interlinear glosses; parts of the sentences that exemplify the construction are in bold. We use the terms *entity expression* and *situation expression* for expressions denoting or describing entities and situations, respectively. Our analysis of Oneida syntactic constructions thus relies on a fundamental distinction between two semantic types. We leave a justification of these two particular semantic types to another venue. Finally, to increase readibility, we indicate graphically the semantic import of constructions on the semantic translation of the (relevant portion of) examples. ▣ indicates which of the daughters' index is the index of the entire expression; ▭ indicates identification of indices across daughters; ▣ indicates that a predication was added by the construction itself; finally, ▭ indicates the output of "previous" semantic composition.

3.1 Entity expression apposition

This construction states (1) that two entity expressions can, generally, co-occur in either order, (2) that the meaning of the whole bears the index of both daughters (which must be the same) and (3) that the meaning of the whole is the conjunction of the contents of the daughters. It is represented in (19) and an example is provided in (20a). Note that this construction applies more widely than apposition in English, as Oneida demonstratives are fully referential entity expressions that can occur in apposition to another fully referential entity expression, as shown in the example in (21).

(19)



(20) a. **Lake?kʌ́ha Leo**, né· kʌs né· wahatkátho? thikʌ́
‘**My brother Leo** saw it’
(Rose Antone, What My Brother Saw, recorded 2011)

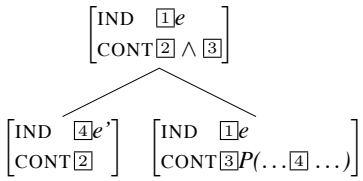
b. brother’(‘I’, x) ∧ Leo’(x)

(21) nʌ ki? ok kwí· wa?ekwe?talu·kó· **ka?ikʌ́ kaná·talok**,
‘and then she cut into chunks **this bread** ,’ (Norma Kennedy, The Bird, recorded 2008)

3.2 Entity expression adjunction

This construction states (1) that two entity expressions can co-occur in either order (which is simplifying somewhat for reasons of space), (2) that the meaning of the whole bears the index of one daughter (the semantic head), (3) that the index of the other daughter is an argument of the semantic head’s content, and (4) that the meaning of the whole is the conjunction of the daughters’ contents. A representation of the construction is provided in (22) and an example in (23a). This construction illustrates the importance of INDex selection for semantic composition in a language where syntactic selection and functional composition do not ensure the proper matching of variables and argument positions. The construction must specify that the index of one of the daughters is the index of the entire expression so that hearers can determine upon hearing (23a) who died, the person referred to via *aknulhá·* or the person referred to via *onulha?kʌ́*.

(22)



(23) a. Tahnú· **aknulhá· onulha?kʌ́** tshahanáklate? Bill ne? thó·ne? né· t-yakaw-ʌhe·yú. ‘And **my mother’s mother** died when Bill was born.’
(Olive Elm, Visits to my Auntie’s, recorded 1993))

b. mother’(‘I’, x) ∧ late.mother’(x , y)

3.3 Clausal constructions

Oneida has several constructions that build clauses out of a situation expression and various other kinds of expressions. Figure 1 summarizes the hierarchy of clausal constructions we mention in this paper.

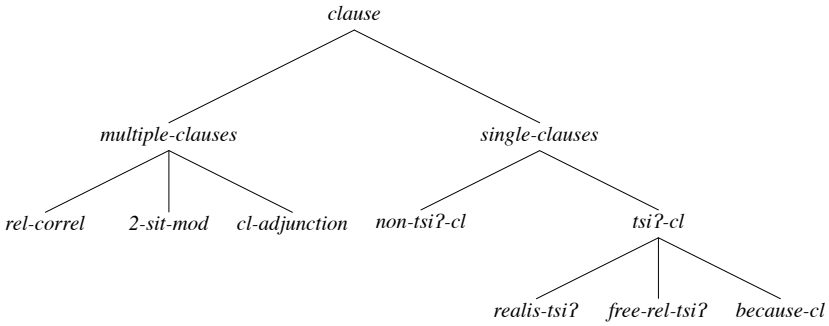


Figure 1: A hierarchy of Oneida clausal constructions

Single clauses. This is the basic construction Oneida uses to build clauses. Its formulation is complex because our analysis of clauses in Oneida is "flatter" than the analysis of clauses in many approaches, and as a consequence expressions of distinct semantic types are sisters to the semantic head. Flatter VPs have been posited for a long time in HPSG (see Bouma, Malouf, and Sag 2001 or Kim and Sag 2002, among others), but our analysis of Oneida clauses is even flatter. We have two main reasons for adopting such a flat structure. First, we know of no evidence to posit more structure; in the absence of such evidence positing additional structure would be imposing onto Oneida what is relevant to other languages. Second, the order of expressions of distinct semantic types can vary and the number of possible orderings is quite large (see Section 4). We could, of course, make use of domains (see Kathol 1999 among others), but in the absence of evidence for more hierarchical structure the introduction of this rather heavy mechanism would be *ad hoc*. Furthermore, the motivation for distinguishing linearization issues from "structural" combinations, which is at the root of linearization-based approaches, is absent in Oneida if, as we argue, there is no syntactic selection. Oneida thus lacks the very motivation for positing a level of representation in which functors and arguments combine that is distinct from their linear order. The net effect of a flat clausal structure and the absence of functional application (or its derivatives) is that semantic composition is case-based for this construction, as shown in (24). The construction is represented in (25) and an example is given in (26a). Note that semantic underspecification makes it relatively easy to have a case-based definition of semantic composition.

(24) A situation expression can consist of a situation-describing word (the semantic head) preceded by zero or more expressions and followed by zero or more expressions. The index of the whole is that of the situation-describing word and the semantic content of the whole is determined as follows:

1. If a non-head daughter is an entity expression, its index must be included in the content of the semantic head (co-indexed with one of the head's argument or an argument of an argument ...of the head), and the content of the whole includes the conjunction of the content

of the non-head daughter and the content of the head

2. If a non-head daughter is a situation expression, its content must be included in the content of the head
3. If a non-head daughter is a time or location expression, it takes the index of the semantic head as argument, and the content of the whole includes the conjunction of the content of the non-head daughter and the head
4. If a non-head daughter is a scopal operator (e.g. negation) its argument must include the content of all expressions to its right that are scope sensitive

(25) $\begin{bmatrix}\text{IND} & \boxed{1}s\\ \text{CONT} & \alpha\end{bmatrix} \rightarrow (\begin{bmatrix}\text{IND} & none\\ \text{CONT} & Op(\beta)\end{bmatrix})^*, (\begin{bmatrix}\text{IND} & \boxed{2}e\\ \text{CONT} & \boxed{3}\end{bmatrix})^*, \begin{bmatrix}\text{IND} & \boxed{1}\\ \text{CONT} & \boxed{4}P(\boxed{1},\ldots,\boxed{5})\end{bmatrix},$
$(\begin{bmatrix}\text{IND} & l\\ \text{CONT} & \boxed{6}P'(\boxed{1},\ldots)\end{bmatrix})^*, (\begin{bmatrix}\text{IND} & s'\\ \text{CONT} & \boxed{5}\end{bmatrix})$

$\boxed{2} \lhd \boxed{4}$; $(\boxed{4} \wedge \boxed{6}) \lhd \alpha$; $(\boxed{3} \wedge \boxed{4}) \lhd \alpha$; α = *leftmost*(Op(β)) (where *leftmost* selects the semantic content of the leftmost daughter that contributes a scope-sensitive operator)

(26) a. **Lake?kʌ́ha Leo**, né· kʌs né· **wahatkátho?** thikʌ́
'My brother Leo saw it' (Rose Antone, What My Brother Saw, recorded 2011)

b. $\boxed{\text{brother'('I', } \boxed{\text{x}}) \wedge \text{Leo'}(\boxed{\text{x}})} \wedge \text{see'}(\boxed{\text{s}}, \boxed{\text{x}}, \text{y})$

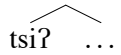
The statement of the construction in (24) and (25) stipulates that scope-sensitive expressions follow a left-to-right order so that expressions on the left take the semantic content of expressions on the right as arguments. Determining which expressions are scope-sensitive is a difficult issue we cannot go into in this paper and our assumption that scopal relations follows a left-to-right order should be considered provisional. Sentences (27) and (28) illustrate the semantic effect of inverting the order of the quantificational expression *akwekú* and the negative particle *yah*, which partially motivates our tentative hypothesis.

(27) nʌ kyuni? wí· né· **akwekú yah** kánike? te?shʌ·né·se?.
'I guess **all of them** (those named in the preceding sentence) are **not** around anymore.' (Pearl Cornelius, Family and Friends, recorded 1993)

(28) **Yah akwekú** tehone·ká·se? onu?úseli?.
'**Not everyone** likes squash (but some do like it).' (Elicited)

One subset of clausal constructions consist of constructions that begin with *tsi?*; they are all subtypes of the *tsi?-cl* construction whose definition (minus inherited properties, of course) is provided in (29). This is the first construction we have mentioned that includes a formal component, i.e. a constraint that restricts the form of one of the daughters.

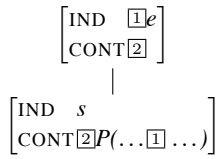
(29) *tsi?-cl* ⇒



3.4 Relative clause formation (type-shifting)

Oneida has three kinds of relative clauses, internally-headed relatives, free relatives, and relative-correlatives. We only discuss internally-headed clauses and relative-correlative constructions in this paper. Relative clauses in Oneida have a purely semantic effect, as the language does not require syntactic part-of-speech information: Relative clauses type-shift a situation expression into an entity expression, provided that this entity expression is an argument (sometimes a semantic adjunct) of the predicate associated with the situation expression. The construction states (1) that an entity expression can have as sole daughter a situation expression, (2) that the content of the whole is that of the daughter, (3) that its index is that of one argument of the content. A representation of the construction is provided in (30) and an example in (31a). Interestingly, type-shifting internally-headed relative clauses (or their lexicalized equivalents) is used to encode quantification in Oneida (see (32a) and Koenig and Michelson 2012)

(30)



(31) a. yah né· té·yʌlhe? a·yutekhu·ní· **kʌ? niyaká·,**
'the little one doesn't want to eat,' (Olive Elm, Visits to My Auntie's, recorded 1993)

b. small'(s, x)

(32) a. **Áhsʌ nikanláhtake** ʌtésku?
'**The [tobacco] leaves that amount to three**, you are to hand them to me.' [You are to hand me three leaves] (Olive Elm, Learning to Work in Tobacco, recorded 1998)

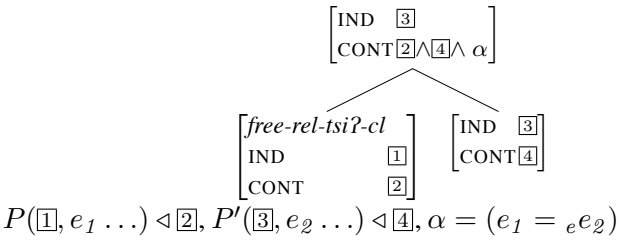
b. leaf'(x) ∧ amount'(s, x , y) ∧ three'(y)

3.5 Relative-correlative construction

Oneida has an interesting relative-correlative construction. The construction states (1) that a clause can consist of two clauses, a clause that describes a situation and a free-relative *tsi?* clause, (2) that the content of the clause is the conjunction of the contents of each clause together with an equality between the index of an entity expression within the situation expression clause and the index of the free relative,

(3) that the index of the whole is that of one entity in the clause that is not the free relative clause. A representation of the construction is provided in (33) and an example in (34a).

(33)



(34) a. **tho yahunáklate? tsi? nú· yehotiyo?tʌsta?**.
'**they would settle over there, where they were working**.' (Mercy Doxtator, Some Woodcutters Get a Visitor, recorded 1996)

b. settle'(s , x, l) ∧ work'(s', x, l') ∧ l=$_o$l'

3.6 Modification

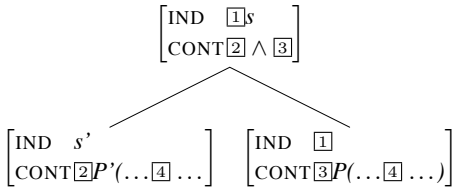
Oneida has several constructions encoding semantic modification in a broad sense of the term. We describe a few of them in this section.

Two situations modification. We begin with an interesting construction illustrated in (35a). This construction states (1) that two situation expressions (that belong to certain categories; e.g. one of them is a wearing situation) can combine, (2) that the meaning of the whole is the conjunction of the meaning of the parts, (3) that the index of the whole is the index of one of the expressions, and (4) that the content of each expression must include a shared argument. A representation of the construction is provided in (36). Readers might wonder why we do not analyze the text in (35a) as a sequence of two independent sentences. We have two reasons for tentatively assuming that the kind of sentences illustrated in (35a) exemplifies a stored pattern, i.e. a construction. First, the order of clauses would be pragmatically odd if the two clauses did not form a construction, since what is shiny is only introduced in the second clause (see the oddity of the English discourse *They$_i$ are shiny. She is wearing shoes$_i$.*) Second, it seems this pattern is restricted to a couple of semantic classes of relations, in particular wearing relations (among perhaps a few others), a restriction that seems incompatible with the assumption that we are dealing with two independent clauses.

(35) a. Kwahikʌ́ **teyostaláthe? teyakohtáli?**.
'**She's wearing shiny shoes**.' (lit. really it's shiny she's wearing shoes) (Georgina Nicholas, The Flirt, recorded 1980)

b. shiny'(s', [y]) ∧ shoes'(y) ∧ wear'([s], x, [y])

(36)

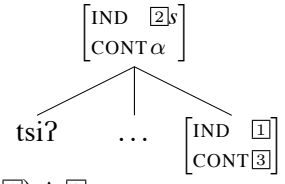


3.7 "Adjunct"-clauses

Until now, the semantic import of the constructions we have encountered was to identify variables in the semantic content of daughters and conjoin the semantic content of the daughters or embed the semantic content of one daughter into part of the content of another daughter (i.e., the semantic content of one daughter is an argument of the content of another daughter). We now consider a construction that adds a specific predication on top of the predications contributed by each of the daughters.

Adding the relation between situations. We define a *because-cl* and other kinds of "adjunct" clauses as an expression that describes a causal relation between two situations one of which is specified by the situation expression that is part of the *because-cl*. (37) represents the construction and sentence (38a) is an example of the construction.

(37)



α = causal-rel([2], s', [1]) ∧ [3]

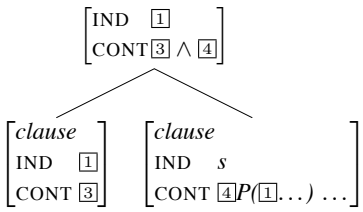
(38) a. ya·wét kyuhte wí· yakotyanlustákhwaʔ ká·slet, **né· tsiʔ** tho **lawʌheyú úhkaʔ ok**.
'it was kind of like a haunted car, **because someone died** there.' (Norma Kennedy, A Haunted Car, recorded 2010)

b. cause'([s], s_1, s_2) ∧ die'(s_2, x)

Clausal situation modifiers. The last construction we discuss in this paper puts together two clauses that describe situations. The construction states (1) that a situation expression can consist of two clauses that describe situation expressions, (2) that the content of the whole is the conjunction of the content of the two situation

expressions, (3) that the index of the whole is the index of one of them (let us call it *the main clause*), and (4) that the index of the main clause is an argument of part of the content of the other clause. The construction is represented in (39) and an example is provided in (40a).

(39)



(40) a. ya·wét kyuhte wí· **yakotyanlustákhwaʔ ká·slet**, **né· tsiʔ** tho **lawʌheyú úhkaʔ ok**.
'it was kind of like a haunted car, because someone died there.'
(Norma Kennedy, A Haunted Car, recorded 2010)

b. haunted(s_1 , y) ∧ car'(y) ∧ cause'(s, s_1, s_2) ∧ die'(s_2, x)

4 The fun aspects of the compositional and formal syntax of Oneida

The previous section described the major Oneida constructions we know of; the constructions we omitted for reasons of space have similar characteristics. The crucial aspect of the descriptions of these constructions is that no syntactic features were needed to model all the constructions we listed and get the semantics right. The fragment of Oneida we described thus constitutes a good example of what compositional syntax looks like. Interestingly, the number of constructions we needed was not that numerous. The fact that in the absence of syntactic selection we could not rely on some very general constructions (Head-Complement; Head-Subject, . . .) did not lead to a proliferation of constructions, as one might have feared. The constructions were also "minimal" and included nothing but the construction's semantics, except in the case of *tsiʔ* clauses: Specifying the index and content of the combination of expressions, identifying variables across the contents of the combining expressions, adding a predication constructionally in a few cases was all that was needed. But Oneida syntax also requires some idiosyncratic ordering constraints, some a little odd. Some simple linear constraints are stated in (41)-(43). It should be rather obvious how such constraints can be stated within HPSG. But Oneida also includes a large array of linear order constraints associated with particles. Because particles and their orders are an important characteristic property of Oneida, the rest of this section is devoted to a brief overview of particles in Oneida.

(41) A word that introduces a question variable (Gronendijk and Stockhoff 1997) must be clause initial.

(42) An argument clause must follow the situation expression it further specifies an argument of.

(43) In a *tsi?-cl*, *tsi?* must be initial.

Particles in Oneida are *a morphological class*, the class of uninflected words. In Michelson and Doxtator (2002), there are around 170 particles, 435 noun stems, and 2775 verb stems. Particles belong to various semantic types, including referring expressions (first person and second person pronouns, for example, are particles). Particles need not be utterance- or clause-initial, but they can "bunch together." There are 165 distinct particles in Michelson, Kennedy, and Doxtator (To appear) and there are 2,059 *distinct* utterance-initial sequences of particles (it should be kept in mind that particles need not be utterance- or clause-initial, though). The order of particles is not arbitrary, but it is particularly complex. First, entity expressions and situation expressions can both occur between particles. Second, in the 2,059 distinct sequences of utterance-initial particles in Michelson, Kennedy, and Doxtator (To appear) there are on average for *each* particle, 10 particles that *only* occur before it, 10 particles that *only* occur after it, and 12 particles that occur *both* before and after it. Third, a very preliminary examination of particle orders suggests that scope follows a left-to-right linear order for particles whose meaning is scope-sensitive. This is what we model in (24). Now, given the number of possible combinations of order of particles and entity or situation expressions, this conclusion is provisional. But looking at the scope of all sequences of utterance-initial particles that include the negative particle *yah*, it appears correct. Fourth, "discourse particles" must be excluded from any order-sensitive scopal relations, as their semantic type should make them appear before, for example, particles that denote propositional operators, but they need not precede propositional operators (and sometimes are required to follow them). Fifth, some particles are complex (think compounds) in that the meaning they convey require *two* phonological words. These words occur in a strict order, but in some cases the two components need not be adjacent (see (44) for *úhka? . . . ok* 'someone'). A statement of the constraint needed to model the strict ordering required by *úhka? . . . ok* (and other similar words) is provided in (45).

(44) **úhka?** ki? **ok** uhte wí· luwa?áshʌ? thikʌ́,
'they stabbed **someone**,' (i.e. someone got stabbed) (Norma Kennedy, A Haunted Car, recorded 2010)

(45) If *úhka?* is a daughter and its content is that of an animate indefinite, *ok* must also be a daughter and must follow it.

5 Discussion

The goal of this paper was to distinguish two kinds of constraints syntactic rules can include, constraints on how to compose the meaning of two or more expressions and constraints on the form of combining expressions. Our current syntactic vocabulary and syntactic thinking is built around the second, formal, kind of constraints and particularly concepts of syntatic selection and percolation of categorial/selectional information. Without selection and category information, syntactic features are almost useless and syntax reduces (almost exclusively!) to what we call *compositional syntax*. The bulk of this paper described the broad outlines of the syntax of a language that comes close to being a strictly compositional syntax language and whose formal syntax does not require the introduction of syntactic features. In this paper's closing paragraphs, we want to place our work on Oneida in the larger context of recent work on the architecture of grammars.

Jackendoff and Wittenberg (2014) describes grammars without syntax of increasing complexities. In their terminology, Oneida would exemplify a language with a recursive phrase grammar, but that is not a 'fully complex' language, as it does not have syntax in their sense (although it is quite complex morphologically, see Lounsbury 1953 and Koenig and Michelson 2015). From their perspective, the sketch of Oneida we present constitutes a demonstration of how much you can do 'without syntax.' To some extent, the difference between our way of describing Oneida and how they would describe it is terminological: What we call *compositional syntax* corresponds to their semantic structure cum interface rules between phonology and meaning, and what we call *formal syntax* corresponds to syntax proper in their approach. But slightly more than terminology is at stake here.

Jackendoff and Wittenberg's approach is implicitly and at times explicitly, teleological: Their grammatical hierarchy is meant to correspond partly to the evolution of language, as their discussion of creoles, homesigning, and young children's grammars makes clear. There is little doubt that languages that consist only of one-word utterances or two-word utterances are simpler than English. But we do not believe that the difference in complexity between compositional syntax and formal syntax is on a par with differences in complexity between two-word grammars and recursive grammars. If we define grammatical expressiveness as the range of semantic combinations licensed through grammatical means, Oneida grammatical expressiveness is roughly on a par with English. It does not leave much more to pragmatic enrichment than English does. Oneida's lack of formal syntax merely means that it lacks the crud that formal syntax has grafted on compositional syntax by vagaries of history. For us, what Oneida syntax illustrates is not a less complex grammatical system (one without syntax), but rather that what linguists typically think of as syntax is actually not the essential role of syntax in human languages. Compositional syntax is what is essential to human linguistic abilities; syntactic features are not, even though they are needed in the overwhelming majority of cases and an important aspect of syntactic theory.

References

Ackerman, F. (1992). Complex predicates and morpholexical relatedness: Locative alternation in Hungarian. In I. Sag and A. Szabolcsi (Eds.), *Lexical Matters*, pp. 55–83. Stanford: CSLI Publications.

Bach, E. (1976). An extension of classical transformational grammar. In *Problems of Linguistic Metatheory (Proceedings of the 1976 Conference)*, Michigan State University, pp. 183–224.

Baker, M. (1996). *The Polysynthesis Parameter*. Oxford: Oxford University Press.

Bickel, B. (2011). Grammatical relations typology. In J. J. Song (Ed.), *The Oxford Handbook of Language Typology*, pp. 399–444. Oxford: Oxford University Press.

Boas, H. and I. Sag (Eds.) (2012). *Sign-Based Construction Grammar*. Stanford, CA: CSLI Publications.

Chafe, W. (In press). *A Grammar of the Seneca Language*. Berkeley: University of California Press.

Groenendijk, J. and M. Stokhof (1997). Questions. In J. Van Benthem and A. Ter Meulen (Eds.), *Handbook of Logic and Language*, pp. 1055–1124. Cambridge, MA: MIT Press.

Jackendoff, R. and E. Wittenberg (2014). What you can say without syntax: A hierarchy of grammatical complexity. Tufts University.

Kamp, H. (1981). A theory of truth and semantic representation. In A. G. Groenendijk, T. M. Janssen, and M. B. Stockhof (Eds.), *Formal Methods in the Study of Language*, pp. 1–41. Amsterdam: Foris.

Kamp, H. and U. Reyle (1993). *From Discourse to Logic*. Dordrecht: Kluwer.

Kathol, A. (1999). *Linearization-based Syntax*. Oxford: Oxford University Press.

Klein, E. and I. Sag (1985). Type-driven translation. *Linguistics and Philosophy 8*, 163–201.

Koenig, J.-P. and K. Michelson (2010). Argument structure of Oneida kinship terms. *International Journal of American Linguistics 76*, 169–205.

Koenig, J.-P. and K. Michelson (2012). The (non)universality of syntactic selection and functional application. In C. Piñón (Ed.), *Empirical Studies in Syntax and Semantics*, Volume 9, pp. 185–205. Paris: Centre National de la Recherche Scientifique.

Koenig, J.-P. and K. Michelson (2014). Invariance in argument realization: The case of Iroquoian. *Language*.

Koenig, J.-P. and K. Michelson (2015). Morphological complexity à la Oneida. In M. Baerman and G. Corbett (Eds.), *Understanding and Measuring Morphological Complexity*. Oxford: Oxford University Press.

Lounsbury, F. (1953). *Oneida Verb Morphology. Yale University Publications in Anthropology 48*. New Haven, CT: Yale University Press.

Michelson, K. (2011). Iroquoian languages. In M. Aronoff (Ed.), *Oxford Bibliographies Online: Linguistics*. New York: Oxford University Press.

Michelson, K. and M. Doxtator (2002). *Oneida-English/English-Oneida Dictionary*. Toronto: University of Toronto Press.

Michelson, K., N. Kennedy, and M. Doxtator. (To appear). *Glimpses of Oneida Life*. Toronto: University of Toronto Press.

Mithun, M. (1987). Is basic word order universal? Grounding and coherence in discourse. In R. Tomlin (Ed.), *Typological Studies in Language*, Volume 11, pp. 281–328. Amsterdam: John Benjamins.

Peters, S. and D. Westerståhl (2006). *Quantifiers in Language and Logic*. Oxford: Clarendon Press.

Pollard, C. and I. Sag (1994). *Head-Driven Phrase-Structure Grammar*. Chicago, IL: Chicago University Press.

Richter, F. and M. Sailer (2004). Basic concepts of lexical resource semantics. In *ESSLLI 2003 Course Material*, Volume 5 of *Collegium Logicum*. Kurt Gödel Society Wien.

Two types of serial verb constructions in Korean: Subject-sharing and index-sharing

Juwon Lee
The University of Texas at Austin

Proceedings of the 21st International Conference on
Head-Driven Phrase Structure Grammar

University at Buffalo

Stefan Müller (Editor)

2014

Stanford, CA: CSLI Publications

pages 135–155

Lee, Juwon. 2014. Two types of serial verb constructions in Korean: Subject-sharing and index-sharing. In Stefan Müller (ed.), *Proceedings of the 21st International Conference on Head-Driven Phrase Structure Grammar, University at Buffalo*, 135–155. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2014.8.
CC-BY 4.0

Abstract

In this paper I present an account for the lexical passive Serial Verb Constructions (SVCs) in Korean. Regarding the issue of how the arguments of an SVC are realized, I propose two hypotheses: i) Korean SVCs are broadly classified into two types, subject-sharing SVCs where the subject is structure-shared by the verbs and index-sharing SVCs where only indices of semantic arguments are structure-shared by the verbs, and ii) a semantic argument sharing is a general requirement of SVCs in Korean. I also argue that an argument composition analysis can accommodate such the new data as the lexical passive SVCs in a simple manner compared to other alternative derivational analyses.

1. Introduction

Serial verb construction (SVC) is a structure consisting of more than two component verbs but denotes what is conceptualized as a single event, and it is an important part of the study of complex predicates. A central issue of SVC is how the arguments of the component verbs of an SVC are realized in a sentence. In the literature, it is generally assumed that the constituent verbs of an SVC share the subject (Foley and Olson 1985, Sebba 1987, Lee 1992, Andrews 1997, Chung and Kim 2008, Müller and Lipenkova 2009, Kim 2010, Lee 2011, among others) or they share the object (Baker 1989) or an internal argument (e.g. themes, instruments, goals) (Collins 1997). In the Korean SVC (1a), for instance, both the subject *akma-ka* 'demon-Nom' and the object *wenswungi-lul* 'monkey-Acc' are shared by the constituent verbs, but in (1b) only the subject *Jane-i* 'Jane-Nom' is shared by the first verb (V1) *chac-a* 'search.for-Comp' and the second verb (V2) *ka-ss-ta* 'go-Pst-Dec' (contra Baker 1989, Collins 1997).

(1) a. akma-ka wenswungi-lul **cap**-a **mek**-ess-ta.
demon-Nom monkey-Acc catch-Comp eat-Pst-Dec[1]
'The demon caught the monkey$_j$ and then ate it$_j$.'

* I would like to thank Stephen Wechsler for his insightful comments on the draft of this paper. The helpful comments, questions and criticisms of the anonymous reviewers and audiences (Doug Arnold, Philippa Cook, Berthold Crysmann, Ray Jackendoff, Philip LeSourd, Stefan Müller, Sang-Hee Park, Robert D. Van Valin, Jr.) of the HPSG 2014 Conference at University at Buffalo, The State University of New York are also gratefully acknowledged. Any remaining errors are of course mine.

1 Abbreviations: Acc = Accusative, Comp = Complementizer, Conj = Conjunction, Dec = Declarative, Gen = Genitive, Neg = Negative, Nom = Nominative, Pass = Passive, Plu = Plural, Pres = Present, Pst = Past

b. Jane-i hakkyo-ey Tom-ul **chac**-a **ka**-ss-ta.
Jane-Nom school-To Tom-Acc search.for-Comp go-Pst-Dec
'Jane went to the school searching for Tom.'

Interestingly, if we examine the canonical SVC (1a) more carefully, we can find that it has its passive counterparts in (2) below which violate the subject- and object-sharing, and the argument saturation. Only the passive verb with *hi* (a passive affix in Korean) in (2) can take the NPs as its subject and complement, whose CASE values are compatible only with it; the other active form of the verb doesn't share them in the sentences. That it, in (1a), the V1 *cap-a* 'catch' shares the nominative subject and accusative object with the V2 *mek-ess-ta* 'eat-Pst-Dec', but in (2a), the same V1 *cap-a* 'catch' does not have in the sentence its nominative subject and accusative object (which should be *akma-ka* 'demon-Nom' and *wenswungi-lul* 'monkey-Acc', respectively). This entails no subject and complement sharing and no argument saturation in the SVC. The same kind of problems applies to the V2 *mek-ess-eyo* 'eat-Pst-Dec' in (2b).

(2) a. wenswungi-ka akma-eykey cap-a mek-**hi**-ess-ta.
monkey-Nom demon-By catch-Comp eat-Pass-Pst-Dec
'The monkey was caught and then eaten by the demon.'

b. wenswungi-ka akma-eykey cap-**hi**-e mek-ess-eyo.
monkey-Nom demon-By catch-Pass-Comp eat-Pst-Dec
'The monkey was caught and then eaten by the demon.'

Note that the SVCs in (2) are largely compositional, since the lexical semantics of the verbs compose the basic meanings of the verbal serializations (with the constructional meaning of the SVCs, a sequence of the subevents denoted by the verbs). They are not idiomatic or metaphorical: the verb *mek-* 'eat' has many metaphorical uses in Korean as shown in (3), but the SVCs in (2) denote the event of the monkey being caught and then literally eaten by the demon.

(3) ku-ka noymwul-ul/ ton-ul **mek**-ess-ta.
he-Nom bribe-Acc/ money-Acc eat-Pst-Dec
'He received the bribe/ money.'

Due to the lack of the subject and complement required for the V1 in (2a) and the V2 in (2b), the SVCs are predicted to be ill-formed in the literature.

However, they are well-formed SVCs in Korean. For some native speakers of Korean, the SVC (2b) may sound somewhat awkward, but in the next section I present the empirical grounds of the SVCs like (2b).

The question that naturally arises is then how to account for the lexical passive SVCs in (2).[2] I propose two hypotheses: i) Korean SVCs are broadly classified into two types, subject-sharing SVCs like (1) where the subject is structure-shared by the verbs and index-sharing SVCs such as (2) where only indices of semantic arguments are structure-shared by the verbs, and ii) an argument index sharing is a general requirement of SVCs in Korean. I also argue that an argument composition analysis can accommodate the novel data like (2b) straightforwardly compared to other alternative derivational analyses.

2. The existence of index-sharing SVCs

In this section I explicitly show that the sentence (2b) is a real SVC; it doesn't belong to other constructions like coordination, subordination, resultative or auxiliary construction.

2.1 Basic properties of SVCs

It seems not easy to precisely define SVC of all serializing languages, and scholars may have different ideas about what is SVC and what is not. Van Valin (2005) classifies English resultative (construction type: serial verb and juncture: core) and English obligatory control constructions (construction type: serial verb and juncture: nuclear) as a type of serial verb. Resultative constructions in Thai can be arguably a kind of SVC (Thepkanjana and Uehara 2009). Coordination, subordination and auxiliary constructions share some grammatical properties with typical SVCs.

So all these related constructions can be plausible candidates for the identity of the sentence (2b). I here discuss three main properties of canonical Korean SVCs based on which I argue that the sentences like (2b) are genuine SVCs in Korean.

First, the negative marker *an* that immediately precedes V1 can have wide scope over V1 and V2 as shown in (4a). In the SVC (4b), where the same form of the verbal serialization has the idiomatic interpretation of 'forget', the negative marker *an* also has wide scope over V1 and V2.

---

2 There are two types of passives in Korean, lexical passives using a passive affix (*-i*, *-hi*, *-li*, *-ki*) as in (2) and syntactic passives using the passive auxiliary *ci-* as shown in the following:

i) wenswungi-ka akma-eykey cap-a mek-e **ci**-ess-ta.
monkey-Nom demon-By catch-Comp eat-Comp become-Pst-Dec
'The monkey became caught and then eaten by the demon.'

In this paper I focus on the lexical passive constructions.

(4) a. Tom-i sakwa-lul **an** kka mek-ess-ta.
Tom-Nom apple-Acc Neg peel.Comp eat-Pst-Dec
'It is not the case that Tom peeled the apple$_j$ and then ate it$_j$.'

b. Ryan-i yaksok-ul **an** kka mek-ess-ta.
Ryan-Nom promise-Acc Neg peel.Comp eat-Pst-Dec
'Ryan did not forget his promise.'

By contrast, the negative marker *an* cannot have wide scope in the coordination (5a), subordination (5b), and resultative construction (5c). But the auxiliary construction (5d) allows wide scope of *an*.

(5) a. Tom-i sakwa-lul **an** kka-ko mek-ess-ta.
Tom-Nom apple-Acc Neg peel-and eat-Pst-Dec
'Tom did not peel the apple$_j$ and ate it$_j$.'

b. Jane-i Mary-lul **an** yeyppu-ta-ko sayngkakhay-ss-ta.
Jane-Nom Mary-Acc Neg pretty-Dec-Comp think-Pst-Dec
'Jane thought that Mary was not pretty.'

c. Hank-ka soy-lul **an** pyengpyengha-key twutulki-ess-ta.
Hank-Nom metal-Acc Neg flat-Key hammer-Pst-Dec
'Hank hammered the metal not flat.'

d. Bob-i mwul-lul **an** masi-ko siph-ess-ta.
Bob-Nom water-Acc Neg drink-Comp want-Pst-Dec
'It is not the case that Bob wanted to drink water.'

Second, a separate tense marking on V1 is not permitted in SVC, whether it be non-idiomatic as in (6a) or idiomatic as in (6b).

(6) a. Tom-i koki-lul kwu(*-**ess**)-e mek-**ess**-ta.
Tom-Nom meat-Acc bake-Pst-Comp eat-Pst-Dec
'Tom baked the meat$_j$ and then ate it$_j$.'

b. Tom-i Jane-ul kwu(*-**ess**)-e salm-**ass**-ta.
Tom-Nom Jane-Acc bake-Pst-Comp boil-Pst-Dec
'Tom coaxed Jane.'

The first verb *kwu-* 'bake' in coordination (7a) and the adjective *yeypp-* 'beautiful' in the embedded clause of the subordination (7b) can also have the separate tense marking *-ess* '-Pst'. But the secondary predicate *pyengpyengha-* 'flat', which appears before the verb in the resultative (7c), and the main verb *masi-* 'drink' in the auxiliary construction (7d) cannot have a separate tense marking.

(7) a. Tom-i koki-lul kwu-**ess-ko** mek-**ess**-ta.
Tom-Nom meat-Acc bake-Pst-and eat-Pst-Dec
'Tom baked the meat$_j$ and ate it$_j$.'

b. Tom-i aki-ka yeypp-**ess**-ta-ko sayngkakhay-**ss**-ta.
Tom-Nom baby-Acc beautiful-Pst-Comp think-Pst-Dec
'Tom thought that the baby was beautiful.'

c. Hank-ka soy-lul pyengpyengha(*-**yess)**-key twutulki-**ess**-ta.
Hank-Nom metal-Acc flat-Pst-Key hammer-Pst-Dec
'Hank hammered the metal flat.'

d. Jane-i mwul-lul masi(*-**ess)**-ko siph-**ess**-ta.
Jane-Nom water-Acc drink-Pst-Comp want-Pst-Dec
'Jane wanted to drink water.'

Third, the delimiter *-man* 'only' can be attached to the first verb in the non-idiomatic SVC (8a), but not in the idiomatic SVC (8b). Another delimiter *-to* 'also' has the same distributions as *-man* 'only' in SVCs.

(8) a. Tom-i hakkyo-ey kel-e-**man** ka-ss-ta.
Tom-Nom school-to walk-Comp-only go-Pst-Dec
'Tom went to school only by walking.'

b. Ryan-i yaksok-ul kka(*-**man)** mek-ess-ta.
Ryan-Nom promise-Acc peel.Comp-only eat-Pst-Dec
'Ryan forgot his promise.'

The first verb of coordination (9a), the adjective in the embedded clause of the subordination (9b), the secondary predicate of the resultative (9c) and the main verb of the auxiliary construction (9d) can also have the delimiter *-man* 'only'. Note that it can be replaced with *-to* 'also' in those sentences.

(9) a. Tom-i pica-lul sa-ko-**man** ka-ss-ta.
Tom-Nom pizza-Acc buy-and-only go-Pst-Dec
'Tom only bought a pizza, and went.'

b. Tom-i aki-ka yeyppu-ta-ko-**man** sayngkakhay-ss-ta.
Tom-Nom baby-Acc beautiful-Dec-Comp-only think-Pst-Dec
'Tom thought that the baby was only beautiful.'

c. Hank-ka soy-lul pyengpyengha-key-**man** twutulki-ess-ta.
Hank-Nom metal-Acc flat-Key-only hammer-Pst-Dec
'Hank hammered the metal only flat.'

d. Jane-i mwul-lul masi-ko-**man** siph-ess-ta.
Jane-Nom water-Acc drink-Comp-only want-Pst-Dec
'Jane wanted only to drink water.'

The three properties of the constructions are summarized in the following table (1 = Yes, 0 = No):

(10) Three properties of the constructions:

	Wide negation scope	Separate tense marking	Delimiter
SVC (non-idiomatic)	1	0	1
SVC (idiomatic)	1	0	0
Coordination/Subordination	0	1	1
Resultative construction	0	0	1
Auxiliary construction	1	0	1

In (10), we can see that some properties are shared by some constructions; there is no single unique property of SVCs. Particularly, the auxiliary constructions are the same as non-idiomatic SVCs in terms of the three properties. So the auxiliary constructions can arguably be a type of SVC. However, I assume here that they are a type of complex predicate, but not SVC, since another important property of SVCs is that the component verbs can be used on its own with its lexical meaning in other sentences, but the final verb of auxiliary construction is simply a dependent auxiliary verb (see Zwicky 1990, Aikhenvald 2006, Kim 2010).

Excluding auxiliary construction, the wide scope of *an* is a unique property of SVCs in the table. The combinations of the three properties (codified as 101 or 100) are also unique to SVCs, which can be used as diagnostics to test whether the sentences such as (2) are SVC or not.

2.2 Index-sharing SVCs

Before the combination of the properties of SVCs are applied to the sentence (2b), I discuss the empirical grounds of it. The sentences like (2b) and (11) below are found in the Web. Of course, the appearances in the Web do not guarantee themselves that they are grammatical. However, if the findings from the Web are associated with a survey result, then we can have a combined support for well-formedness of the sentences. In the survey I have conducted[3], most participants judged (2b) acceptable (mean: 1.73, standard deviation: 1.10), and about half the participants judged the sentences in (11a) (mean: 2.45, standard deviation: 1.21) and (11b) (mean: 2.73, standard deviation: 1.19) acceptable. But another serialization in (11c) was judged unacceptable by most participants (mean: 3.09, standard deviation: 1.22).

(11) a. ?wenswungi-ka akma-eykey ssip-**hi**-e mek-ess-eyo.
monkey-Nom demon-By chew-Pass-Comp eat-Pst-Dec
'The monkey was chewed and then eaten by the demon.'

b. ?ku-uy phi-ka akma-eykey ppal-**li**-e mek-ess-eyo.
he-Gen blood-Nom demon-By suck-Pass-Comp eat-Pst-Dec
(lit.) 'His blood was sucked and then eaten by the demon.'

c. *sasum-i akma-eykey ccic-**ki**-e mek-ess-eyo.
sasum-Nom demon-By tear-Pass-Comp eat-Pst-Dec
(int.) 'The deer was torn and then eaten by the demon.'

---

3 The 11 participants of the survey were native speakers of Korean living in Korea. The survey was to collect their acceptability judgments of stimulus sentences. For example, a participant should choose one of the four acceptability grades regarding a given sentence (instructions and stimulus sentences were given to participants in Korean):

i) wenswungi-ka akma-eykey cap-hi-e mek-ess-eyo.

1. Clearly acceptable
2. Seems acceptable
3. Seems unacceptable
4. Clearly unacceptable

Even though it was a small informal survey, the results indicate that some sentences like (2b) are fairly acceptable for some speakers.

So it seems not implausible to assume that some sentences like (2b) are grammatical.

I show now that the sentences such as (2b) are genuine SVCs which have only an argument index sharing in the system of Korean SVCs. For instance, the sentence (2b) has the unique combination of the non-idiomatic SVC properties that other non-SVCs do not have: the negation immediately preceding the first verb can scope over the whole verbal serialization, as shown in (12a),[4] the first verb cannot have a separate tense marking, as in (12b), and it seems that the delimiter *-man* 'only' can be attached to the first verb, as in (12c).

(12) a. wenswungi-ka akma-eykey **an** cap-**hi**-e mek-ess-eyo.
monkey-Nom demon-By Neg catch-Pass-Comp eat-Pst-Dec
'It is not the case that the monkey was caught and then eaten by the demon.'

b. wenswungi-ka akma-eykey cap-**hi**(***-ess**)-e mek-ess-eyo.
monkey-Nom demon-By catch-Pass-Pst-Comp eat-Pst-Dec
'The monkey was caught and then eaten by the demon.'

c. ?wenswungi-tul-i akma-eykey cap-**hi**-e-**man** mek-ess-eyo.
monkey-Plu-Nom demon-By catch-Pass-Comp-only eat-Pst-Dec
'The monkeys were only caught and then eaten by the demon.'

This combination of the properties strongly indicates that the sentence (2b) is a genuine SVC in Korean.

SVCs can have more than two verbs. We can predict that SVCs like (2) involving more than two verbs systematically have at least one argument index sharing (i.e. one semantic argument sharing). This is verified below:

(13) a. wenswungi-ka akma-eykey cap-a mek-**hi**-e
monkey-Nom demon-By catch-Comp eat-Pass-Comp
cwuk-ess-eyo.
die-Pst-Dec
'The monkey was caught and then eaten by the demon and then died.'

4 One may think that the reason why the negation cannot have narrow scope over the V1 in (12a) is that it is infelicitous for the monkey being eaten even without it being caught first. But this is not actually implausible pragmatically: we can imagine a situation where the monkey was not caught, but just found dead and then eaten by the demon.

b. wenswungi-ka akma-eykey cap-**hi**-e mek-e
monkey-Nom demon-By catch-Pass-Comp eat-Comp
cwuk-ess-eyo.
die-Pst-Dec
'The monkey was caught and then eaten by the demon and then died.'

In (13), the semantic argument (i.e. the referent of the monkey) is shared by all the three verbs. In the next section, I further discuss the requirement of the semantic argument sharing in SVCs.

3. Requirement of an index sharing in SVCs

The lexical passive SVCs lead us to posit the hypothesis that a semantic argument sharing is necessary for SVCs in Korean.

The coordination construction in (14a) and the subordination constructions in (14b) and (14c) (without e.g. a pronoun and its antecedent) do not have an argument index sharing. So generally, they do not necessarily have a semantic argument sharing.

(14) a. Jenny-ka mwul-ul sa-ss-ko,
Jenny-Nom water-Acc buy-Pst-Conj
Tom-i cusu-lul sa-ss-ta.
Tom-Nom juice-Acc buy-Pst-Dec
'Jenny bought the water, and Tom bought the juice.'

b. Mary-ka Tom-i ttokttokhata-ko sayngkakhay-ss-ta.
Mary-Nom Tom-Nom smart-Comp think-Pst-Dec
'Mary thought that Tom was smart.'

c. Mary-ka tolawa-se Bill-i kippe-ss-ta.
Mary-Nom return.Comp-since Bill-Nom happy-Pst-Dec
'Since Mary returned, Bill was happy.'

So it seems to hold that if a construction doesn't have an argument index sharing, then it is not an SVC.

However, some constructions that have an argument index sharing like resultative, auxiliary construction, and typical subject or object control

constructions don't belong to SVCs in Korean: SVCs have the unique combination of the properties that distinguishes them from other constructions like resultative, auxiliary construction (as already discussed above), and control. For example, the negative marker *an* cannot have wide scope in control constructions as follows:

(15) a. ku-ka **an** o-n-ta-ko yaksokhay-ss-ta.
he-Nom Neg come-Pres-Dec-Comp promise-Pst-Dec
'He promised not to come.'

b. ku-ka Mary-lul **an** o-tolok kangyohay-ss-ta.
he-Nom Mary-Acc Neg come-Tolok force-Pst-Dec
'He forced Mary not to come.'

No wide scope of the negation in controls should be enough to falsify the classification of controls as a kind of SVCs. Hence it follows that a semantic argument sharing does not entail SVCs.

In sum, no argument index sharing seems to entail non-SVCs in Korean, which supports the necessity of an argument index sharing (not the subject, an object or an internal argument) in SVCs.

4. Previous approaches

In this section I show that some previous analyses are not appropriate for an account of the lexical passive SVCs in question.

First, it may be argued that two different underlying sentences are combined to derive an SVC (e.g. Stewart 1963, Bamgose 1974). If this is true, in order to generate the lexical passive SVCs, *cap-ass-ta* in (16a) must be changed to *cap-a* and *cap-hi-ess-ta* in (17a) to *cap-hi-e* through some kind of complex derivational operations (i.e. replacing *-ass-ta* with *-a* or *-ess-ta* with *-e* in syntax). The more serious problem of this kind of analysis is that an ill-formed sentence like (16a) or (17b) should be licensed first in order to generate the relevant lexical passive SVCs. In (16a) and (17b), the NP *akma-eykey* 'demon-By' is the complement which is not required by the active forms of the verbs. In addition, the subject *wenswungi-ka* 'monkey-Nom' must be the agent in (16a) and (17b), but in the lexical passive SVCs, it is the patient.

(16) a. wenswungi-ka (*akma-eykey) cap-ass-ta.
monkey-Nom demon-By catch-Pst-Dec
'The monkey caught something.'

b. wenswungi-ka akma-eykey mek-**hi**-ess-ta.
monkey-Nom demon-By catch-Pass-Pst-Dec
'The monkey was eaten by the demon.'

(17) a. wenswungi-ka akma-eykey cap-**hi**-ess-ta.
monkey-Nom demon-By catch-Pass-Pst-Dec
'The monkey was caught by the demon.'

b. wenswungi-ka (*akma-eykey) mek-ess-ta.
monkey-Nom demon-By eat-Pst-Dec
'The monkey ate something.'

In order to circumvent these problems, the analysis should invent a much more complex derivational system.

Baker (1989) argues that SVCs require the object sharing, and the component verbs co-head the shared object. However, as already shown in the SVC in (1b) and the lexical passive SVCs in (2), there is no shared object, and thus the object sharing is not necessary in Korean SVCs. Similarly, Collins (1997) argues that the internal argument sharing is the requirement of SVCs in Ewe, and V2 combines with an empty category coindexed with the explicit object of V1. However, in SVCs like (2a), *akma-eykey* 'demon-By' is not the object of the V1, and also it is not immediately clear how the passive V2 assigns its CASE values to the subject and complement. If we assume that V2 somehow assigns its CASE values to the subject and complement to account for (2a), then we also need to explain why in (2b) V2 does not assign its CASE values to the subject and complement.

Choi (2003) assumes that the index-sharing SVCs like (2b) are ill-formed. However, it seems plausible to consider them genuine SVCs (at least for some speakers), as illustrated above. According to Choi (2003), the subject in [Spec *v*1] and object in [Spec V1] are moved to [Spec *v*2] and [Spec V2], respectively. Then this analysis seems to need to explain how in (2a) the subject and object of V1 should be moved to the complement and subject of V2, respectively, and how the CASE values of V1 are changed to the CASE values of V2. It should also account for how in (2b) V2 may not assign its CASE values to the moved arguments unlike V2 in (2a).

Sohn and Ko (2010) categorize Korean SVCs into two types: H(igh)-SVC where passive *v* head is merged to a verbal stem before it is serialized with another verb, and L(ow)-SVC where the verbal serialization occurs prior to the merger of the *v* head. Then they argue for Distributed Morphology. For instance, the lexical passive SVCs like (2a) can be analyzed as L(ow)-SVCs involving the passive form of the verbal serialization (i.e. [*cap-a mek*]-***hi***).

However, the data like (2b) seem to be a considerable theoretical problem for their analysis, since the passive affix *hi* is inside the V1 (i.e. [*cap-**hi**-e mek-ess-ta*]), which makes a bit more difficult to derive in syntax the serialization with the appropriate meaning. In addition, they didn't talk about the CASE assignments in SVCs.

Although I do not prove that a new derivational analysis accommodating the lexical passive SVCs like (2b) is impossible, I believe that an argument composition analysis (e.g. Andrews 1997, Chung 1998, Chung and Kim 2008, Kim 2010, Lee 2011) is able to account for the phenomenon of the lexical passive SVCs with ease. We can simply add a new SVC type of lexical passive SVCs requiring that the arguments of active verb be coindexed with those of passive verb and only the subject and complement of the passive verb be passed up, respectively, to the subject and complement of the resulting combination in a similar manner of controls.

5. HPSG formalization

I present a formal analysis of the lexical passive SVCs, focusing on those SVCs that have only two component verbs, in Head-Driven Phrase Structure Grammar (Pollard and Sag 1994, Sag *et al*. 2003).

5.1 Lexical rule and lexical items

A VP- or S-complement analysis violates the locality constraint of CASE assignment of, say, the passive V2 to its arguments:

(18) a. wenswungi-ka [$_{VP}$ akma-eykey cap-a] mek-**hi**-ess-ta.
monkey-Nom demon-By catch-Comp eat-Pass-Pst-Dec
'The monkey was caught and then eaten by the demon.'

b. [$_{S}$ wenswungi-ka akma-eykey cap-a] mek-**hi**-ess-ta.
monkey-Nom demon-By catch-Comp eat-Pass-Pst-Dec
'The monkey was caught and then eaten by the demon.'

Rather, I adopt and adapt the argument composition analysis (e.g. Andrews 1997, Chung 1998, Chung and Kim 2008, Kim 2010, Lee 2011) which captures the generalizations and idiosyncrasies via the type hierarchy of SVCs. Passive lexemes with active form (e.g. *mek*-1 'eat' vs. *mek*-2 'be eaten') may be posited in the lexicon or generated by a lexical rule, but it seems very unintuitive that the active form of a verb has a passive meaning and this also appears to lack independent motivation in Korean. Note also that the passive meaning of the active form of a verb is created in the context

of the SVCs, rather than in isolation. So I assume that the passive lexeme (19b) is licensed from (19a) by the general Passive Lexical Rule in (20) adopted from Sag *et al*. (2003) and Kim (2004).

(19) a. *cap-* 'catch':

$$\begin{bmatrix} \text{PHON } cap- \\ \text{PASSIVE } - \\ \\ \text{ARG-ST } \langle \text{NP}_i, \text{ NP}_j \rangle \\ \\ \text{RELS} < \begin{bmatrix} catch_rel \\ \text{ARG1 } i \\ \text{ARG2 } j \end{bmatrix} > \end{bmatrix}$$

b. *cap-hi-* 'caught':

$$\begin{bmatrix} \text{PHON } cap-hi- \\ \text{PASSIVE } + \\ \\ \text{ARG-ST } \langle \text{NP}_j, \text{ NP}_i \rangle \\ \\ \text{RELS} < \begin{bmatrix} catch_rel \\ \text{ARG1 } i \\ \text{ARG2 } j \end{bmatrix} > \end{bmatrix}$$

(20) Passive Lexical Rule:

$$\begin{bmatrix} \text{INPUT} \left\langle \boxed{1}, \begin{bmatrix} v-tr \\ \text{ARG-ST} \langle \text{NP}_i, \text{ NP}_j, \ldots \rangle \\ \text{CONT } \boxed{2} \end{bmatrix} \right\rangle \\ \\ \text{OUTPUT} \left\langle \text{F}_{PASS}\left(\boxed{1}\right), \begin{bmatrix} v-pass \\ \text{PASSIVE } + \\ \text{ARG-ST} \langle \text{NP}_j, \text{ NP}_i, \ldots \rangle \\ \text{CONT } \boxed{2} \end{bmatrix} \right\rangle \end{bmatrix}$$

When the Passive Lexical Rule in (20) is applied to the verb lexeme *cap-* 'catch' in (19a) as an input, then the verb lexeme *cap-hi-* 'caught' attached with the passive affix *hi* and with the semantics unchanged is generated as the output. The arguments arrangement of *cap-hi-* 'caught' is different from that of *cap-* 'catch'; the NP_j, which should be the patient of the verb, comes now first (leftmost) in the ARG(UMENT)-ST(RUCTURE) list of *cap-hi-* 'caught'.[5]

Some lexemes relevant to the SVCs under discussion are presented below. (21b) is licensed from (21a) by the Passive Lexical Rule. The verb lexemes in (22) are listed in the lexicon.

(21) a. *mek-* 'eat'

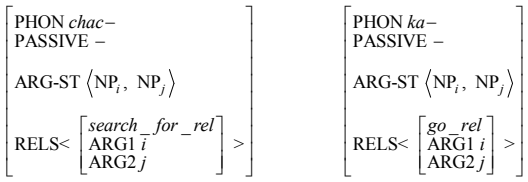
$$\begin{bmatrix} \text{PHON } mek- \\ \text{PASSIVE } - \\ \\ \text{ARG-ST } \langle \text{NP}_i, \text{ NP}_j \rangle \\ \\ \text{RELS} < \begin{bmatrix} eat_rel \\ \text{ARG1 } i \\ \text{ARG2 } j \end{bmatrix} > \end{bmatrix}$$

b. *mek-hi-* 'eaten':

$$\begin{bmatrix} \text{PHON } mek-hi- \\ \text{PASSIVE } + \\ \\ \text{ARG-ST } \langle \text{NP}_j, \text{ NP}_i \rangle \\ \\ \text{RELS} < \begin{bmatrix} eat_rel \\ \text{ARG1 } i \\ \text{ARG2 } j \end{bmatrix} > \end{bmatrix}$$

5 If we want to make the arguments arrangement unchanged in a passive lexeme since it may sound odd that the patient comes first in the ARG-ST list, then we may need a different type of Argument Realization Principle (see Sag *et al*. 2003) that can apply only to passive words. Rather, I choose not to make multiple types of Argument Realization Principle.

(22) a. *chac-* 'search for': b. *ka-* 'go':

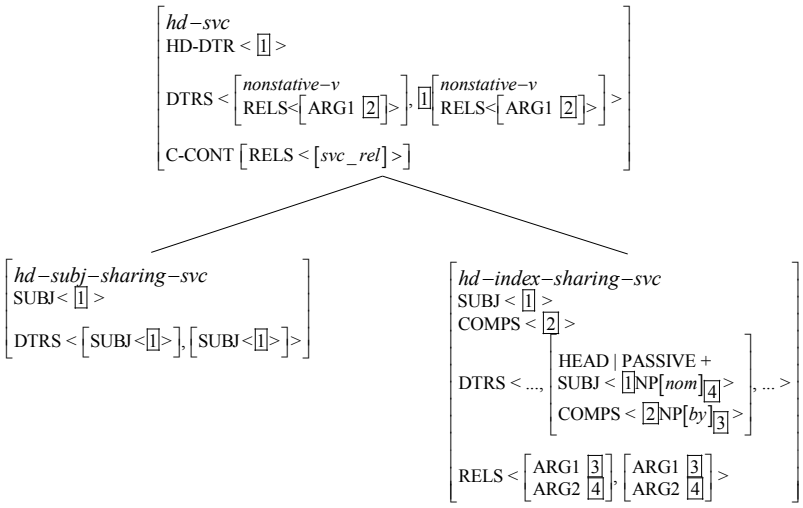


I propose below the relevant constructional rules of SVCs stated in a type-hierarchy for combinations of the given lexical items.

5.2 Type hierarchy of SVCs

In the type hierarchy of SVCs described in (23), I state the generalization of a semantic argument sharing (the structure-shared index of ARG1) as constraint on the type *hd-svc* with the final verb as the morphosyntactic head. I claim this type has two subtypes, *hd-subj-sharing-svc* and *hd-index-sharing-svc*.

(23) *hd-svc*:

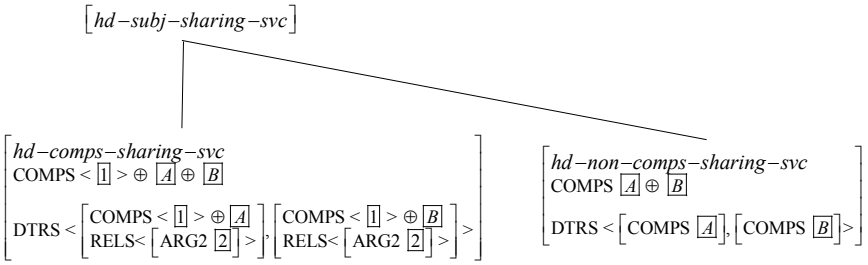


Since subject-sharing SVCs basically require the subject sharing, in the first subtype *hd-subj-sharing-svc* the SUBJ values (tagged 1) of the two component verbs are structure-shared, and they are then identified with the SUBJ value of the resulting combination. In the other subtype *hd-index-sharing-svc*, one component verb is passive (marked with PASSIVE +) and one more semantic argument (ARG2) is structure-shared in addition to the structure-shared argument (ARG1) inherited from its supertype *hd-svc*. Besides, the SUBJ value and COMPS value of the passive verb are

indentified with the SUBJ value and COMPS value of the resulting construction, respectively. Note that the index of the SUBJ value is the same as that of ARG2 in both verbs, and the index of the COMPS value is the same as that of ARG1 of both verbs. These co-indexations have the effect that the other verb contributes its semantics to the verbal serialization.

The type *hd-subj-sharing-svc* in turn has two subtypes, *hd-comps-sharing-svc* and *hd-non-comps-sharing-svc*, as shown in (24) (see a similar analysis in Müller and Lipenkova 2009).

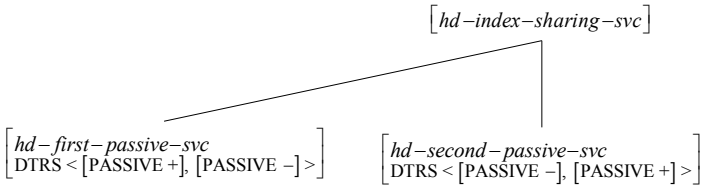
(24) *hd-subject-sharing-svc*:



The two subtypes in (24) has the structure-shared SUBJ value inherited from their supertype *hd-subj-sharing-svc*. The first subtype *hd-comps-sharing-svc* has one complement that is structure-shared (tagged 1). So in addition to this structure-shared COMPS value, the unshared COMPS values (boxed A and B) are added to the COMPS list of the resulting combination via the list append operation. However, the other subtype *hd-non-comps-sharing-svc* does not have a structure-shared complement, and so it is the unshared COMPS values that compose the COMPS list of the combination.

The type *hd-index-sharing-svc*, which is the sister of *hd-subj-sharing-svc*, also has two subtypes, *hd-first-passive-svc* whose first verb is passive and *hd-second-passive-svc* whose second verb is passive:

(25) *hd-index-sharing-svc:*



The type *hd-index-sharing-svc* requires the index sharing of ARG2 (the structure-shared 4) as declared on it in (23); and thus in both *hd-first-passive-svc* and *hd-second-passive-svc*, the unrealized SUBJ value and COMPS value of the other active verb (marked with PASSIVE –) are coindexed with the

realized COMPS value and SUBJ value of the passive verb, respectively. In other words, the active form of the other verb contributes its semantics to the verbal serialization via the co-indexations.

Summarizing, I added the new type of *hd-index-sharing-svc* and its two subtypes to the system of Korean SVCs employing the previous mechanism of argument composition of SVCs.

5.3 Analyses of verbal serializations

Equipped with the lexical items and combination rules, we can derive the verbal serializations under discussion in a straightforward way. First, the active sentence in (1a), *akma-ka wenswungi-lul* ***cap****-a* ***mek****-ess-ta.* 'The demon caught the monkey$_j$ and then ate it$_j$,' is a typical form of SVC, whose verbs share both the subject and object. This verbal serialization is licensed by the construction rule *hd-comps-sharing-svc*:

(26) [***cap****-a* ***mek****-ess-ta*] in (1a)

$$\left[\begin{array}{l} hd\text{-}comps\text{-}sharing\text{-}svc \\ \text{SUBJ} < \boxed{2} > \\ \text{COMPS} < \boxed{3} > \\ \text{HD-DTR} < \boxed{1} > \\ \text{DTRS} < \left[\begin{array}{l} \text{PHON } cap\text{-}a \\ \text{PASSIVE} - \\ \text{SUBJ} < \boxed{2}\text{NP}[nom]_i > \\ \text{COMPS} < \boxed{3}\text{NP}[acc]_j > \\ \text{RELS} < \left[\begin{array}{l} catch_rel \\ \text{ARG1 } i \\ \text{ARG2 } j \end{array}\right] > \end{array}\right], \boxed{1}\left[\begin{array}{l} \text{PHON } mek\text{-}ess\text{-}ta \\ \text{PASSIVE} - \\ \text{SUBJ} < \boxed{2} > \\ \text{COMPS} < \boxed{3} > \\ \text{RELS} < \left[\begin{array}{l} eat_rel \\ \text{ARG1 } i \\ \text{ARG2 } j \end{array}\right] > \end{array}\right] > \\ \text{C-CONT} \left[\text{RELS} < [svc_rel] >\right] \end{array}\right]$$

In the active SVC (1b), *Jane-i hakkyo-ey Tom-ul* ***chac****-a* ***ka****-ss-ta*. 'Jane went to the school searching for Tom,' only the subject is shared by the verbs. So it is an instance of the type *hd-non-comps-sharing-svc*:

(27) [***chac****-a* ***ka****-ss-ta*] in (1b)

$$\left[\begin{array}{l} hd\text{-}non\text{-}comps\text{-}sharing\text{-}svc \\ \text{SUBJ} < \boxed{2} > \\ \text{COMPS} < \boxed{3}, \boxed{4} > \\ \text{HD-DTR} < \boxed{1} > \\ \text{DTRS} < \left[\begin{array}{l} \text{PHON } chac\text{-}a \\ \text{PASSIVE} - \\ \text{SUBJ} < \boxed{2}\text{NP}[nom]_i > \\ \text{COMPS} < \boxed{3}\text{NP}[acc]_j > \\ \text{RELS} < \left[\begin{array}{l} search_for_rel \\ \text{ARG1 } i \\ \text{ARG2 } j \end{array}\right] > \end{array}\right], \boxed{1}\left[\begin{array}{l} \text{PHON } ka\text{-}ss\text{-}ta \\ \text{PASSIVE} - \\ \text{SUBJ} < \boxed{2} > \\ \text{COMPS} < \boxed{4}\text{NP}[to]_k > \\ \text{RELS} < \left[\begin{array}{l} go_rel \\ \text{ARG1 } i \\ \text{ARG2 } k \end{array}\right] > \end{array}\right] > \\ \text{C-CONT} \left[\text{RELS} < [svc_rel] >\right] \end{array}\right]$$

In the lexical passive SVC (2a), *wenswungi-ka akma-eykey* ***cap-a mek-hi-****ess-ta*. 'The monkey was caught and then eaten by the demon,' only the second passive verb takes the NPs as its subject and complement, which is an instantiation of *hd-second-passive-svc*:

(28) [***cap-a mek-hi-****ess-ta*] in (2a)

$$\left[\begin{array}{l} hd\text{-}second\text{-}passive\text{-}svc \\ \text{SUBJ} < \boxed{2} > \\ \text{COMPS} < \boxed{3} > \\ \text{HD-DTR} < \boxed{1} > \\ \text{DTRS} < \left[\begin{array}{l} \text{PHON } cap\text{-}a \\ \text{PASSIVE} - \\ \text{SUBJ} < \text{NP}[nom]_i > \\ \text{COMPS} < \text{NP}[acc]_j > \\ \text{RELS} < \left[\begin{array}{l} catch_rel \\ \text{ARG1 } i \\ \text{ARG2 } j \end{array}\right] > \end{array}\right], \boxed{1} \left[\begin{array}{l} \text{PHON } mek\text{-}hi\text{-}ess\text{-}ta \\ \text{PASSIVE} + \\ \text{SUBJ} < \boxed{2}\text{NP}[nom]_j > \\ \text{COMPS} < \boxed{3}\text{NP}[by]_i > \\ \text{RELS} < \left[\begin{array}{l} eat_rel \\ \text{ARG1 } i \\ \text{ARG2 } j \end{array}\right] > \end{array}\right] > \\ \text{C-CONT} \left[\text{RELS} < [svc_rel] >\right] \end{array}\right]$$

Now the lexical passive SVC in (2b), *wenswungi-ka akma-eykey* ***cap-hi-e mek-****ess-eyo*. 'The monkey was caught and then eaten by the demon,' is analyzed as an example of the type *hd-first-passive-svc*:

(29) [***cap-hi-e mek-****ess-eyo*] in (2b)

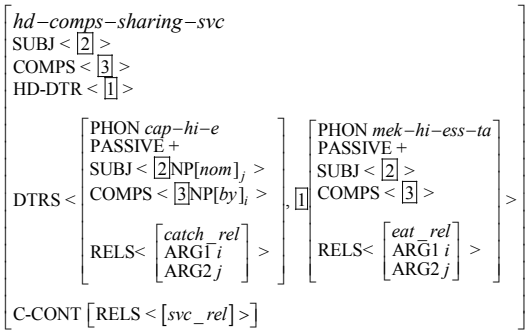
$$\left[\begin{array}{l} hd\text{-}first\text{-}passive\text{-}svc \\ \text{SUBJ} < \boxed{2} > \\ \text{COMPS} < \boxed{3} > \\ \text{HD-DTR} < \boxed{1} > \\ \text{DTRS} < \left[\begin{array}{l} \text{PHON } cap\text{-}hi\text{-}e \\ \text{PASSIVE} + \\ \text{SUBJ} < \boxed{2}\text{NP}[nom]_j > \\ \text{COMPS} < \boxed{3}\text{NP}[by]_i > \\ \text{RELS} < \left[\begin{array}{l} catch_rel \\ \text{ARG1 } i \\ \text{ARG2 } j \end{array}\right] > \end{array}\right], \boxed{1} \left[\begin{array}{l} \text{PHON } mek\text{-}ess\text{-}eyo \\ \text{PASSIVE} - \\ \text{SUBJ} < \text{NP}[nom]_i > \\ \text{COMPS} < \text{NP}[acc]_j > \\ \text{RELS} < \left[\begin{array}{l} eat_rel \\ \text{ARG1 } i \\ \text{ARG2 } j \end{array}\right] > \end{array}\right] > \\ \text{C-CONT} \left[\text{RELS} < [svc_rel] >\right] \end{array}\right]$$

In (29), it is the first passive verb that takes the NPs as its subject and complement. The second active verb contributes its semantics to the SVC through the argument index sharings.

If the two verbs of an SVC are all passive, as exemplified in (30), it is licensed by the type *hd-comps-sharing-svc*, as in (31), just like (1a) is.

(30) sasum-i saca-eykey cap-**hi**-e mek-**hi**-ess-ta.
deer-Nom lion-By catch-Pass-Comp eat-Pass-Pst-Dec
'The deer was caught and then eaten by the lion.'

(31) [*cap-**hi**-e mek-**hi**-ess-ta*]



The current system of Korean SVCs proposed in this paper can interact with pragmatic theories to restrict what specific verb combinations can appear in SVCs (see cultural factors noted in Durie 1997, Kroeger 2004). This interaction may be related to why some lexical passive SVCs like (2b) are quite acceptable, but some other lexical passive SVCs such as (11c) are highly unacceptable.

6. Conclusion

I presented an argument composition analysis of the lexical passive SVCs in Korean by adding the new construction type of index-sharing SVCs, *hd-index-sharing-svc*, and its two subtypes, *hd-first-passive-svc* and *hd-second-passive-svc*, to the grammar of Korean SVCs. I also showed that in a Korean SVC, V1 and V2 must share a semantic argument (i.e. an argument index) rather than the subject, an object, or an internal argument.

I believe the conclusion has promise, and the prediction (the existence of such an index-sharing SVCs) would be cross-linguistically valid. I leave to future research examining this prediction and a formalization of the interaction between the current SVC system and pragmatic theories.

References

Aikhenvald, Alexandra Y. 2006. Serial Verb Constructions in Typological Perspective. In Alexandra Y. Aikhenvald and R. M. W. Dixon (ed.), *Serial Verb Constructions: A Cross-Linguistic Typology*, 338-350. Oxford: Oxford University Press.

Andrews, Avery D. 1997. Complex Predicates and Nuclear Serial Verbs. In *Proceedings of the LFG97 Conference.* Stanford: CSLI Publications.

Baker, Mark C. 1989. Object Sharing and Projection in Serial Verb Constructions. *Linguistic Inquiry* 20.4, 513-553.

Bamgbose, Ayo. 1974. On serial verbs and verbal status. *Journal of West African Languages* 9, 17-48.

Choi, Seongsook. 2003. Serial verbs and the empty category. In *Proceedings of the Workshop on Multi-Verb Constructions*. Trondheim: Norwegian University of Science and Technology.
Chung, Chan. 1998. Argument Composition and Long-Distance Scrambling in Korean: An Extension of the Complex Predicate Analysis. In Hinriches *et al*. (eds.), *Complex Predicates in Nonderivational Syntax*, 159-220. New York: Academic Press.
Chung, Chan and Jong-Bok Kim. 2008. Korean Serial Verb Constructions: A Construction-Based Approach. *Studies in Generative Grammar* 18.4, 559-582.
Collins, Chris. 1997. Argument Sharing in Serial Verb Constructions. *Linguistic Inquiry* 28.3, 461-497.
Durie, Mark. 1997. Grammatical structures in verb serialization. Grammatical Structures in Verb Serialization. In Alex Alsina, Joan Bresnan, and Peter Sells (ed.), *Complex Predicates*, 289-354. Stanford: CSLI Publications.
Foley, William A. and Mike Olson. 1985. Clausehood and verb serialization. In Johanna Nichols and Anthony C. Woodbury (ed.), *Grammar Inside and Outside the Clause: Some Approaches to Theory from the Field*, 17-60. Cambridge: Cambridge University Press.
Kim, Jong-Bok. 2004. *Korean Phrase Structure Grammar*. [in Korean] Seoul: Hankook Munhwasa.
Kim, Jong-Bok. 2010. Argument Composition in Korean Serial Verb Constructions. *Studies in Modern Grammar* 61, 1-24.
Kroeger, Paul R. 2004. *Analyzing Syntax: A Lexical-functional Approach*. Cambridge: Cambridge University Press.
Lee, Juwon. 2011. Change of State Verb and Syntax of Serial Verb Constructions in Korean: An HPSG Account. In *Proceedings of the Poster Session of the West Coast Conference on Formal Linguistics*, *Coyote Papers: Working Papers in Linguistics* 20, 57-65.
Lee, Sookhee. 1992. *The Syntax and Semantics of Serial Verb Constructions*. Ph.D. Dissertation. University of Washington.
Müller, Stefan and Janna Lipenkova. 2009. Serial Verb Constructions in Chinese: An HPSG Account. In *Proceedings of the 16th International Conference on Head-Driven Phrase Structure Grammar*, University of Göttingen, Germany, 234–254. Stanford: CSLI Publications.
Pollard, Carl and Ivan A. Sag. 1994. *Head-driven Phrase Structure Grammar*. Stanford: CSLI Publications and Chicago: The University of Chicago Press.
Sag, Ivan A., Thomas Washow, and Emily M. Bender. 2003. *Syntactic Theory: A Formal Introduction, 2nd edition*. Stanford: CSLI Publications.
Sebba, Mark. 1987. *The Syntax of Serial Verbs*. Amsterdam: John Benjamins.
Sohn, Daeyoung and Ko Heejeong. 2010. The Role of Merger and Typology

of *v* Heads in Serialization. *Japanese/Korean Linguistics* 19. Stanford: CLSI Publications.
Stewart, John M. 1963. Some restrictions on objects in Twi. *Journal of African Languages* 2.2, 145-149.
Thepkanjana, Kingkarn and Satoshi Uehara. 2009. Resultative constructions with ‘‘implied-result’’ and ‘‘entailed-result’’ verbs in Thai and English: a contrastive study. *Linguistics* 47.3, 589-618.
Van Valin, Robert D., Jr. 2005. *Exploring the Syntax-Semantics Interface*. Cambridge: Cambridge University Press.
Zwicky, Arnold M. 1990. What are we talking about when we talk about serial verb constructions? In Brian D. Joseph and Arnold M. Zwicky (eds.), *When Verbs Collide: Papers from the 1990 Ohio State Mini-Conference on Serial verb constructions*. *The Ohio State University*, *Working Papers in Linguistics* 39, 1-13.

Complement raising, extraction and adposition stranding in Dutch

Frank Van Eynde
University of Leuven

Liesbeth Augustinus
University of Leuven

Proceedings of the 21st International Conference on
Head-Driven Phrase Structure Grammar

University at Buffalo

Stefan Müller (Editor)

2014

Stanford, CA: CSLI Publications

pages 156–175

Van Eynde, Frank & Liesbeth Augustinus. 2014. Complement raising, extraction and adposition stranding in Dutch. In Stefan Müller (ed.), *Proceedings of the 21st International Conference on Head-Driven Phrase Structure Grammar, University at Buffalo*, 156–175. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2014.9.
CC-BY 4.0

Abstract

In Dutch, adpositions can be stranded, typically if their complement is an R-pronoun. The complement usually appears in the left part of the Mittelfeld or in the Vorfeld. In HPSG this is canonically modeled in terms of extraction, making use of nonlocal devices such as SLASH and BIND. This paper argues that the extraction analysis is indeed appropriate for cases in which the complement is realised in the Vorfeld, but proposes an alternative for the cases in which the complement is realised in the Mittelfeld. The new treatment is based on argument inheritance, as complement raising in the Mittelfeld involves a middle distance dependency rather than a long distance dependency.

1 Introduction

In Dutch, adpositions canonically precede their complement.

(1) Ze zegt dat ze soms nog [aan hem/Hans] denkt.
she says that she sometimes still [of him/Hans] thinks
'She says that she still thinks of him/Hans from time to time.'

(2) * Ze zegt dat ze soms nog [hem/Hans aan] denkt.
* she says that she sometimes still [him/Hans of] thinks

However, if the complement is a demonstrative pronoun, such as *dit* 'this' or *dat* 'that', it takes another form, the so-called R-form, and precedes the adposition.[1]

(3) * Ze zegt dat ze soms nog [aan dat/dit] denkt.
* she says that she sometimes still [of that/this] thinks

(4) Ze zegt dat ze soms nog [daar/hier aan] denkt.
she says that she sometimes still [that+R/this+R of] thinks
'She says that she still thinks of that/this from time to time.'

The same holds for the impersonal *het* 'it' and the interrogative/relative *wat* 'what', which alternate with *er* and *waar* respectively. The alternation also applies to the quantifying *iets* 'something', *niets* 'nothing' and *alles* 'everything', but for these pronouns it is optional: (5) and (6) are equally well-formed.[2]

(5) Ze zegt dat ze soms gewoon [aan niets] denkt.
she says that she sometimes simply [of nothing] thinks
'She says that she simply thinks of nothing from time to time.'

[†] We thank the audience of the HPSG 2014 conference (Buffalo, August 28-29) for their comments. The research presented in this paper is part of a project on complement raising and cluster formation in Dutch, sponsored by FWO Vlaanderen (2011-2015, G.0.559.11.N.10).

[1] The sequence of the R-pronoun and the adposition is often treated as an orthographic unit, as in *daaraan* and *hieraan*.

[2] The sequence of *(n)ergens/overal* and an adposition is not treated as a single unit in the orthography.

(6) Ze zegt dat ze soms gewoon [nergens aan] denkt.
she says that she sometimes simply [nothing+R of] thinks
'She says that she simply thinks of nothing from time to time.'

Table 1 provides a survey of the pronouns which show the [–/+ R] alternation. What they have in common is that they all denote a thing rather than a person: They are [–HUMAN], see Van Riemsdijk (1978, 37–40).

Pronoun	[-R]	[+R]	
Impersonal	*het*	*er*	'it'
Demonstrative	*dat*	*daar, d'r*	'that'
	dit	*hier*	'this'
Interrogative/Relative	*wat*	*waar*	'what'
Quantifying	*iets*	*ergens*	'something'
	niets	*nergens*	'nothing'
	alles	*overal*	'everything'

Table 1: The Dutch pronouns with an R-form

A peculiar property of the R-pronouns is that they tend to be realized out of the PP: They typically end up in the left part of the Mittelfeld, preceding the VP adjuncts, as in (7–8), or in the Vorfeld, as in (9–10).

(7) Ze zegt dat ze *daar* soms nog [__ aan] denkt.
she says that she that+R sometimes still [__ of] thinks
'She says that she still thinks of it from time to time.'

(8) We hebben *er* toen een lied [__ over] gezongen.
we have it+R then a song [__ about] sung
'We have sung a song about it.'

(9) *Waar* denk je dat ze [__ op] wachten?
what+R think you that they [__ for] wait
'What do you think they are waiting for?'

(10) *Hier* kunnen we echt niet [__ op] wachten.
this+R can we really not [__ for] wait
'This we really cannot wait for.'

The result of this non-local realization is that the adposition is left alone: It is stranded in the right part of the Mittelfeld. The phenomenon has been studied extensively. Descriptive surveys are provided in Haeseryn et al. (1997) and Broekhuis (2013), transformational treatments in Van Riemsdijk (1978) and Bennis (1986),

and HPSG treatments in Rentier (1993) and Bouma (2000).[3] Both of the latter treat the phenomenon in terms of extraction, employing non-local devices such as SLASH and BIND.

This paper endorses the extraction treatment for sentences in which the R-pronoun ends up in the Vorfeld, as in (9–10), but proposes an alternative for the sentences in which the R-pronoun occurs in the left part of the Mittelfeld, as in (7–8). Since the latter is a middle distance (or bounded) dependency rather than a long distance (or unbounded) dependency, we argue that its proper treatment requires an analysis that is based on argument inheritance, rather than on non-local devices. Section 2 presents the analysis, section 3 compares it with the uniform extraction analysis, especially with Bouma's version, and section 4 draws some conclusions.

2 The analysis

In HPSG middle-distance dependencies are typically dealt with in terms of argument inheritance, also known as generalized raising. It was first proposed in Hinrichs & Nakazawa (1989, 1994) for a treatment of the German verb clusters, and it was adopted and adapted by various authors to deal with similar phenomena in other languages, such as the Dutch verb clusters in Bouma & van Noord (1998) and clitic climbing in French and Italian, see Abeillé et al. (1998) and Monachesi (1998). We will adopt it here to deal with the adposition stranding in (7–8), albeit with a twist, in the sense that we adopt the treatment of argument inheritance that is proposed in Van Eynde & Augustinus (2013). A characteristic property of that treatment is that it differentiates complement raising from subject raising.

We first show how this treatment deals with scrambling in the Mittelfeld (section 2.1), and then apply it to the phenomenon of adposition stranding (section 2.2). Next, we discuss a constraint on adposition stranding (section 2.3) and show that it extends to scrambling in general (section 2.4). Finally, we discuss a second constraint on adposition stranding and argue why complement raising must be differentiated from complement extraction (section 2.5).

2.1 Scrambling as a result of complement raising

A typical instance of scrambling in the Mittelfeld is given in the bracketed subordinate clause of (11).

(11) Het schijnt [dat ze hem nog niet had ontmoet].
it seems [that she him still not had met]
'It seems that she had not met him yet.'

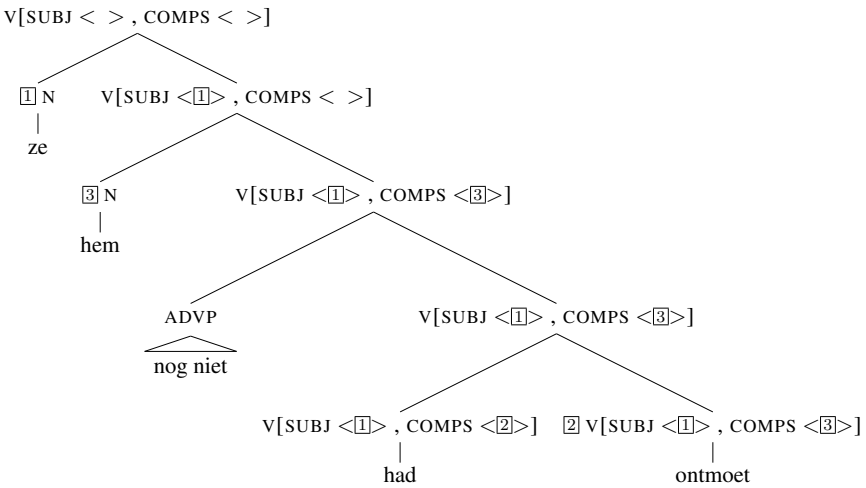
[3]The phenomenon also occurs in German, albeit on a smaller scale. See Fleischer (2002) and Kunkel-Razum & Münzberg (2006) for a descriptive overview, and Müller (1995) for an HPSG analysis.

The main verb *ontmoet* 'met' is separated from its arguments *ze* 'she' and *hem* 'him' by the auxiliary of the perfect *had* and the VP-adjunct *nog niet* 'not yet'.

To link the verb to its arguments, the generalized raising treatment assumes that the auxiliary inherits the unfulfilled expectations of its participial complement. As applied to the example, *had* inherits the SUBJ list of *onmoet* 'met' and adds the latter's COMPS list to its own COMPS list.

In the treatment of Van Eynde & Augustinus (2013), the auxiliary inherits the SUBJ list of *ontmoet* 'met', but not its COMPS list. Instead, the non-realized COMPS requirement of the participle is propagated directly from the participle to the mother node, as in (12).[4]

(12)



The auxiliary selects a participial VP complement (2) and inherits its unrealized SUBJ requirement (1), but not its unrealized COMPS requirement (3). The latter is propagated directly to the mother.

To model this we employ a lexical constraint for subject raising and a phrasal constraint for complement raising.

(13) *s-rsg-lx* $\Rightarrow$ [ARG-ST $<\boxed{1}, [\text{SUBJ} <\boxed{1}>]>$]

(14) *hd-ph* $\Rightarrow$
$$\begin{bmatrix} \text{SS} \mid \text{LOC} \mid \text{CAT} \mid \text{COMPS} & \boxed{B} \oplus \boxed{A} \\ \text{HEAD-DTR} \mid \text{SS} \mid \text{LOC} \mid \text{CAT} \mid \text{COMPS} & \boxed{A} \\ \text{NONHD-DTR} \mid \text{SS} \mid \text{LOC} \mid \text{CAT} \mid \text{COMPS} & \boxed{B} \end{bmatrix}$$

The lexical constraint is identical to the one for English, see Ginzburg & Sag (2000, 22) and Sag et al. (2003, 367).[5] The phrasal constraint subsumes all headed phrases.[6]

[4]The SUBJ and COMPS lists of *ze* 'she', *hem* 'him' and *nog niet* 'not yet' are all empty.

[5]There is a similar lexical constraint for the object raising lexemes, such as *expect* and *make*.

[6]In non-headed phrases, such as coordinate structures, the COMPS list of the mother is identified with the COMPS lists of each of the conjunct daughters separately. In *he buys and sells cars*, for instance, the coordinate phrase *buys and sells* has the same COMPS list as its conjunct daughters, *buys* and *sells*.

Cancellation of elements from the COMPS list is modeled in the definition of the phrases of type *head-complement*.[7]

(15) *hd-comp-ph* $\Rightarrow$ $\begin{bmatrix} \text{SS} \mid \text{LOC} \mid \text{CAT} \mid \text{COMPS} & \boxed{A} \\ \text{HEAD-DTR} \mid \text{SS} \mid \text{LOC} \mid \text{CAT} \mid \text{COMPS} & \boxed{A} \oplus \left\langle \boxed{1} \right\rangle \\ \text{NONHD-DTR} \mid \text{SS} & \boxed{1} \end{bmatrix}$

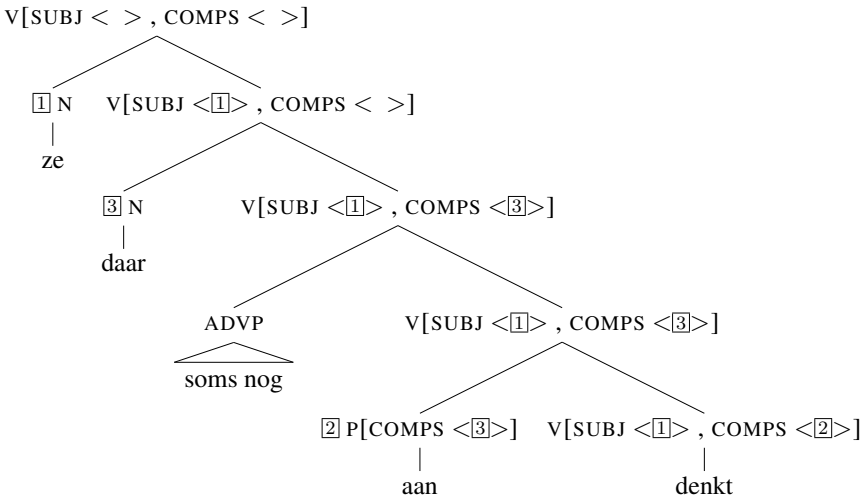
Since *head-complement-phrase* is a subtype of *headed-phrase*, it follows that the COMPS list can shrink and expand at the same time. The combination of *had* with *ontmoet* in (12), for instance, involves the cancelation of the requirement for a participial complement (2) and the addition of the unrealized requirement for an accusative nominal (3).

Empirical evidence for treating complement raising in another way than subject raising is provided in Van Eynde & Augustinus (2013). It hinges on the interaction of raising with the argument realization principle, the binding principles and the passive lexical rule. In a nutshell, while the integration of unrealized SUBJ requirements in the ARG-ST list of the raising lexemes meshes well with the independently motivated treatments of argument realization, binding and passivization, the integration of unrealized COMPS requirements in the ARG-ST lists of the selecting lexemes causes complications.

2.2 Adposition stranding as an instance of complement raising

Clauses in which an R-pronoun is realized in the left part of the Mittelfeld can be analyzed in the same way as the scrambling data in the previous section. The application of complement raising to (7), for instance, yields the structure in (16).

(16)



[7](15) assumes that complements are added one at a time, from the most to the least oblique. To allow the combination with two or more complements at once, as in a flat structure, the COMPS list of the head daughter has to be reformulated as in Ginzburg & Sag (2000, 34).

The verb's requirement for an adpositional complement (2) is immediately saturated, but the adposition's requirement for a nominal complement (3) is not. It is appended to the one of the mother and canceled after the addition of *daar* 'there'.[8]

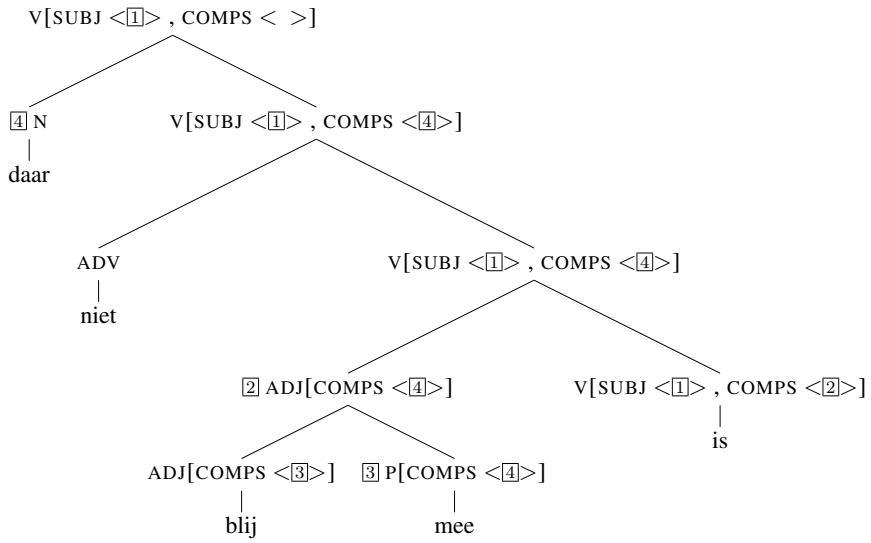
This treatment not only deals with raising out of PP complements of verbs, but also out of PP complements of non-verbal categories, such as the predicative adjective in (17) and the noun in (18).

(17) ...dat ze *daar* niet blij [__ mee] is.
...that she that+R not glad [__ with] is
'...that she is not glad about that.'

(18) ...dat ze *er* een boek [__ over] wil lezen.
...that she it+R a book [__ about] wants read
'...that she wants to read a book about it.'

The structure of (17) is spelled out in (19).

(19)



The adjective's requirement for an adpositional complement (3) is immediately saturated, and so is the verb's requirement for a predicative complement (2), but the adposition's requirement for a nominal complement (4) is not. It is appended to the one of the mother and propagated up the tree, till the point where the addition of *daar* triggers its cancellation. Notice that the requirement for a nominal complement (4) figures in the COMPS lists of the adposition and the nodes which dominate it, but not in the COMPS lists of the adjective or the verb.

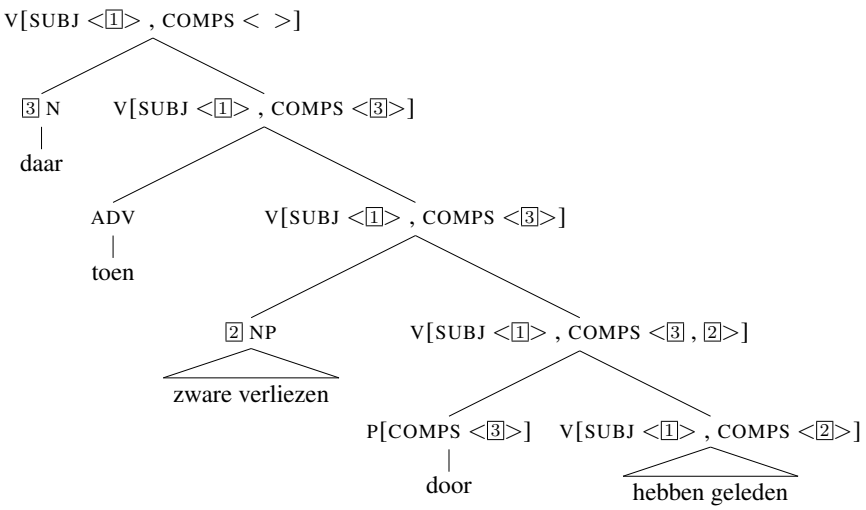
An advantage of this treatment of complement raising is that it also copes with the raising out of PP adjuncts, as in (20).

[8]The SUBJ value of the adposition is the empty list, in accordance with the canonical HPSG treatment of argument marking adpositions.

(20) ...dat we *daar* toen zware verliezen [__ door] hebben geleden.
...that we that+R then heavy losses [__ by] have suffered
'...that we suffered heavy losses because of that.'

The *door*-phrase in this sentence is not a complement of the verb, but a VP adjunct. It specifies the cause of the losses. Its COMPS requirement is not immediately saturated, but propagated in the by now familiar way.

(21)



In the generalized raising treatment, Hinrichs-Nakazawa style, this would require special measures: Since adjuncts are not selected by their head sister, the latter cannot inherit the former's COMPS requirement in the usual way. No such complications are needed in our treatment of complement raising.

In sum, the device which we use to model scrambling in the Mittelfeld can be used to model the raising out of PPs as well, no matter whether those PPs are complements of verbs, complements of other categories, or adjuncts.

2.3 A constraint on adposition stranding

A general constraint on Dutch adposition stranding is that it only affects complements which precede the adposition. Complements which follow the adposition must be realized within the PP.

(22) * Ze zegt dat ze hem/Hans soms nog [aan __] denkt.
* she says that she him/Hans sometimes still [of __] thinks

This is confirmed by the locative adverbs. They are homophonous to the R-pronouns, but in contrast to the latter they follow the adposition in PPs and must be realized within the PP.

(23) ... dat ze volgens ons niet [van hier] zijn.
... that they according-to us not [from here] are
'... that they are not from here according to us.'

(24) * ... dat ze volgens ons niet [hier van] zijn.
* ... that they according-to us not [here from] are

(25) * ... dat ze hier volgens ons niet [van __] zijn.
* ... that they here according-to us not [from __] are

The relevance of the linear order is also clear from the contrast between (26) and (27).

(26) ... dat de auto dagenlang [in de garage] stond.
... that the car days-long [in the garage] stood
'... that the car stood in the garage for days.'

(27) ... dat ze achteruit [de garage in] reed.
... that she backward [the garage in] drove
'... that she drove backward into the garage.'

The prepositional PP in (26) requires in situ realization of its complement, but its postpositional counterpart in (27) allows raising.[9]

(28) * ... dat de auto de garage dagenlang [in __] stond.
* ... that the car the garage days-long [in __] stood

(29) ... dat ze de garage achteruit [__ in] reed.
... that she the garage backward [__ in] drove
'... that she drove backward into the garage.'

This suggests that P-initial PPs are islands for complement raising. This is confirmed by the contrast between (30) and (31).

(30) Heb jij *daar* al [een boek [__ over]] gelezen?
have you that+R already [a book [__ about]] read?
'Have you already read a book about that?'

(31) * Heb jij *daar* al [aan een boek [__ over]] meegewerkt?
* have you that+R already [on a book [__ about]] collaborated?

(30) is well-formed, but (31) is not: The addition of the preposition *aan* 'on' blocks the raising of the complement.

In sum, complements can be raised out of a P-final PP, but not out of a P-initial PP. Besides, we have seen that the raised complement cannot only be an R-pronoun, but also a full NP, as in (29).

[9]The distinction corresponds to a difference in interpretation: While the prepositional PP has a locational interpretation, the postpositional one has a directional interpretation.

2.4 A generalization

The conclusion of the previous section naturally extends to verbs: While it is possible to raise complements out of a V-final VP, as shown in (12), it is not possible to raise a complement out of a V-initial VP. Indeed, complements which precede a V-initial VP, as in (32), are standardly treated as extracted, rather than as raised.

(32) Zo iemand [had ik nog nooit __ ontmoet].
so someone [had I still never __ met]
'Such a person I had never met before.'

In terms of the canonical HPSG treatment of extraction, the requirement of *ontmoet* 'met' for a direct object NP is subtracted from its COMPS list and added to its SLASH value (Ginzburg & Sag, 2000, 170-171).

Empirical evidence for differentiating complement raising from complement extraction is provided by the contrast between (33) and (34).

(33) Ze zouden jou/je volgens haar meteen moeten ontslaan.
they should you according-to her immediately must fire
'They should fire you immediately according to her.'

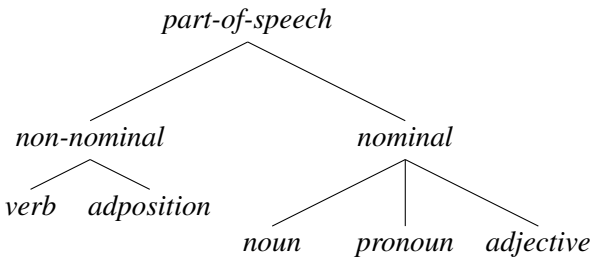
(34) Jou/*je zouden ze volgens haar meteen moeten ontslaan.
you should they according-to her immediately must fire
'It is you that they should fire immediately according to her.'

While both the full form and the phonologically reduced form of the object pronoun can be raised to the left part of the Mittelfeld, it is only the full form (*jou*) that can be extracted. The non-extractability of the reduced form (*je*) is due to the fact that non-subject constituents in the Vorfeld must be able to bear stress. Pronouns with a clear vowel can bear stress and, hence, occur in the Vorfeld, but pronouns with a mute vowel or without vowel cannot. The other non-extractable forms include the non-nominative personal pronouns *me, ze, d'r, 'r, 'm*, the impersonal *het, 't* and the reflexive *zich*, see (Van Eynde, 1999).

Assuming then that complement extraction is different from complement raising, in the same way as subject extraction is different from subject raising, we can formulate the ban on complement raising out of V-initial VPs and P-initial PPs in terms of a single constraint. To pave the way for its formulation we first add a feature, called POSITION, to the HEAD values of the verbs and the adpositions. As in X-bar theory, we assume that they are both non-nominal, i.e. [– N] (Chomsky, 1970; Jackendoff, 1977).[10]

[10] In X-bar theory, the lexical categories are analyzed in terms of the boolean features N and V: Verbs are [–N, +V], nouns are [+N, –V], adjectives are [+N, +V] and adpositions are [–N, –V].

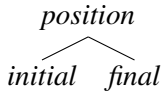
(35)



(36) *non-nominal* : [POSITION *position*]

The inventory of POSITION values is given in (37).

(37)



In terms of this dichotomy, the Dutch adpositions come in three types. Some are inherently initial, such as *met* 'with', *tot* 'to, till', *te* 'at, to' and *sinds* 'since, for', some are inherently final, such as *mee* 'with', *toe* 'to, till', *af* 'from' and *heen* 'towards', and some are used either way, such as *in* 'in', *op* 'up, on', *aan* 'on' and *van* 'of'. The verbs can be partitioned in the same way: Assuming that *initial* subsumes both the V1 and the V2 order, the inherently initial ones include the imperatives, the inherently final ones include the participles and the infinitives, and the underspecified ones include the non-imperative finite forms. Table 2 provides a survey.

POSITION	Adpositions	Verbs
Initial	*met, tot, te, sinds*	imperative
Final	*mee, toe, af, heen*	non-finite
Underspecified	*in, op, aan, van*	non-imperative finite

Table 2: The POSITION values of Dutch adpositions and verbs

Assuming that the underspecified values are resolved contextually, the constraint which blocks complement raising can now be formulated as follows:

(38) $\begin{bmatrix} \textit{hd-ph} \\ \text{SS} \mid \text{LOC} \mid \text{CAT} \mid \text{HEAD} \mid \text{POSITION} \quad \textit{initial} \end{bmatrix} \Rightarrow \begin{bmatrix} \text{SS} \mid \text{LOC} \mid \text{CAT} \mid \text{COMPS} \quad \langle\, \rangle \end{bmatrix}$

What (38) says, is that phrases whose POSITION value is resolved to *initial* must have an empty COMPS list. From this it follows that complements cannot be raised out of V-initial VPs nor out of P-initial PPs. Technically, the restriction to verbs and adpositions is due the fact that only these have the POSITION feature. Empirically, it is motivated by the fact that the other lexical categories do not abide by the

constraint. APs and NPs, for instance, allow complement raising, also if the head precedes its dependents. This was shown for the R-pronouns in (17–18), but it also holds for other kinds of dependents of adjectives and nouns, such as the bracketed PPs in (39–40).

(39) Ze zullen [met die resultaten] volgens mij niet [blij __] zijn.
they will [with those results] according-to me not [happy __] be
'They will not be happy with those results according to me.'

(40) Ze hebben [van elk dier] om het uur [foto's __] gemaakt.
they have [of each animal] about the hour [pictures __] made
'Every hour they made pictures of each animal.'

The restriction to verbs and adpositions is, hence, justified.

Broadening the scope to the functional categories, there is one that could be claimed to show the same behavior as the verbs and the adpositions, i.e. the complementizers. They take the same position in V-final clauses as the finite verbs in V-initial clauses, i.e. the first pole, also known as the *linke Satzklammer*. Since that position separates the Vorfeld from the Mittelfeld, complementizers are a barrier for complement raising. As a consequence, if we add the complementizers to the non-nominal parts of speech, they also have the POSITION feature, and since complementizers invariably precede their clausal complement (in Dutch), the value of that feature is always *initial*, so that the CPs are subsumed by the constraint in (38). Appealing as it is, we present this extension as hypothetical, since it presupposes that complementizers are heads of CPs. If one adopts the marker treatment of the complementizers instead, as in Pollard & Sag (1994, 44–46), the constraint has to be formulated in another way, for instance, by adding the requirement that the clausal sister of the complementizer must have an empty COMPS list. Since we do not know of any conclusive evidence in favour of the CP treatment, as opposed to the marker treatment, we leave the issue open.

2.5 A second constraint on adposition stranding

As for extraction out of PPs, it is clear that P-final PPs allow it, see (9–10). Some other examples are those in (41–42).

(41) Welke garage denk je dat hij toen achteruit [__ in] reed?
which garage think you that he then backward [__ in] drove?
'Which garage do you think he drove into backward?'

(42) Daar hebben we toen met de baas [__ over] gesproken.
that+R have we then with the boss [__ about] spoken
'That we talked about with the boss then.'

P-initial PPs, by contrast, are islands for extraction.

(43) * Welke garage denk je dat ze [in __] liggen?
* which garage think you that they [in __] lie?

(44) * Daar hebben we toen [met de baas [__ van]] gesproken.
* that+R have we then [with the boss [__ of]] talked

As indicated by the bracketing, the stranded *van* 'of' in (44) is a PP-adjunct of *baas* and, hence, included in the PP that is introduced by *met* 'with'. It thus contrasts with the stranded *over* 'about' in (42), which is a PP-complement of the verb *gesproken* 'spoken'.

In contrast to the P-initial PPs, the V-initial VPs do allow complement extraction, as shown in (32) and (45).

(45) Wie [heb je gisteren in Gent __ ontmoet]?
who [have you yesterday in Gent __ met]?
'Who did you meet in Gent yesterday?'

The same holds for clauses which are introduced by a complementizer.

(46) Wie denk je [dat ik gisteren in Gent __ ontmoet heb]?
who think you [that I yesterday in Gent __ met have]?
'Who do you think I met in Gent yesterday?'

The constraint on complement extraction is, hence, less restrictive than the one on complement raising:

(47) $\begin{bmatrix} \textit{hd-ph} \\ \text{SS} \mid \text{LOC} \mid \text{CAT} \mid \text{HEAD} \begin{bmatrix} \textit{adposition} \\ \text{POSITION} \quad \textit{initial} \end{bmatrix} \end{bmatrix} \Rightarrow \begin{bmatrix} \text{SS} \mid \text{NONLOC} \mid \text{SLASH} \quad \{\ \} \end{bmatrix}$

In plain words, P-initial PPs must have an empty SLASH set, but this constraint does not extend to V-initial VPs nor to clauses which are introduced by a complementizer.

3 A comparison with the uniform extraction analysis

A distinctive property of our analysis of adposition stranding is that we see it as the result of either complement raising or complement extraction. In this respect it differs from the existing HPSG treatments which see it as the result of complement extraction only. Rentier (1993) and Müller (1995) take the uniform extraction analysis for granted and focus mainly on the issue of how it can be spelled out in formal detail. Bouma (2000), by contrast, considers argument inheritance as an alternative for the uniform extraction analysis, but then argues against it. His four arguments will be discussed in this section.

3.1 Raising versus extraction

> "Prepositions which do not allow extraction (such as *met*) cannot be associated with an R-pronoun in the Mittelfeld either. If two different mechanisms are used to account for these two phenomena, such generalizations are easily lost." (Bouma, 2000, p.69)

Our answer to this objection is threefold. First, it is true that we have separate constraints on complement raising and complement extraction, see (38) and (47) respectively. This, however, is motivated by the fact that the former also subsumes the verbs, while the latter does not. If we use a single constraint, we lose the generalization that the constraint on complement raising also subsumes V-initial VPs.

Second, the empirical argument for differentiating raising from extraction is also valid for the R-pronouns. The reduced forms *er* and *d'r* can be raised, as in (48), but they cannot be extracted, as shown in (49).

(48) We hebben *daar/er/d'r* een liedje [__ over] gezongen.
we have that+R/it+R a song [__ about] sung
'We sang a song about that.'

(49) *Daar/*er/*d'r* hebben we een liedje [__ over] gezongen.
that+R/*it+R have we a song [__ about] sung
'That we sang a song about.'

Third, there are languages, such as English, which allow adposition stranding as a result of extraction, but not as a result of raising.

(50) a. *What* did you say she sang a song [about __]?
b. *That man* I never want to talk [to __] again.

(51) a. * I once heard *it* a song [about __].
b. * You should never *that* talk [about __] again.

The ban on complement raising follows from the Empty COMPS Constraint (ECC), see Ginzburg & Sag (2000, p.33).

(52) $\textit{phrase} \Rightarrow \left[\text{SS} \mid \text{LOC} \mid \text{CAT} \mid \text{COMPS} \left\langle \right\rangle\right]$

This constraint requires all phrases to have an empty COMPS list and is, hence, much more restrictive than (38), which requires this only for V-initial VPs and P-initial PPs.[11]

In sum, the use of separate constraints on complement raising and complement extraction is motivated by the fact that they have a different range of application, as well as by the fact that there are languages which have one but not the other.

[11] It might make sense to restrict the ECC to *headed* phrases, since coordinate phrases may consist of unsaturated words, as in *he buys and sells cars* and *are you for or against the war on terror*.

3.2 PP-internal order

> "As argument inheritance normally involves the composition of two COMPS lists, R-pronouns would have to be allowed on COMPS, even though they can, apart from a few exceptional cases, never appear in a position following the preposition." (ibid.)

This objection is based on the assumption that a nominal can only be a complement of an adposition if it follows that adposition, as in (1) and (5). This assumption, though, is hardly tenable in view of the fact that R-pronouns canonically precede the adposition when they are realized within the PP, as in (4) and (6), repeated in (53–54).

(53) Ze zegt dat ze soms nog [daar aan] denkt.
she says that she sometimes still [that+R on] thinks
'She says that she still thinks about it from time to time.'

(54) Ze zegt dat ze soms gewoon [nergens aan] denkt.
she says that she sometimes simply [nothing+R of] thinks
'She says that she simply thinks of nothing from time to time.'

It is also contradicted by the PP-internal order in (27), repeated in (55).

(55) ... dat ze achteruit [de garage in] reed.
... that she backward [the garage in] drove
'... that she drove backward into the garage.'

Rentier (1993, 116), who just like Bouma assumes that Dutch PPs must be *pre*positional, mentions (55) as a possible counterexample for his claim that Dutch has no postpositions, but then casts doubt on the adpositional status of *in*, claiming that it might be a particle. We do not share this doubt, since the adposition in (55) is clearly distinct from the separable verb particle in *inrijden*, a transitive verb denoting the activity of preparing a vehicle (car, bike, bus, ...) for use on the road. For detailed argumentation that postpositions like the one in (55) are distinct from particles, see Van Riemsdijk (1978, 90-108).

In addition, given that Dutch has V-final VPs and A-final APs, as shown in (56), the existence of P-final PPs is just what one expects.

(56) ... dat hij [[haar fratsen beu] is].
... that he [[her antics fed-up] is]
'... that he is fed up with her antics'.

In fact, Dutch is widely assumed to be predominantly head-final.

3.3 Argument inheritance

> "The set of argument inheritance verbs must now not only contain auxiliaries and modals, but all verbs which select a (prepositional) complement. Examples such as *Kim is er tevreden mee* introduce further complications for an argument inheritance approach, as it suggests that predicative adjectives and nouns must be argument inheritors as well." (ibid.)

This is a concern which we share. In fact, it is one of the reasons why we have chosen to model complement raising in terms of a constraint on headed phrases rather than in terms of a lexical constraint. The latter is only used to model subject raising and is, hence, limited to auxiliaries, modals and a few other verbs. It is not necessary to extend this to all the verbs, adjectives and nouns which select a PP complement, since the unsaturated COMPS requirements are propagated directly from the nonhead-daughter to the mother, see (16), (19) and (21).

3.4 Amalgamation of syntactic functions

> "In an argument inheritance approach, the relationship between valence and syntactically realized arguments has to be one-on-one, and thus there is no room for amalgamation of syntactic functions." (ibid.)

This objection requires a more lengthy rebuttal. To see what is meant with amalgamation, notice that *er* and *d'r* are not only used as R-pronouns and locative adverbs, but also as the semantically vacuous subject of existential clauses and impersonal passives, as in (57–58).

(57) Er/d'r staat een artikel over die mislukte aanslag in de krant.
there stands an article about that failed coup in the newspaper
'There is an article about that failed coup in the newspaper.'

(58) Er/d'r wordt nog elke dag over die mislukte aanslag geschreven.
there is still every day about that failed coup written
'That failed coup is still written about every day.'

These uses of *er* and *d'r* can be seen as the nominative counterparts of the non-nominative R-pronouns in PPs. If a clause contains both a nominative and a non-nominative R-pronoun, there is a tendency to drop the latter.

(59) Er/d'r staat (er) een artikel [__ over] in de krant.
there stands (it+R) an article [__ about] in the newspaper
'There is an article about it in the newspaper.'

(60) Er/d'r wordt (er) nog elke dag een artikel [__ over] geschreven.
there is (it+R) still every day an article [__ about] written
'Every day an article is written about it.'

If the two occurrences are adjacent, the elision is even obligatory.

(61) ... dat er/d'r (*er) een artikel [__ over] in de krant staat.
... that there (*it+R) an article [__ about] in the newspaper stands
'... that there is an article about it in the newspaper.'

(62) ... dat er/d'r (*er) een artikel [__ over] geschreven wordt.
... that there (*it+R) an article [__ about] written is
'... that there is an article written about it.'

Bouma (2000, 73) treats the clauses with a single occurrence of *er/d'r* as instances of function amalgamation: He assumes that the pronoun simultaneously fulfills two functions in such clauses.[12] This amalgamation, he claims, is impossible to model in terms of argument inheritance, since that device does not allow for discrepancies between valence and syntactically realized arguments.

We see this differently. In our analysis, there is no function amalgamation. Instead, we assume that the first *er* tokens in (59–62) have only one function, i.e. subject of the verb. The homophonous raised pronouns are not identified with that subject, but simply omitted.[13] Independent evidence for this analysis is provided by the fact that the omission also occurs in clauses which do not contain another instance of *er*, as in (63).

(63) Wie is (er) voor? En wie is (er) tegen?
who is (it+R) for? And who is (it+R) against?
'Who is in favor? And who is against?'

In fact, the optional omission in (59–60) and (63) is comparable to the kind of elision that is also attested by the post-auxiliary ellipsis in *yes we can*, by the intransitive use of verbs like *eat* and *read*, and by the intransitive use of the adposition in (64).

(64) We kunnen niet meer zonder.
we can not more without
'We can't do without it anymore.'

The obligatory omission in (61–62), for its part, is due to a constraint which blocks adjacent instances of *er/d'r*.

Similar remarks apply to clauses which contain the locative *er* 'there', such as (65).

(65) We gaan er de ontsnapte papegaai met een groot net vangen.
we go there the escaped parrot with a large net catch
'We are going to catch the escaped parrot there with a large net.'

[12] Technically, the amalgamation is modeled in terms of structure sharing: The LOCAL value of the subject is identified with the SLASH value of the adposition as well as with the BIND value of the verb.

[13] A similar assumption is made in the transformational treatment of Bennis (1986).

If the locative *er* is followed by the homophonous non-nominative R-pronoun, as in (66), the latter is omitted.

(66) We gaan er (*er) de ontsnapte papegaai [__ mee] vangen.
we go there (*it+R) the escaped parrot [__ with] catch
'We are going to catch the escaped parrot there with it.'

In sum, we do not need any room for amalgamation of syntactic functions, since the relevant data can be modeled in terms of the independently motivated omissibility of the (nominal) complements of adpositions.

4 Conclusions

The existing HPSG treatments of adposition stranding in Dutch provide a uniform extraction analysis, employing such nonlocal devices as SLASH and BIND, see Rentier (1993) and Bouma (2000). We endorse this analysis for the cases in which the extracted pronouns end up in the Vorfeld, but not for the cases in which they end up in the left part of the Mittelfeld. Since the latter concerns a middle-distance (bounded) dependency, we propose a treatment that is based on argument inheritance. More specifically, we employ the version of argument inheritance in Van Eynde & Augustinus (2013), which differentiates subject raising from complement raising, and show how the treatment of complement raising, originally motivated to model scrambling, can be used to model adposition stranding as well. In order to avoid overgeneration, we added two constraints: (38) blocks complement raising out of P-initial PPs and V-initial VPs, while (47) blocks complement extraction out of P-initial PPs. Having spelled out the treatment, we discussed Bouma's objections against the use of argument inheritance for the analysis of adposition stranding, and demonstrated that none of them sticks.

The resulting tretament is not only economical, it also accounts for the fact that languages which abide by the Empty COMPS Constraint, such as English, not only lack the kind of scrambling that we find in Dutch and German, but also the kind of adposition stranding that results from complement raising (as opposed to the kind of adposition stranding that results from complement extraction).

References

Abeillé, Anne, Danièle Godard, Philip Miller & Ivan Sag. 1998. French bounded dependencies. In S. Balari & L. Dini (eds.), *Romance in HPSG*, 1–54. Stanford: CSLI Publications.

Bennis, Hans. 1986. *Gaps and Dummies*: Katholieke Hogeschool Tilburg dissertation.

Bouma, Gosse. 2000. Argument realization and Dutch R-pronouns. In R. Cann, C. Grover & P. Miller (eds.), *Grammatical Interfaces in HPSG*, 51–76. Stanford: CSLI Publications.

Bouma, Gosse & Gertjan van Noord. 1998. Word Order Constraints on Verb Clusters in German and Dutch. In E. Hinrichs (ed.), *Complex Predicates in Nonderivational Syntax*, New York: Academic Press.

Broekhuis, Hans. 2013. *Syntax of Dutch. Adpositions and Adpositional Phrases*. Amsterdam: Amsterdam University Press.

Chomsky, Noam. 1970. Remarks on Nominalization. In R.A. Jacobs & P.S. Rosenbaum (eds.), *Readings in English transformational grammar*, 184–221. Waltham, Mass.: Ginn and Company.

Fleischer, Jürg. 2002. Preposition Stranding in German Dialects. In S. Barbiers, L. Cornips & S. van der Kleij (eds.), *Proceedings of the Workshop on Syntactic Microvariation, August 2000*, 116–151. Amsterdam: Meertens Instituut.

Ginzburg, Jonathan & Ivan Sag. 2000. *Interrogative Investigations*. Stanford: CSLI Publications.

Haeseryn, Walter, Kirsten Romijn, Guido Geerts, Jos de Rooij & Maarten van den Toorn. 1997. *Algemene Nederlandse Spraakkunst*. Nijhoff and Wolters Plantyn.

Hinrichs, Erhard & Tsuneko Nakazawa. 1989. Flipped out: AUX in German. In *Papers from the 25th Regional Meeting of the Chicago Linguistic Society*, 193–202. Chicago.

Hinrichs, Erhard & Tsuneko Nakazawa. 1994. Linearizing AUXs in German Verbal Complexes. In J. Nerbonne, K. Netter & C. Pollard (eds.), *German in Head-driven Phrase Structure Grammar*, 11–37. Stanford: CSLI Publications.

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

Kunkel-Razum, Kathrin & Franziska Münzberg (eds.). 2006. *DUDEN. Die Grammatik. Unentbehrlich für richtiges Deutsch.* Mannheim/Leipzig/Vienna/Zürich: Dudenverlag.

Monachesi, Paola. 1998. Decomposing Italian clitics. In S. Balari & L. Dini (eds.), *Romance in HPSG*, 305–357. Stanford: CSLI Publications.

Müller, Stefan. 1995. Scrambling in German – Extraction into the *Mittelfeld*. In Benjamin K. T'sou & Tom B. Y. Lai (eds.), *Proceedings of the tenth Pacific Asia Conference on Language, Information and Computation*, 79–83. City University of Hong Kong.

Pollard, Carl & Ivan Sag. 1994. *Head-driven Phrase Structure Grammar*. Stanford/Chicago: CSLI Publications and University of Chicago Press.

Rentier, Gerrit. 1993. Dutch Clitic Pronouns and Preposition Stranding. In *Computational Linguistics in the Netherlands. Papers from the Third CLIN Meeting*, 109–120. Tilburg.

Sag, Ivan, Thomas Wasow & Emily Bender. 2003. *Syntactic Theory. a Formal Introduction. Second Edition*. Stanford: CSLI Publications.

Van Eynde, Frank. 1999. Major and minor pronouns in Dutch. In G. Bouma, E. Hinrichs, G.-J. Kruijff & R. Oehrle (eds.), *Constraints and Resources in Natural Language Syntax and Semantics*, 137–151. Stanford: CSLI Publications.

Van Eynde, Frank & Liesbeth Augustinus. 2013. Why and How to Differentiate Complement Raising from Subject Raising in Dutch. In Stefan Müller (ed.), *Proceedings of the 20th International Conference on Head-Driven Phrase Structure Grammar*, 222–242. Stanford: CSLI Publications.

Van Riemsdijk, Henk. 1978. *A Case Study in Syntactic Markedness*. Lisse: The Peter de Ridder Press.

Focus case outside of Austronesian: An analysis of Kolyma Yukaghir

Olga Zamaraeva
University of Washington

Emily M. Bender
University of Washington

Proceedings of the 21st International Conference on
Head-Driven Phrase Structure Grammar

University at Buffalo

Stefan Müller (Editor)

2014

Stanford, CA: CSLI Publications

pages 176–196

Zamaraeva, Olga & Emily M. Bender. 2014. Focus case outside of Austronesian: An analysis of Kolyma Yukaghir. In Stefan Müller (ed.), *Proceedings of the 21st International Conference on Head-Driven Phrase Structure Grammar, University at Buffalo*, 176–196. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2014.10.
CC-BY 4.0

Abstract

Case is traditionally approached as a lexical phenomenon in HPSG. The LinGO Grammar Matrix customization system, an HPSG-based grammar engineering toolkit and also a typological meta-resource, includes several options for case assignment, and one of them, 'focus case', assumes that case of the participants in basic clauses is handled via lexical rules rather than lexical entries. This phenomenon was previously only attributed to a group of Austronesian languages, and thus the focus case differed from all other case options in the Matrix which were attested for across language families. Our analysis of Kolyma Yukaghir, a nearly extinct language of North-Eastern Russia, shows that focus case can be successfully used outside of Austronesian family and therefore that the option is more universal than it was previously thought.

1 Introduction

Since at least Pollard & Sag, 1994, case assignment in HPSG has been seen as a lexical phenomenon: "assignment of case to complements [...] is simply treated as part of subcategorization" (*Ibid.* p.30). Simple subcategorization isn't enough to capture all facts of case assignment (Przepiórkowski, 1996), but it still serves as a useful core: On the one hand, lexicalist analyses of valence alternations add lexical rules (e.g. for passive or causative alternations) which produce new case assignments as part of new subcategorization frames (e.g. Müller, 2001) and on the other hand, many authors have proposed a distinction between lexical and structural case (e.g. Heinz & Matiasek, 1994; Przepiórkowski, 1996), allowing the actual morphological form of structural cases to be sensitive to the syntactic environment. However, even with all of these extensions, the analysis of case still crucially involves the subcategorization frames of verbs.

Drellishak, 2009 adds a library for case to the LinGO Grammar Matrix customization system (Bender et al., 2002, 2010).[1] This library, based on a rigorous review of typological literature, provides nine choices of general case system (including nominative-accusative, ergative-absolutive, several kinds of splits, and none) which in turn provide basic case subcategorization frames appropriate to the language type as well as facilities for defining additional case values and additional case frames (e.g. quirky case). The ninth of these types, called 'focus-case' is of particular interest here: Unlike the other case systems, it does not involve underlying case frames for verbs. Rather, verbs are required to undergo inflectional rules which constrain the case on the arguments. It also stands out from the others, as far as Drellishak was able to determine, in only being attested in one language family (Austronesian).

In this paper, we describe an analysis of the case system of Kolyma Yukaghir (ISO 639-3: yux), a language of North-Eastern Russia. Strikingly, though this

[1] Drellishak's library only concerns lexical case.

system has some clear differences to that found in Tagalog and other Austronesian languages, the focus case option of Drellishak's library supports an elegant implementation. In §2, we outline the fundamentals of Drellishak's analysis. §3 presents a brief sketch of Yukaghir. We present our analysis (implemented in a Grammar Matrix based grammar) in §4 and evaluate it against a hand-constructed testsuite in §5. Finally, §6 describes how the grammar can be extended to handle the interaction of case marking and definiteness.

2 Focus case in the Grammar Matrix

In Austronesian languages (including Tagalog), the case frame of the verb depends on an inflectional marker that picks out which argument is in 'focus'.[2] This is illustrated in (1)–(2), where the AGENT.FOCUS and PATIENT.FOCUS markers on the verb indicate which role the *ang*-marked element should be interpreted as filling.

(1) Bumili ang babae ng baro
bought-AGENT.FOCUS FOCUS woman PATIENT dress
'The woman bought a dress.' [tgl] (Drellishak, 2009, p.54)

(2) Bimili ng babae ang baro
bought-PATIENT.FOCUS AGENT woman FOCUS dress
'A/the woman bought the dress.' [tgl] (Drellishak, 2009, p.54)

Drellishak's analysis of this system leaves the case of arguments underspecified in the lexical entries for the verbs, and then uses lexical rules to fill in case frames. More specifically, these lexical rules form a required 'position class' that every verb must go through.[3] These lexical rules attach the affixes for AGENT.FOCUS, PATIENT.FOCUS etc. while also constraining the case values of all elements of the verb's arguments structure appropriately.

3 Kolyma Yukaghir

Yukaghir languages[4] are considered either a small isolated family or a distant relative of Uralic languages (Fortescue, 1996, p.17). Kolyma Yukaghir is nearly extinct (estimates vary from just 5 to 300-400 speakers, depending on the definition of fluency) (Maslova, 2003, p.1). It is basically a SOV agglutinating language (Fortescue, 1996, p.17), though systematic deviations from SOV word order are attested

[2] It is not clear whether this actually corresponds to the information-structural notion of focus in these languages, as a 'focused' constituent other than the verb is required in every clause; see §3.2.

[3] A position class in a Grammar Matrix-derived grammar is a type describing a set of lexical rules which take the same inputs and in turn can serve as inputs to the same set of further lexical rules (Goodman, 2013).

[4] There are two varieties: Kolyma and Tundra Yukaghir. They are typologically very similar, but whenever we say "Yukaghir" in this paper, we mean Kolyma Yukaghir.

(Maslova, 2003, p.341) and pragmatics often determines the word order as well, so that even in simple clauses, different word orders are possible (Maslova, 2003, p.17). It has singular and plural number, first, second, and third person, no gender system, and fairly developed morphology. There is inflection for future and non-future tense, and a periphrastic construction for past tense, using a nominalized verb form (*Ibid*. p.6). There are also aspect distinctions (including imperfective, ingressive, resultative, and habitual), and four major moods (inferential, prospective, irrealis, and periphrastic prospective) (*Ibid*. p.6). Subjects and objects are easily dropped when recoverable from context (*Ibid*. p.9), unless they are in focus (*Ibid.* p.326).

Focus, which is expressed grammatically, is one of the most notable features of the language. Of particular interest is the role of information structure in syntax of the clause. In literature about Yukaghir, the well-established and traditional use of the term 'focus' is justified primarily by "grammaticalized association of Focus role with canonical 'focus-presupposition' contexts, where the information about the situation being described is (directly or indirectly) 'activated' by the time of utterance or can be viewed as a part of extra-linguistic context, the referential identity of one participant being the only unknown piece of information about the situation" (Maslova, 2005, p.600). Thus in Yukaghir, the term 'focus' seems to basically correspond to the usual information structure sense, as summarized, for example, in Song, 2014. The focus marking on the nouns is referred to as predicative case in Maslova, 2003.

3.1 Case in Kolyma Yukaghir

Yukaghir exhibits a fairly complex system of case marking. Maslova distinguishes 9 cases for nouns and 12 for pronouns, of which the following seem to be most involved in the basic intransitive and transitive verb patterns: nominative, predicative, accusative, instrumental, and a form called 'pronominal accusative' which we analyze as object non-focus marker ('NFO') for non-3rd person pronouns (in contrast to nouns which use the unmarked form both in positions requiring nominative case and where pronouns would take NFO). As we will describe further below, the choice of the case frame in simple clauses typically depends not on the particular verb, but rather on the information structure of the clause, as well as on the person value of the subject in transitive clauses.

3.1.1 Intransitive clauses

In intransitive clauses, the subject can either be in focus or it can be neutral to focus. If the subject is focused, it takes predicative case, and the verb takes the marker *-l* glossed SF for 'subject focus' (3). When an intransitive subject is not focused, it takes nominative case and the verb agrees with it in person and number (4).[5]

[5]Abbreviations used in Yukaghir examples:

(3) tāt touke-lek jede-l
CA dog-PRED appear-SF
'Then {a dog}$_{Foc}$ appeared.' [yux] (Maslova, 2003, p.9)

(4) met ejre-je
I.NOM walk-1SG
'I walked.' [yux]

3.1.2 Transitive clauses

In transitive clauses, the verb registers whether or not the object is in focus (OF) (5). The marking of this information interacts with the marking for subject agreement: There are are two sets of subject agreement markers, one used when the object is focused and one used when it is not. The choice of the case frame for most verbs that we considered depends on whether the verb has the OF marker, whether the subject is 3rd person (6) or not (7), and additionally whether the object is definite (8). The paradigms are summarized in Table 1. The subject is always in nominative case; in focused transitive clauses, the object is always in predicative case. In non-focused transitive clauses, if the subject is 1st or 2nd person, the object appears in the NFO (non-focused object) form: the so-called 'pronominal accusative' for first and second person pronouns and the zero-marked form (equivalent to nominative) for nouns and third person pronouns (9). If the subject is 3rd person, a definite object will be accusative (e.g. (8)) and an indefinite instrumental (e.g. (6)).

(5) tet-ek aŋči-nu-ŋile
you-PRED search-IPFV-3PL.OF

12PER	1st and 2nd Person
1SG	1st Person Singular
ACC	Accusative case
CA	connective adverbial
FUT	Future
NONFUT	Nonfuture
INSTR	Instrumental case
INTR	Intransitive
IPFV	Imperfective aspect
NEG	Negation
NF	Non-focus
NFO	Non-focused Object case; 'Pronominal Accusative' in (Maslova, 2003)
NOM	Nominative case
OF	Object Focus
PERNUM	Person and Number marking
PL	Plural
PRED	Predicative (Focus) case
SF	Subject Focus
TR	Transitive
NON3PL	Not 3rd plural

All glosses except 'NFO' are from Maslova, 2003. Examples without citations are constructed by the authors and have been verified by a Yukaghir expert.

OF		Non-Focus	
S 3rd	S non-3rd	S 3rd	S non-3rd
NOM-PRED		NOM-ACC/INSTR	NOM-NOM/NFO

Table 1: Transitive clause case frames

'It is {you}$_{Foc}$ whom they are seeking.' [yux] (Maslova, 2003, p.153)

(6) tudel tolow-le kudde-m
he.NOM deer-INSTR kill-3SG
'He killed a deer.' [yux] (Maslova, 2003, p.10)

(7) met tolow kudede
I.NOM deer.NOM kill.1SG
'I killed a deer.' [yux] (Maslova, 2003, p.10)

(8) tudel met kønme-gele juø-m
he.NOM my friend-ACC see-3SG
'He saw my friend.' [yux] (Maslova, 2003, p.10)

(9) met tet-ul juø
I.NOM you-NFO see.1SG
'I saw you.' [yux] (Maslova, 2003, p.10)

We leave ditransitive, chained, and non-finite clauses out of the discussion. Generally only finite clauses can be marked in this way for focus. There are other case frames which involve other cases such as ablative, as well as classes of verbs which require locative, but the most basic intransitive and transitive patterns can be summarized as above.

3.2 Focus in Kolyma Yukaghir and in Tagalog

Not surprisingly, given the great distance separating their geographic distributions, Kolyma Yukaghir and Tagalog are quite different typologically and in particular in the details of their focus case systems. Tagalog uses adpositional marking to mark nouns for case (Comrie, 1989, p.121); Yukaghir uses inflectional marking. The Yukaghir verbal markers implicated in this system also carry person and number information, where the Tagalog markers do not (Cruz & Shkarban, 1966). Finally, where every Tagalog clause must have an element which is marked with the so-called 'focus marker' *ang* (Comrie, 1989, p.121), Yukaghir, allows for clauses where no argument bears focus marking. We believe that this means that the Yukaghir system is more likely to actually be marking information structural contrasts: It seems implausible that a language would systematically disallow verbs to be in (narrow) focus, but if Tagalog *ang* were to strictly represent information structural focus, narrow verbal focus would be impossible. Yukaghir, on the other hand, allows clauses where the NPs are not marked for focus, leaving open the possibility of the verb carrying focus.

In the initial development of our Yukaghir grammar, the focus case option was passed over, at least in part because it was presented as only appropriate for Austronesian languages. We ended up developing by hand a system that was quite close to Drellishak's analysis, and so went back to the customization system and created a grammar fragment using the focus case option. This is the grammar fragment presented in §4 and evaluated in §5 below, and it shows that despite its typological and geographical distance from Austronesian, when viewed through the framework of HPSG, Yukaghir shares a certain typological similarity with Tagalog and its kin.

4 Analysis of Kolyma Yukaghir in the Grammar Matrix

In order to test our analysis of the case system, we needed our grammar fragment to handle enough other basic facts to parse our test sentences which exemplify choice of case frame in intransitive and transitive clauses. We created the fragment through the Grammar Matrix customization system (Bender et al., 2002, 2010).[6] In the process, we made a few simplifying decisions: In particular, we said that word order is 'free', though in some cases some orders are not possible.[7] We also did not implement any of the moods and only implemented habitual aspect as an example. Generally, in this fragment we only try to cover the basic case assignment in simple clauses. For number, person and gender the Matrix provides suitable options (sg/pl, 1/2/3 and no gender). It is also possible to model sentential negation (which is simple negation expressed by prefix *el-* on the verb), but this doesn't affect the analysis of case.

4.1 Case

We picked the focus case option for Kolyma Yukaghir despite the note in the Grammar Matrix customization system questionnaire which presented it as relevant only to the Austronesian family. Unlike the other case system options in the Grammar Matrix customization system, this one does not provide a set of default argument structures with case frames pre-defined. Instead, it supports the implementation of verbal lexical rules which fill in case requirements on the verb's arguments. As described below, this option supports an effective analysis of the Yukaghir system.

We restricted our analysis to the subset of cases described in the previous section (nominative, predicative, accusative, instrumental, and NFO). The customization system questionnaire requires specification of 'focus case', 'A-case' and 'O-case' when the focus case option is chosen.[8] We specified these as predicative,

[6]The grammar fragment and our testsuite are available online at http://depts.washington.edu/uwcl/matrix/yukaghir/.

[7]Maslova (2003) says that the language is basically SOV, though generally other orders are possible. In our testsuite, only VSO and VOS were deemed ungrammatical by the Yukaghir expert who gave the judgments.

[8]A and O refer to the subject/object of a transitive verb, respectively.

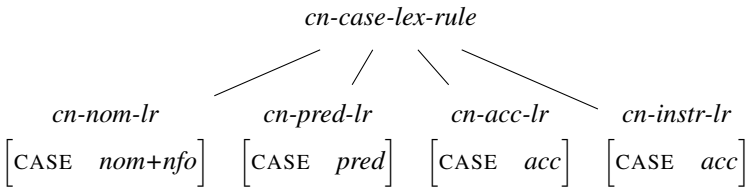


Figure 1: Noun Case lexical rules hierarchy

accusative and nominative respectively. Instrumental and NFO were listed as 'additional cases'. Later in the morphology section, we were able to model the case frames based on this choice. The customized grammar was able to handle the facts of Yukaghir case (as detailed in §5 below) without further modification, with one exception: The customization system does not yet provide facilities for constraining discourse/cognitive status (Borthen & Haugereid, 2005) of arguments, and so our grammar overgenerates with respect to the distribution of accusative and instrumental objects. This can of course be remedied by hand-editing of the grammar (see §6).

4.2 Lexicon

We populated the lexicon with all the personal pronouns and a few basic common nouns, to be able to test example sentences. We did not include possessives or other parts of speech. This means that for testing, we used constructed examples such as 'He saw a friend' rather than 'He saw my friend'. In order to model a definite NP, we included a demonstrative determiner, so it is possible to parse a sentence 'He saw this friend'. We included a few basic verbs: *ejre* ('walk'), *jede* ('appear'), *juø* ('see'), *kudede* ('kill'). Intransitive and transitive verbs are the only verb classes that we included in the grammar fragment (these are the only types supported at present by the Grammar Matrix customization system questionnaire); a more extensive grammar would of course require more classes (for example, verbs of cognition, qualitative verbs).

4.3 Morphology

The main goal of our grammar fragment is to implement case requirements on verbs and the associated marking of case on nouns (and pronouns), which requires lexical rules for both. We assume a morphological analyzer to account for some of the morphophonological phenomena, which are described in Maslova (2003). Accordingly, our grammar targets morpheme-segmented, regularized forms such as those shown in the examples in this paper.

4.3.1 Noun and Pronoun inflection rules

The key observation for the analysis for nouns and pronouns is that the distribution of the NFO case marker on the pronouns is exactly like the zero nominative case marker for nouns as direct objects (zero marker on the object noun if the subject is non-3rd person). This is exemplified in (7) and (9), repeated here as (10) and (11). Compare with (12) and (13) to see that the surface forms of the NFO and the NOM are the same for nouns, but different for first and second person pronouns. Examples (14), (15), and (16) show the 3rd person pronoun forms (which are the same for NOM, NFO as well as PRED).

(10) met tolow kudede
I.NOM deer.NOM kill.1SG
'I killed a deer.' [yux] (Maslova, 2003, p.10)

(11) met tet-ul juø
I.NOM you-NFO see.1SG
'I saw you.' [yux] (Maslova, 2003, p.10)

(12) tolow met-gele kudede-m
deer.NOM I.ACC kill-3SG
'The deer killed me.' [yux]

(13) tet met-ul juø-mek
you.NOM I.NFO see-2SG
'You killed me.' [yux]

(14) tudel met-gele kudede-m
he.NOM I.ACC kill-3SG
'He killed me.' [yux]

(15) met tudel-gele kudede
I.NOM he.ACC kill.1SG
'I killed him.' [yux]

(16) met tudel kudede-me
I.NOM he.NOM kill-OF.1SG
'I killed him$_{Foc}$.' [yux]

Thus, it makes sense to say the zero-marker (Maslova's 'nominative') marks nouns for nominative ('A-case' in the focus case library terms) as well as for NFO (see Figure 1). In contrast, pronouns must have separate lexical rules for nominative and NFO (Figure 2). Then, in the verb lexical rules section, it is sufficient to constrain the *tr-nf* object to be in NFO case (shown in Figure 4).

We analyze both the (surface) instrumental and accusative as marking nouns as [CASE *acc*].[9] This allows for the instrumental forms but does not properly

[9]This analysis leaves open the possibility of a separate, homophonous, instrumental case rule that produces [CASE *instr*] nouns with definiteness unspecified.

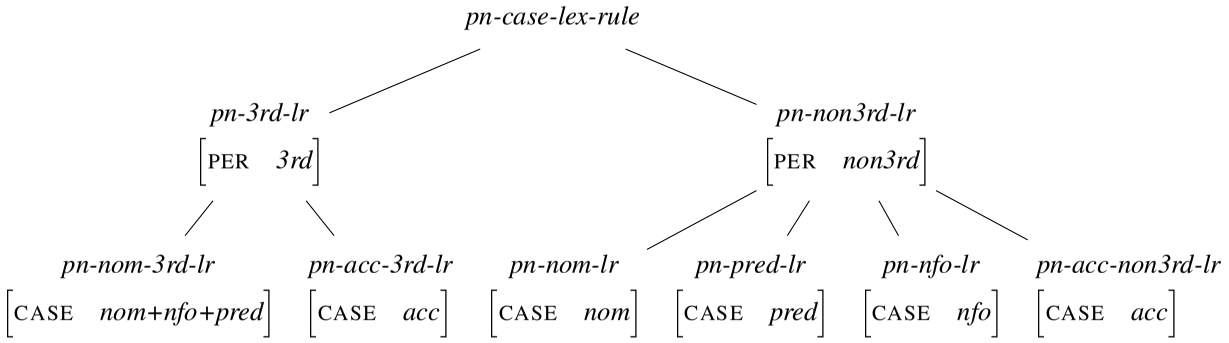


Figure 2: Pronoun Case lexical rules hierarchy

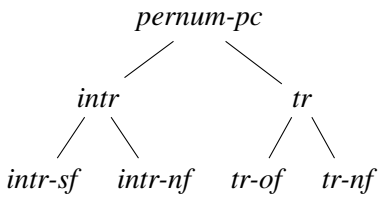


Figure 3: Verb lexical rules

limit their distribution: The additional constraints required (on definiteness) can be added directly via tdl editing, but are not at present supported by the Grammar Matrix customization system (see section 6).

4.3.2 Verb inflection rules

Unlike other options for case systems in the Grammar Matrix customization system, focus case does not provide a set of argument structure values with case specified to use in the definition of verb classes. Rather, picking the option creates case frames that can be used in verb lexical rules as constraint on the verb's argument structure. (The high level hierarchy for these rules is presented in Figure 3; further detail is given below.) In case of Yukaghir, the automatically generated 'nom-pred' case frame option can be used in modeling the *tr-of* lexical rule type. In general, three lexical rule types give rise to the majority of the lexical rules: intransitive non-focused subject (*intr-nf*), transitive focused object (*tr-of*), and transitive non-focused object (*tr-nf*). Since there is only one marker for any verb with a focused subject, intransitive focused subject (*intr-sf*) type can be realized via a single lexical rule, much like Drellishak's treatment of Tagalog transitive clauses (Drellishak, 2009, p.66). In order to model the various patterns in transitive clauses found in Yukaghir, additional lexical rules are required, as discussed below.

In the basic morphology that we consider the position classes are chained as follows (Maslova, 2003, p.149):

(17) NEG-stem-ASPECT-TENSE-AGR

However, in verbs agreeing with 3PL subjects, the tense marker is in between two markers which both mark the verb for person and number.[10] Therefore it is more practical to assume an additional position class, which comes before tense, classifies the verb's subject as 3rd person plural or not, and is typically still accompanied by a person and number marker after the tense marker:

(18) juø-ŋi-te-m
see-3PL-FUT-3PL

[10]We differ from Maslova in analyzing these as both marking both person and number (i.e. an instance of multiple exponence), as this leads to a correct association of form and morphosemantic features in our system.

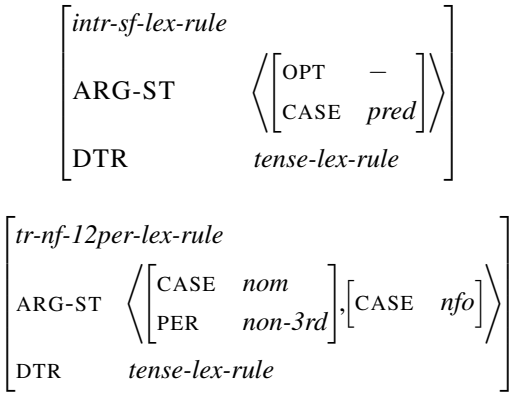


Figure 4: Sample lexical rules

'They will see.' [yux] (Maslova, 2003, p.140)[11]

(19) juø-0-te-m
see-NON3PL-FUT-3SG
'(S)he will see.' [yux] (Maslova, 2003, p.140)

(20) ejre-ŋi-0-0
walk-3PL-NONFUT-3PL
'They walk.' [yux] (Maslova, 2003, p.140)

(21) ejre-ŋi-te-j
walk-3PL-FUT-3PL
'They will walk.' [yux] (Maslova, 2003, p.140)

With the 3PL position class in place, and with the future/non-future position class taking the output of the 3PL rule as input, we are ready to create the rules involved in determining case frames. We do this in terms of a hierarchy where the supertypes *intr-sf*, *intr-nf*, *tr-of*, and *tr-nf* constrain the case frames. *Intr-sf* is instantiated by just one lexical rule instance. The others all have subtypes describing full paradigms of person/number values on the subject, and thus all four supertypes correspond to the AGR position class in (17).

Implementing the paradigm for intransitives is relatively simple: There is one rule for subject focus (with no agreement distinctions), which attaches the subject-focus marker to the verb and constrains its subject to be [CASE *pred*], and a set of rules for non-focused subjects that indicate agreement in person and number and constrain the subject to be [CASE *nom*].

One source of complexity in the transitive paradigms is that the 3rd person marker depends also on whether the tense is future or not, since if it is, the above mentioned 3PL position class marker is nonzero while the AGR marker shows less variety.

[11]Examples (18)–(21) come from a morphological table which we used for reference and were not originally in the form of IGT.

The transitive branch is further complicated by the split on the subject's person value in terms of the case frame. In addition, the *intr-sf* rule also specifies that the subject is [OPT −] (cannot be dropped), and the *tr-of* rule does the same for its object. Figure 4 illustrates how the case frames are constrained for *intr-sf* and *tr-nf-12per*, respectively. Figures 5–7 show the type hiearchies and key constraints contributed by each type for these lexical rules.

The choices file implementing this analysis includes 699 individual choices (pieces of information). Among these, many are dedicated to defining the 35 lexical rules that handle case assignment and person/number agreement with the subject (32 in the AGR position class and 3 in the 3PL position class). While it may seem that this is a lot, it is in fact a manageable analysis of four distinct paradigms offered in Maslova's verb morphology table (Maslova, 2003, p.140). Furthermore, we note that this results in a working, testable grammar fragment for Yukaghir.

4.3.3 Information structure and argument optionality

The Grammar Matrix customization system also provides a library for modeling information structural constraints (Song, 2014). As noted above, it appears that the focus-related morphology in the verb does mark focus in the information structural sense. Using Song's library we are able to model this by selecting the option that creates affixes as focus markers and then adding the specification that the *intr-sf* and *tr-of* rules constrain their subject and object, respectively, to be in focus. This has the effect that the rules add an element to the verb's ICONS list representing the information structural meaning (in Song's terminology) of focus on the relevant argument. More specifically, giving this specification to the customization system causes the type *tr-of*, supertype to all of the object focus rules, to inherit from *add-icons-comp-rule*,[12] shown in Figure 8 and then to further constrain the particular type of ICONS element added to be *focus*.

An added benefit of modeling the information structural meaning is that it interacts with another constraint provided by the information structure library to correctly model the generalization that the focused arguments (subject of SF verbs, object of OF verbs) cannot be dropped, despite fairly free argument optionality within Yukaghir. Song models this as a general constraint on focus cross-linguistically: Focused arguments must be overtly realized and so the argument drop rules (non-branching productions provided by the argument optionality library (Saleem & Bender, 2010) constrain the argument being discharged to be [ICONS-KEY *non-focus*]. Without any further specification on our part, the customized grammar thus correct rules out examples like (22).

(22) *jede-te-l
appear-FUT-SF
Intended: '(someone) will appear.' [yux]

[12]The type *add-icons-comp-rule* is provided in the Matrix core grammar as part of Song's information structure library.

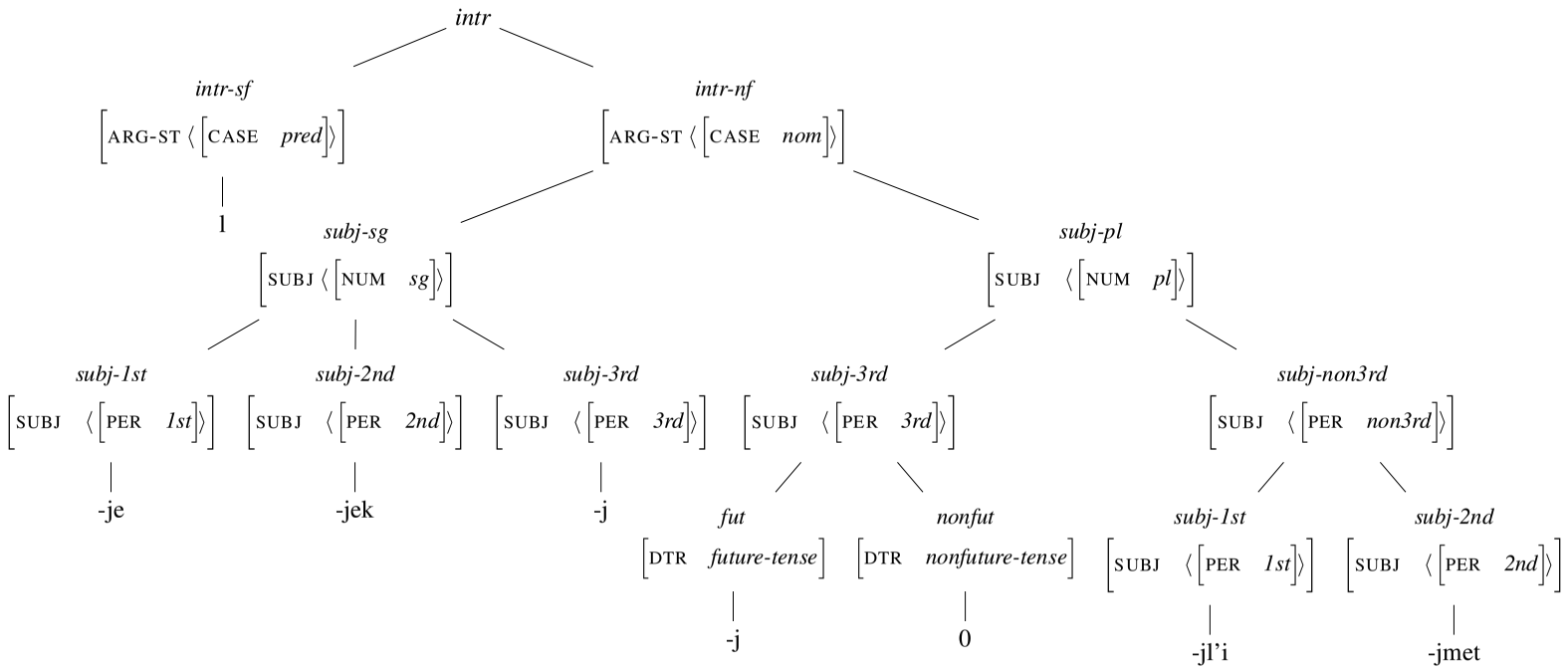


Figure 5: Intransitive verbs lexical rules

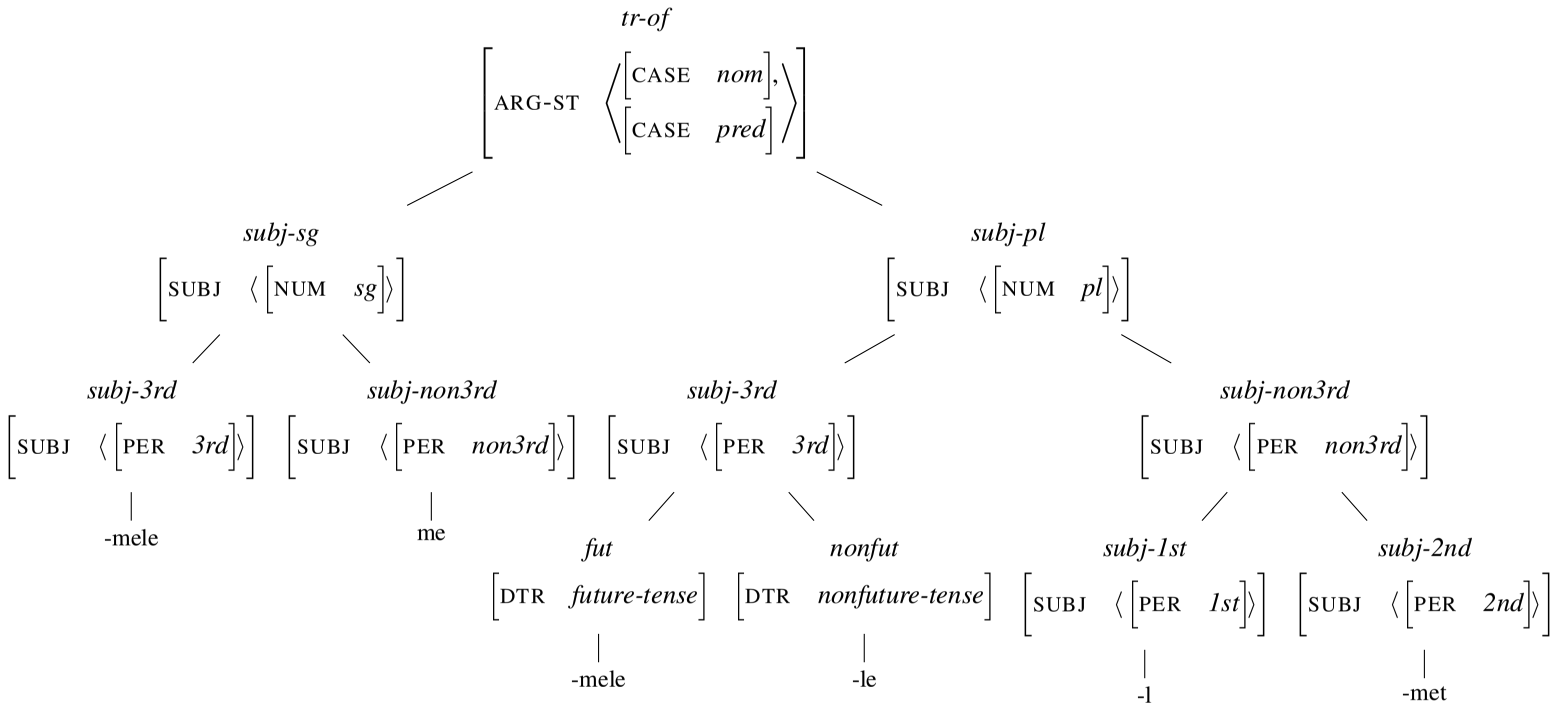


Figure 6: Transitive-object-focus verbs lexical rules

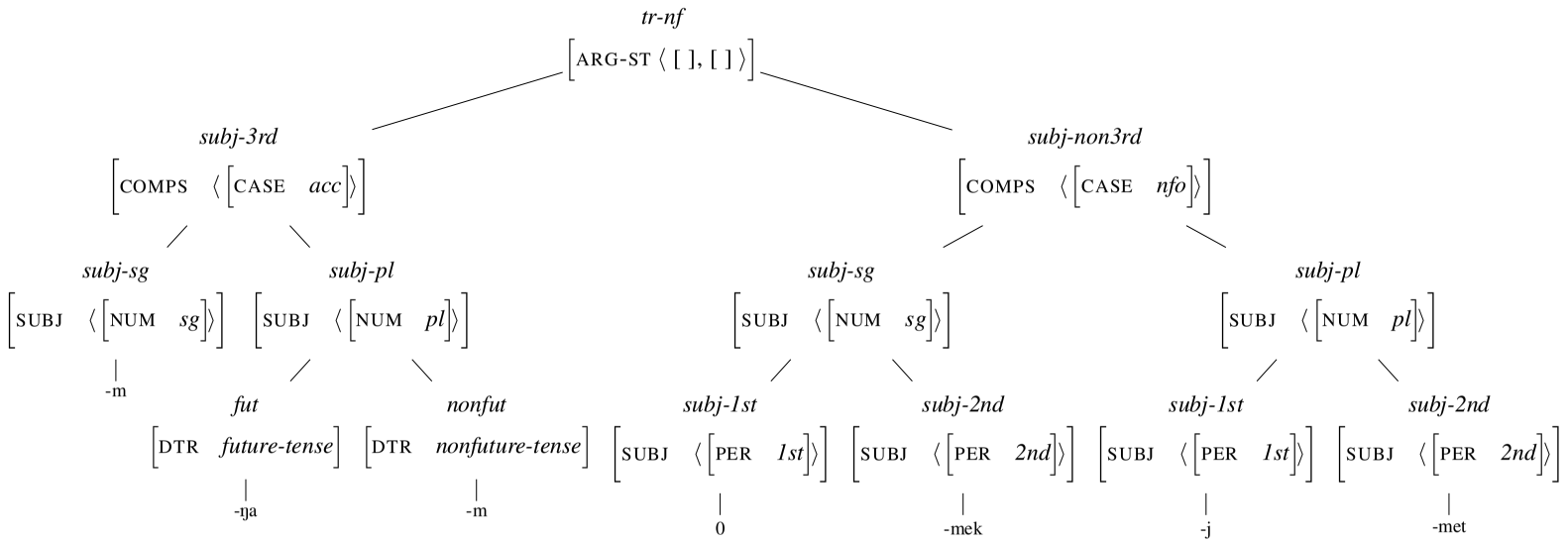


Figure 7: Transitive-object-non-focus verbs lexical rules

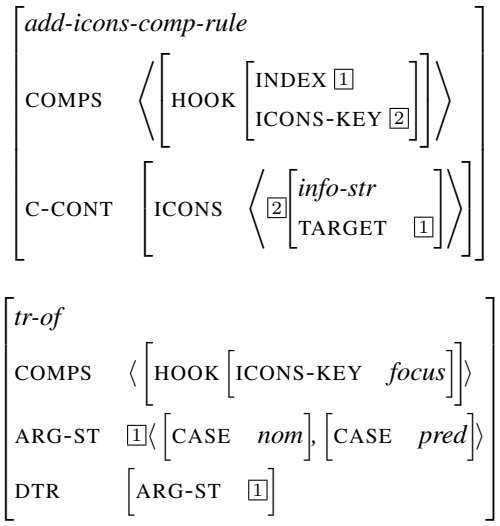


Figure 8: Constraints on ICONS modeling information structure of focused objects

5 Evaluation

We used 161 sentences,[13] 81 grammatical, 80 ungrammatical, to test the grammar that we obtained via the Matrix, using the LKB (Copestake, 2002) and [incr tsdb()] (Oepen, 2001). The performance of the grammar is summarized in Table 2.[14]

Coverage	Overgeneration
84.0 %	6.2%

Table 2: Yukaghir Grammar Performance

Inspection of the testsuite shows that the unanalyzed sentences (undergeneration) are the ones that represent phenomena beyond the scope of our grammar fragment (ditransitives, copula, attributive forms).

The overgeneration falls into two classes. On the one hand, our grammar allows VSO and VOS orders, marked as ungrammatical in the testsuite, due to our simplifying decision to use the free word order option. On the other hand, we also overgenerate because the grammar does not model the correlation between case and definiteness on non-focused objects. Specifically, as shown in (6) and (8) (repeated

[13] Most of the sentences are very simple, only involving a subject, and object, and a verb. Due to the lack of such basic examples in the available literature, most test examples are constructed from more complex sentences by removing possessives (and changing accusative case on the object to instrumental where appropriate), adverbs, and other words not directly involved in the transitive or intransitive pattern. All sentences have been verified by a Yukaghir expert.

[14] These numbers reflect changes to the testsuite compared to the results presented at the conference. In particular, we removed some of the more complex sentences with larger vocabulary, since they were not supposed to be covered by the grammar, and changed the word order from SOV to Free, after consulting with our Yukaghir expert. The version of the grammar and testsuite as evaluated can be found at http://depts.washington.edu/uwcl/matrix/yukaghir/.

here as (23) and (24)), indefinite non-focused objects are marked with instrumental case, while definite non-focused objects bear accusative. Furthermore, as shown in (25), definite objects require an overt determiner.[15] Our grammar fragment, however, accepts (25) and furthermore does not reflect the definiteness contrast in the semantics.

(23) tudel tolow-le kudde-m
he.NOM deer-INSTR kill-3SG
'He killed a deer.' [yux] (Maslova, 2003, p.10)

(24) tudel met kønme-gele juø-m
he.NOM my friend-ACC see-3SG
'He saw my friend.' [yux] (Maslova, 2003, p.10)

(25) *tudel tolow-gele kudde-m
he.NOM deer-ACC kill-3SG
Intended: 'He killed a/the deer.' [yux]

6 Extension: Definiteness of the Accusative

The grammar fragment we created with the customization system does not handle the definiteness distinction discussed immediately above because the Grammar Matrix customization system does not yet provide a library for definiteness marking. However, the Grammar Matrix's core grammar does provide support for adding such an analysis to the grammar by hand, based on Borthen & Haugereid's (2005) analysis of cognitive status of references (Bender & Goss-Grubbs, 2008).

To do this, we need to add demonstrative determiners, as in Figure 9. This type is instantiated by lexical entries such as *tiŋ* 'this', which contribute two elementary predications, an existential quantifier and the demonstrative (adjective) relation. Most relevantly here, they also constrain their SPEC's COG-ST (cognitive status) value to *activ+fam*, the value typical of definite NPs. Meanwhile, the lexical rules for instrumental and accusative case must also be enhanced. The former adds the information that the noun's index is [COG-ST *type-id*]. This value, typical of indefinite NPs, is incompatible with the constraints on the determiner. The latter constrains the noun's specifier requirement to [OPT −], requiring an overt determiner,[16] successfully blocking (25).

7 Conclusion

The case system of Yukaghir is relatively complex: the case frames of verbs depend on both the person of the subject (familiar from languages with split-ergativity)

[15] In this matter, opinions differ, and it is not clear whether examples like (25) would indeed be rejected by the speakers or not. We followed Maslova (2003) and assumed that (25) would be rejected.

[16] There are no indefinite determiners in the language; indefinites are bare NPs.

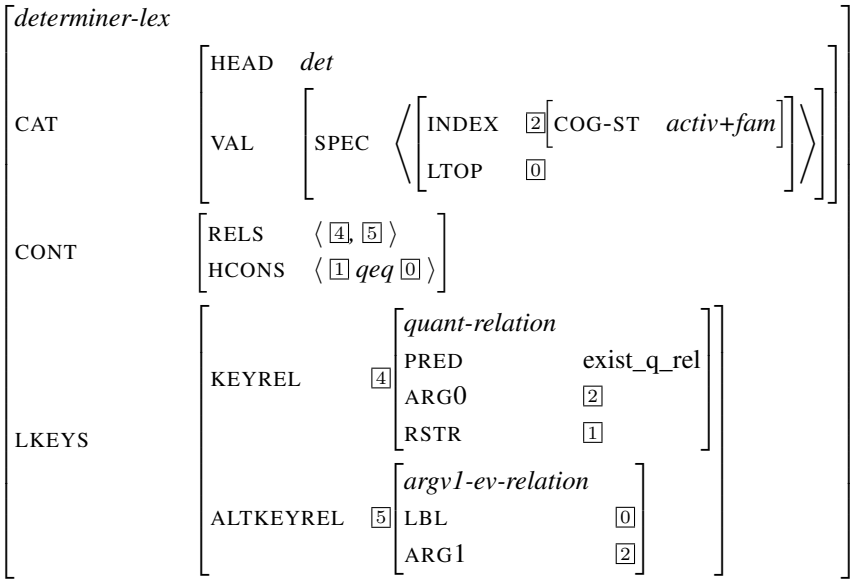


Figure 9: Constraints on *determiner-lex*

and on whether the verb bears focus marking. On the surface, the pattern does not immediately resemble that of the Austronesian languages which inspired the 'focus-case' option in Drellishak's case library for the Grammar Matrix. However, on closer inspection, Drellishak's analysis provides the core of an elegant account of this complex system, whether or not the Austronesian notion of 'focus' corresponds to the traditional information structure sense.

We have tested that analysis by creating a grammar fragment with the Grammar Matrix customization system. This fragment is able to handle all of the patterns described above, with the exception of the association between accusative case and definiteness, for which we offer a solution that requires some editing of the grammar by hand. This grammar fragment can be further extended as well: We find that the customization system's information structure library (Song, 2014) provides a suitable analysis for the information structural effects of focus marking in Yukaghir.

References

Bender, Emily M., Scott Drellishak, Antske Fokkens, Laurie Poulson & Safiyyah Saleem. 2010. Grammar customization. *Research on Language & Computation* 1–50. 10.1007/s11168-010-9070-1.

Bender, Emily M., Dan Flickinger & Stephan Oepen. 2002. The grammar matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. In John Carroll, Nelleke Oostdijk &

Richard Sutcliffe (eds.), *Proceedings of the workshop on grammar engineering and evaluation at the 19th international conference on computational linguistics*, 8–14. Taipei, Taiwan.

Bender, Emily M. & David Goss-Grubbs. 2008. Semantic representations of syntactically marked discourse status in crosslinguistic perspective. In *Proceedings of STEP 2008*, College Publications.

Borthen, Kaja & Petter Haugereid. 2005. Representing referential properties of nominals. *Research on Language and Computation* 3(2). 221–246.

Comrie, Bernard. 1989. *Language universals and linguistic typology: Syntax and morphology*. Chicago, IL: University of Chicago Press.

Copestake, Ann. 2002. *Implementing typed feature structure grammars*. Stanford, CA: CSLI Publications.

Cruz, Manuel & Lina I. Shkarban. 1966. *Tagal'slij yazyk*. Moscow, Russia: Nauka.

Drellishak, Scott. 2009. *Widespread but not universal: Improving the typological coverage of the Grammar Matrix*: University of Washington dissertation.

Fortescue, Michael. 1996. Grammaticalized focus in Yukaghir. is it really grammaticalized and is it really focus? In *Content, expression and structure: Studies in Danish structural grammar*, 17–38. John Benjamins Publishing.

Goodman, Michael Wayne. 2013. Generation of machine-readable morphological rules with human readable input. *UW Working Papers in Linguistics* 30.

Heinz, Wolfgang & Johannes Matiasek. 1994. Argument structure and case assignment in German. *German in Head-Driven Phrase Structure Grammar* 46. 199–236.

Maslova, Elena. 2003. *A grammar of Kolyma Yukaghir* Mouton Grammar Library (Book 27). Berlin: Mouton De Gruyter.

Maslova, Elena. 2005. Information structure in Tundra Yukaghir and typology of focus structures. In M. M. Jocelyn Fernandez-Vest (ed.), *Les langues ouraliennes aujourd'hui*, 599–610. Honore Champion.

Müller, Stefan. 2001. The passive as a lexical rule. In *Proceedings of the 7th international conference on HPSG*, 247–266.

Oepen, Stephan. 2001. [incr tsdb()] — Competence and performance laboratory. User manual. Technical report Saarbrücken, Germany.

Pollard, Carl & Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar* Studies in Contemporary Linguistics. Chicago, IL and Stanford, CA: The University of Chicago Press and CSLI Publications.

Przepiórkowski, Adam. 1996. Case assignment in Polish: Towards an HPSG analysis. In Claire Grover & Enric Vallduví (eds.), *Studies in HPSG*, vol. 12 Edinburgh Working Papers in Cognitive Science, 191–228. Centre for Cognitive Science, University of Edinburgh.

Saleem, Safiyyah & Emily M. Bender. 2010. Argument optionality in the lingo grammar matrix. In *Coling 2010: Posters*, 1068–1076. Beijing, China: Coling 2010 Organizing Committee. http://www.aclweb.org/anthology/C10-2123.

Song, Sanghoun. 2014. *A grammar library for information structure*: University of Washington dissertation.

Part II
Contributions to the Workshop

Language description and the lexicon: Verbs of wearing in two Oaxacan languages

George Aaron Broadwell

University at Albany, State University of New York

Proceedings of the 21st International Conference on
Head-Driven Phrase Structure Grammar

University at Buffalo

Stefan Müller (Editor)

2014

Stanford, CA: CSLI Publications

pages 198–216

Broadwell, George Aaron. 2014. Language description and the lexicon: Verbs of wearing in two Oaxacan languages. In Stefan Müller (ed.), *Proceedings of the 21st International Conference on Head-Driven Phrase Structure Grammar, University at Buffalo*, 198–216. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2014.11.
CC-BY 4.0

Abstract

Verbs of wearing show unusual linking properties in two languages of Oaxaca: San Dionisio Ocotepec Zapotec and Copala Triqui. Several distinct lexical types must be recognized, and their linking to grammatical relation is not predictable on general principles.

1. Two Oaxacan language documentation projects

Oaxaca is a state in southern Mexico with a rich variety of indigenous languages. I and my students at University at Albany have been engaged in language documentation projects on two of these languages since about since about 1998.

The two languages are San Dionisio Ocotepec Zapotec and Copala Triqui. San Dionisio Ocotepec Zapotec is a Zapotecan language spoken in Oaxaca, Mexico by about 2,000 people.[1] It is spoken in the town of San Dionicio Ocotepec, which is 15-20 miles south of Tlacolula and 10-15 miles southwest of Mitla in the Central Valley of Oaxaca.

Copala Triqui is a Mixtecan language, spoken in the mountains of western Oaxaca, Mexico.[2] There about 30,000 speakers in Oaxaca, Mexico (and in other parts of Mexico and the United States).

Zapotecan and Mixtecan languages are part of the larger Otomanguean stock, but Proto-Otomanguean has a time-depth of 4000-7000 years (Kaufman and Justeson 2009). Zapotec and Triqui languages are thus not very closely related to each other. They are also spoken in non-adjacent parts of Oaxaca, so similarities are probably due to influences of the language area, rather than direct contact or borrowing.[3]

[†]I thank the audience at HPSG 2014 for helpful comments and suggestion. I also thank speakers of San Dionisio Ocotepec (Luisa Martinez and Pedro Morales) and Copala Triqui (Román Vidal-López, Monica deJesus Ramírez, and others) for their patience and help in understanding their languages.

[1]San Dionicio Ocotepec Zapotec (ISO 639-3 code: ztu) is written in a practical orthography, which is adapted from the practical orthographies for other Zapotec languages spoken in the Valley of Oaxaca. In this orthography, <x> = /ʃ/, <zh> = /ʒ/, <dx> = /dʒ/, <ch> = /tʃ /, and <qu> = /k/ before front vowels, <c> = /k/ elsewhere. <y> represents IPA /j/. Doubled vowels are long. SDZ is a language with four contrastive phonation types: breathy <Vh>, creaky <V'V>, checked <V'>, and plain <V>. High tone is marked with an acute accent, low with a grave.

2. Basic syntactic properties

Both of these languages are head initial, with Verb Subject Object order and prepositions.

(1) [ztu]
Ù-cà' *Juáàny* *bgùù* *lòò* *dù'ù*
com-put Juan knot to rope

'Juan tied the rope'

(2) [trc]
A'níí *Mariá* *chraa* *rá* *yoó* *a.*
put Maria tortilla in straw:container decl

'Maria puts the tortilla in the tenate (straw container).'

Both have the possessum before the possessor and both show a distinction between inalienable (3, 5) and alienable (4, 6) possession.

(3) [ztu]
lòò *Juààny*
face Juan

'Juan's face'

(4) [ztu]
x-pè'cw *Juààny*
poss-dog Juan

'Juan's dog'

---

[2]The orthography for Copala Triqui (ISO 639-3 code: trc) used in this paper is based on the practical orthography developed by Barbara and Bruce Hollenbach of the Summer Institute of Linguistics for their translation of the New Testament. I follow their usage in the representation of the consonants, including the following conventions: <x> = [ʃ], <xr> = [ʂ] (a retroflex alveopalatal sibilant), <ch> = [tʃ], <chr> = [tʂ], <c> = [k] (before front vowels), <qu> = [k] before back vowels, [v] = [β] and <j> = [h]. <Vn> represents a nasalized vowel. Triqui has five level tones (1, 2, 3, 4, 5) and three contour tones (13, 31, 32), as discussed in Hollenbach (1984), but this paper shows only the tonal marking of the popular orthography, which is detailed enough to show the relevant morphological distinctions of the language

[3]One might compare them to languages in two different branches of Indo-European, such as German and Spanish.

(5) *rihaan xnii*
[trc] face child

'the child's face'

(6) *se cafeé xnii*
[trc] poss coffee child

'the child's coffee'

Unlike Zapotec, Copala Triqui has case marking. An accusative particle *man* is obligatory before pronominal objects and optional before other objects:

(7) *Que-ne'e Mariá (man) Juán.*
[trc] com-see Maria (acc) Juan.

'Maria saw Juan.'

(8) *Que-ne'e Mariá man so'.*
[trc] com-see Maria (acc) 3msg.

'Maria saw him.'

(9) **Que-ne'e Mariá so'.*
[trc] com-see Maria 3msg.

(Intended: 'Maria saw him.')

3. Verbs of wearing and linking theory

Both Zapotec and Triqui have complex sets of verbs that are used to describe 'wearing situations'. Different types of 'clothing/adornment' require different verbs:

(10) *Cáá lèènt lòò Màrìì.*
[ztu] wear glasses face Maria

'Maria is wearing glasses.'

(11) *Bè'cy Juààny pàntlòòn.*
[ztu] wear Juan pants

'Juan is wearing pants.'

Below I will demonstrate that in (10), *lèènt* 'glasses' is the subject of the sentences, while in (11), the wearer *Juààny* 'Juan' is the subject.

For San Dionisio Ocotepec Zapotec, there are five verbs of wearing.

Verb	**Subject**	**Object**	**Clothing/Adornment**
R-àjcw	wearer	clothing	shirts, dresses
Rr-cáá	clothing	poss'd body part	glasses, shoes, watches
R-áày	wearer	clothing	skirts, diapers
Rr-bè'cy	wearer	clothing	pants, underwear
N-ù'ú	clothing	poss'd body part	earrings, false teeth, wigs, stockings, belts, girdles, gloves, ribbons, hats

In order to discuss the two possible syntactic realizations of verbs in this group, I will introduce some terms to distinguish the types.

Zapotec verbs which make the wearer the subject of these verbs have a mapping like English. I will call these W-verbs (mnemonic for "wearer subject"). Verbs for wearing pants, underwear, skirts, shirts, diapers, and dresses show this mapping. The subcategorization for W-verbs is <wearer, clothing>. The body part is not explicit, but must be a large part of the trunk.

Zapotec verbs which make the clothing the subject have the reverse mapping from English. I will call these C-verbs (For "clothing subject"). Verbs for wearing items such as hats, shoes, belts, earrings, etc. show this mapping. The subcategorization for C-verbs is <clothing, <body part <wearer> >. Peripheral body parts use C-verbs, and this is also the case for clothing worn on the trunk that fails to cover all the trunk (belt, girdle).

The following table shows the five Zapotec verbs categorized by type.

Verb	**Type**	**Clothing/Adornment**
R-àjcw	W-verb	shirts, dresses
Rr-cáá	C-verb	glasses, shoes, watches
R-áày	W-verb	skirts, diapers

Verb	Type	Clothing/Adornment
Rr-bè'cy	W-verb	pants, underwear
N-ù'ú	C-verb	earrings, false teeth, wigs, etc.

3.1. The linking problem for Zapotec

Many/most syntactic theories assume that the linking of arguments to grammatical relations is predictable given the semantics or semantic roles. E.g. Lexical Mapping theory in LFG (Bresnan and Kanerva 1989) or Proto-role entailments (Dowty 1991).

However, in what way are the semantic roles for 'John' different in the sentences like the following?

- *'John is wearing pants'*
- *'John is wearing glasses'*

For verbs of wearing, neither argument necessarily displays the semantic properties that have most frequently been cited as diagnostic-- e.g. volition, causation, incremental themes, or movement.

Since these general approaches to linking do not offer clear insight into these verbs, it seems more promising to pursue approaches that use more carefully structured accounts of lexical semantics, such as Simpler Syntax (Culicover & Jackendoff 2005) or some work in HPSG (Davis 2001).

4. The syntax of wearing in San Dionisio Ocotepec Zapotec

4.1. Subject properties in Zapotec

What is the evidence that for some verbs of wearing, the wearer is subject, while for others the clothing is the subject? First, Zapotec is a strictly VSO language so simple word order is an excellent diagnostic. VOS is never grammatical in Zapotec:

(12) *Cáá lèènt lòò Màrìì.*
[ztu] wear glasses face Maria

'Maria is wearing glasses.'

(13) **Cáá lòò Màrìì lèènt.*
[ztu] wear face Maria glasses

(Intended: 'Maria is wearing glasses.')

Zapotec also shows subject raising with negative predicates like *íity* ‘not’ and *cáády* ‘still not’ (Broadwell 2012). Consider these C-verbs (clothing=subject):

(14) *Íity lèènt ní-càà lòò Màrìì.*
[ztu] not glasses neg-wear face Maria

'Maria didn't wear glasses.'

(15) *Cáády lèènt ní-càà lòò Màrìì.*
[ztu] still:not glasses wear face Maria

'Maria didn't wear glasses.'

Raising applies only to the subject; other objects of the verb cannot appear after negatives:

(16) **Íity lòò Màrìì ní-càà lèènt .*
[ztu] not face Maria neg-wear glasses

(Intended: 'Maria didn't wear glasses.')

W-verbs have a VSO order which shows that the wearer is the subject:

(17) *Bé'cy Juààny pàntlòòn*
[ztu] wear Juan pants

'Juan is wearing pants.'

The wearer subject of a W-verb may also undergo raising:

(18) *Cáády Juààny gwé'cy pàntlòòn.*
[ztu] still:not Juan pot:wear pants

'Juan is still not wearing pants.'

The subject of a W-verb passes an additional subject test: ability to be an equi target:

(19) *B-yè'là'z Juààny gwé'cy pàntlòòn*
[ztu] com-forget Juan pot:wear pants

'Juan forgot to wear pants.'

In contrast, C-verbs cannot undergo equi, since the subject of a matrix verb like 'forget' cannot be equal to the clothing which is the subject of its complement:[4]

(20) *B-yè'là'z Juààny n-yù'ú xhùmbrèl ìicy
[ztu] com-forget Juan neg-wear hat head

'Juan forgot to wear pants.'

(21) *B-yè'là'z ìicy Juààny n-yù'ú xhùmbrèl
[ztu] com-forget head Juan neg-wear hat

'Juan forgot to wear pants.'

4.2. Causation and wearing in Zapotec

Because 'wear' represents a stative event, we might expect its mapping to be less predictable. 'Put (clothing) on [SELF]' might be expected to contain CAUSE and an agent in the semantics. Yet even in this case, the agent is not necessarily the subject.

My consultant reports that both C-verbs and W-verbs have another possible reading -- 'put (clothes) on [SELF]'. The mapping is identical for this reading.[5]

Consider the following examples:

(22) Ù-lè'cy Juààny pàntlòòn.
[ztu] perf-wear Juan pants

'Juan wore pants.' OR 'Juan put his pants on.'

(23) Cwáá lèènt lòò Màrìì.
[ztu] perf:wear glasses face Maria

[4]To express a roughly equivalent meaning, speakers have to switch the complement verb to something like 'pick up' or 'take' which has the wearer as subject.

[5]To get the 'put on' reading, it is necessary to shift the examples into an aspect, such as perfective, that is compatible with the eventive reading. 'Wear' readings typically have a verb in the stative aspect, which has either the prefix /n(i)-/ or is the bare stem of the verb.

'Maria wore glasses.' OR 'Maria put on her glasses.'

Thus even when the subject is agentive, as in (23), it is not the subject of a C-verb.

Although there is no morphology that differentiates the stative and eventive versions of these verbs, there are morphological causatives which refer to ‘putting (clothing) on (another person)’. Distinct valence patterns are seen for W-verbs and C-verbs.

The causative of a C-verb is a ditransitive:

(24) [ztu]
Ù-càà	*dòctòrr*	*lèènt*	*lòò*	*Màrìì.*
perf-put	doctor	glasses	face	Maria.

'The doctor put glasses on Maria.'

The causative of a transitive W-verb, however is still a transitive. The ‘wearer’ must undergo ‘possessor lowering’:

(25) [ztu]
B-cwè'cy	*Juààny*	*x-pàntlòòn*	*xì'ny=ní'.*
perf-put	Juan	poss-pants	child=3ref

'Juan put pants on his child.'

(26) [ztu]
Ù-gwááy	*Màrìì*	*x-fààld*	*xì'ny=ní'.*
perf-put	Maria	p-skirt	child=3ref

'Maria put a skirt on his child.'

In these examples, *xpàntlòòn xì'nyní'* 'his child's pants' and *xfààld xì'ny=ní'* 'his child's skirt' are NP constituents, as shown by standard constituency tests.

(27) [ztu]
[Túú	*x-pàntlòòn]*	*b-cwè'cy*	*Juààny.*
who	poss-pants	perf-put	Juan

'Who did Juan put pants on?' (Lit. 'Whose pants did Juan put?')

In the causative of C-verbs, the NP of clothing/adornment and the possessed body part do not form a constituent. This can be seen by the inability of these two elements to undergo pied-piping in wh-questions:[6]

(28) *[Túú lèènt lòò] ù-càà dòctòrr?
[ztu] Who glasses face perf-put doctor

(Intended 'Who did the doctor put glasses on?')

(29) *[Túú lòò lèènt] ù-càà dòctòrr?
[ztu] Who face glasses perf-put doctor

(Intended 'Who did the doctor put glasses on?')

The causative wearing verbs in Zapotec are shown in the following chart.[7]

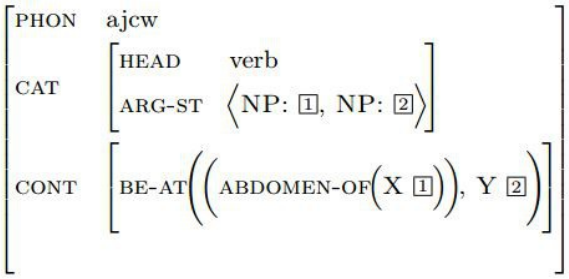
Verb	**Type**	**Subj**	**Obj**	**Second Obj**	**Clothing**
rr-gwàjcw	W-verb	Causer	Poss'd clothing	--	shirt,dress
rr-cwáá	C-verb	Causer	Clothing	Poss'd body part	glasses, etc
rr-gwáày	W-verb	Causer	Poss'd clothing	--	skirt, diaper
rr-cwè'cy	W-verb	Causer	Poss'd clothing	--	pants, underwear
r-gù'ú	C-verb	Causer	Clothing	Poss'd body part	earrings, etc.

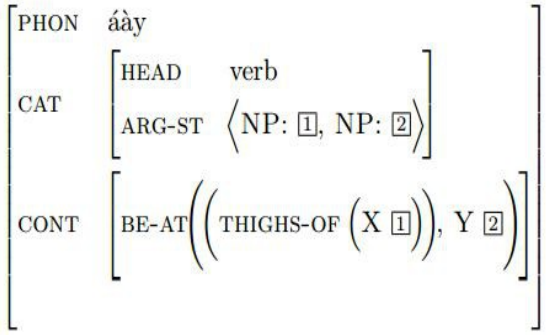
4.3. Tentative analysis of Zapotec wearing

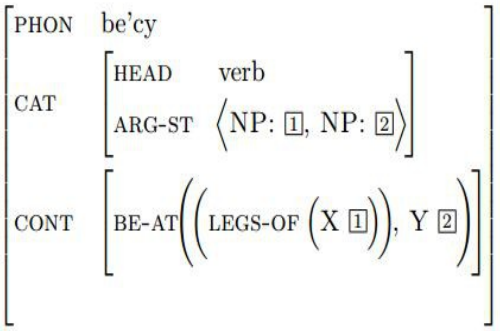
Tentative analyses of the three W-verbs are shown below:

[6]The word orders in pied-piping contexts are discussed in some detail in Broadwell (2001).

[7]The morphological causative is expressed in a number of distinct ways in Zapotec, but for these verbs by a /gw-/ or /g-/ prefix or a /-w-/ infix.



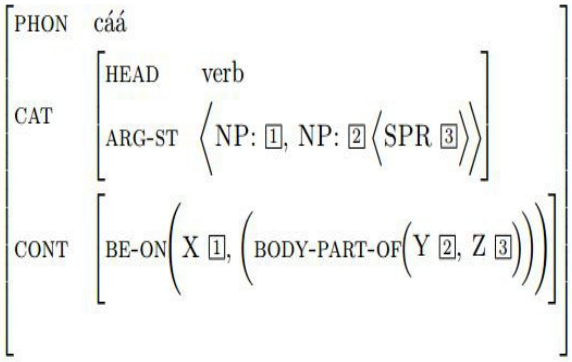


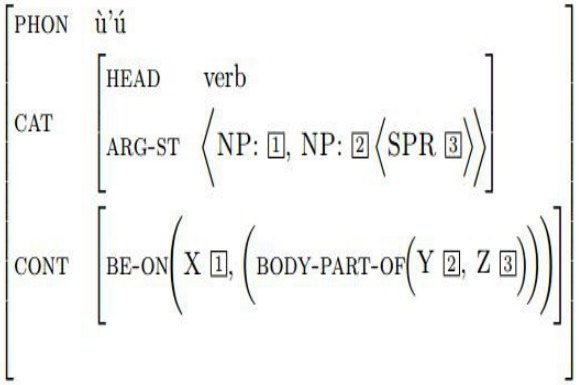


Verbs of this type involve an implied body part (abdomen, legs, thighs) which is not expressed as an argument. We can call these body parts central ; they serve as metonymies for the whole person.

BE-AT predicates are symmetrical, with Figure-Ground relations established by preferences like Person > Non-Person; More mobile > Less mobile. For these verbs, one argument qualifies as a Person, and is thus the Figure.[8]

In contrast, the C-verbs require expression of the body part, thus their ARG-ST shows three elements linked to the lexical semantics. I tentatively give them the following analyses:





For verbs where the body part must be explicit in the argument structure, the wearer is in an embedded structure in both CONTENT and ARG-ST. I have given the same structures for *cáá* and *ù'ú*, on the assumption that the subtle difference between them is probably a matter of selection.

[8] I list the Figure before the Ground in the lexical decomposition.

4.4. Uses of these verbs outside wearing contexts

The three W-verbs of Zapotec are specialized for wearing various items, and don't have other uses. In contrast, the C-verbs are specialized uses of locational verbs that are used for other items as well:

Rr-cáá is used to describe the location of high or hanging objects, e.g. fruit on a tree. *N-ù'ú* is used as a general verb of existence when no more specific verb (sit, stand, hang, …) is appropriate.

The use of these verbs in wearing situations shares the semantics of BE-AT with other uses of these verbs.

5. Verbs of wearing in Copala Triqui

5.1. Types of wearing verbs in Copala Triqui

Copala Triqui is like San Dionisio Ocotepec Zapotec in having multiple verbs of wearing, but shows an even more complex system, with three types of wearing verbs.

In addition to the C-verb and W-verb types, there is a third type, which I label B-verb, in which the subject must be a possessed body part. Consider the following examples:

(30) *'Nij ra'a Juán nuj guanté*
[trc] wear hand Juan glove

'Juan is wearing gloves.'

(31) *Nuu rihaan Juán scuraan.*
[trc] wear face Juan glass

'Juan is wearing gloves.'

The following table shows the four verbs used in wearing expressions in Copala Triqui:

Verb	**Type**	**Subj**	**Obj**	**Items**
'nij	B-verb	poss'd body part	clothing	shoes, gloves, sashes
nuu (1)	W-verb	wearer	clothing	shirt, huipil, pants, dress, underwear
nuu (2)	C-verb	clothing	poss'd body part	hat, mask

Verb	Type	Subj	Obj	Items
nuu (3)	B-verb	poss'd body part	clothing	glasses, ring, watch, hat
táá	C-verb	clothing	poss'd body part	mask, turban
nacutáj	C-verb	clothing	poss'd body part	diaper

Copala Triqui has only one W-verb (wearer =subject), but it is used for a wide variety of clothing.

(32) *Nuu Juán cotoó marę̲e.*
[trc] wear Juan shirt red

'Juan is wearing a red shirt.'

Nuu thus covers the same range as three different Zapotec verbs.

In contrast to Zapotec, C-verbs are rather marginal in the Copala Triqui system. The only examples are *táá* for something that sits on top of the head, but not in the usual 'hatlike' way. (Our speaker suggested a turban or cloth for this.) *Nacutáj* is used for something that encircles a person like a diaper.

5.2. Three uses of *nuu*

The verb *nuu* is remarkable in Copala Triqui, in that it shows three different subcategorizations, depending on the type of clothing and body part. Example (32) above shows its use as a W-verb. However, C-verb and B-verb patterns are also found:

(33) *Nuu navę̲ ra̲a̲ Juán .*
[trc] wear hat head Juan

'Juan is wearing a hat.' (C-verb)

(34) *Nuu rihaan Juán scura̲a̲n.*
[trc] wear face Juan glass

'Juan is wearing gloves.' (B-verb)

It is surprising that 'hat' appears with the verb *nuu* in both the C-verb and B-verb frames. Thus in addition to (33), it is also possible to say

(35) *Nuu raa Juán navẹ.*
[trc] wear head Juan hat

'Juan is wearing a hat.' (B-verb)

This is currently the *only* verb in our lexicon that shows this kind of subject/object variety, and I do not have a good explanation of this variability.

5.3. Textual examples

'Wear' must be translated in several different ways in Copala Triqui, depending on the clothing and body part. For this reason, textual examples translated from European languages often show interesting alternations in verb choice. Consider the following:

(36)... *nẹ nụụ so' yatzíj xcaan , nẹ 'nịj*
nẹ nụụ so' yatzéj xcaan nẹ 'nịj
and wear 3mSg clothes long and wear

catuụn so' merquẹ caxra' ...
catuụn so' marquẹ caxra'
waist 3mSg sash wide

'..and he was dressed in a robe reaching down to his feet and with a golden sash around his chest.' (Rev 1:13)

Because the verbs for wearing a robe and wearing a sash are of different types, it is necessary to coordinate the two clauses here, and it is not possible to have a coordinated NP object.

5.4. Variable treatment of arguments and figure-ground geometry

Copala Triqui verbs of wearing are sensitive to the geometry of wearing verbs, so the grammar reflects clothing/adornment worn in an unusual way. Consider the three different ways in which wearing a crown may be expressed in the texts:

(38) *Gaa ne curihanj Jesucristó xe' ,*
gaa ne c- urihanj Jesucristó lle'
then com leave, quit; go out Jesus Christ outside
ne 'nij raa so' cachriin tanj , ne nuu
ne 'nij raa so' cachriin tanj ne nuu
and wear head of 3mSg crown thorn and wear
so' yatzíj tintá sa' yo' a .
so' yatzéj tintá sa' yo' a
3mSg clothes purple, dark blue good 3n decl

'When Jesus came out wearing the crown of thorns and the purple robe...' (John 19:5)
Lit. 'And then Jesus went out and his head wore a crown of thorns and he wore those good purple clothes.' *B-verb*

(39) *ne táá chi' cachriin raa nij*
ne táá chi' cachriin raa nij
and be in, be on top ten crown head of pl
cúú taj raa xo' a .
cúú taj raa xo' a
bone, horn be in, be on top head of 3an decl

'…and it had ten crowns on the horns on its head.' (Rev 19:14)

(40) *ne nuu raa nij so' cachriin aga' oró míí a.*
and wear head:of 3pl crown gold decl

'and they wore golden crowns.' (Rev 4:4) (*B-verb*)

Notice that in these examples, the choice of the appropriate verb is determined by the appropriate contact relationship between the wearer and the clothing/adornment. *'nij* is appropriate in (38) because the crown is stuck into the skin. When the crowns are on the horns of a monster in (39), *táá* is the appropriate verb because the canonical contact relation between the crown and the skin does not hold. (I.e. the crown is above the head, not on it.) *Nuu* is the most normal verb for 'wear' with a crown in (40), but notice that it is a regular B-verb, in contrast to the (irregular) treatment of 'hat' seen in examples like (33) above where it is a C-verb.

5.5. Analysis of the Triqui verbs

For W-verbs in Copala Triqui, I use a structure very similar to that for Zapotec:

$$\begin{bmatrix} \text{PHON} & \text{nuu} \\ \text{CAT} & \begin{bmatrix} \text{HEAD} & \text{verb} \\ \text{ARG-ST} & \langle \text{NP: } \boxed{1}, \text{ NP: } \boxed{2} \rangle \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \text{BE-IN}\Big(\big(\text{BODY-OF}\big(\text{X } \boxed{1}\big)\big), \text{ Y } \boxed{2}\Big) \end{bmatrix} \end{bmatrix}$$

This structure accounts for examples like (32) above. Triqui differs from Zapotec in having one general 'body-of' element, while Zapotec differentiates upper, lower, and middle body areas.

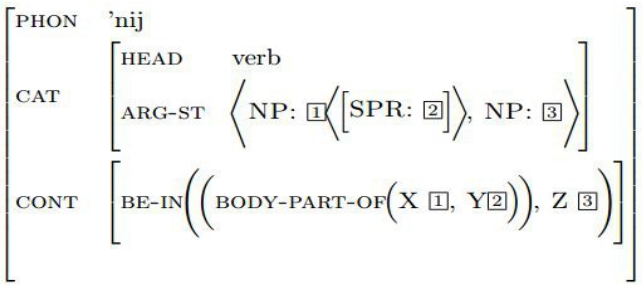
C-verbs also have an analysis similar to that for Zapotec. Consider the structure for *táá* below.

$$\begin{bmatrix} \text{PHON} & \text{táá} \\ \text{CAT} & \begin{bmatrix} \text{HEAD} & \text{verb} \\ \text{ARG-ST} & \langle \text{NP: } \boxed{1}, \text{ NP: } \boxed{2} \langle \text{SPR } \boxed{3} \rangle \rangle \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \text{BE-ON}\Big(\text{X } \boxed{1}, \big(\text{BODY-PART-OF}\big(\text{Y } \boxed{2}, \text{ Z } \boxed{3}\big)\big)\Big) \end{bmatrix} \end{bmatrix}$$

This structure accounts for examples like the following:

(41) *Táá tzej raa Juán.*
be:on cloth head Juan
'Juan is wearing a turban.'
Lit. 'A cloth is on top of Juan's head.'

Finally, B-verbs have a structure like the following:



B- and C-Predicates like *táá* 'be on top of', *'nij* 'be (tightly) inside, *nacutáj* 'encircle' are inherently asymmetrical. To account for the multiple uses of nuu, the least costly account probably posits BE-IN). *Nuu* as a B-verb shows is the most regular use (since the body part is inside the clothing). *Nuu* as a W-verb is probably a metonymy, with an implied element 'body'. I do not have a good account of *nuu* used as a C-verb with 'hat', and have to treat this as irregular.

Unlike Zapotec, Copala Triqui does not seem to have any verbs that are specialized for wearing clothing. All the verbs discussed here function as general locatives for other situations:

- *táá* 'be on top of'
- *nuu* 'be inside'
- '*nij* 'be stuck inside; be inside with tight contact'
- *nacutáj* 'be wrapped in'

6. Conclusions

The kinds of considerations that seem to be relevant to the mapping of Zapotec and Copala Triqui verbs of wearing include the following:

- Which parts of the body may stand as metonomies for the whole

- The geometry (IN, ON, AT) of the relationship between clothing and the body of the wearer
- The relative prominance of arguments for Figure/Ground assignment
- Lexical exceptions (hats in Copala Triqui; shoes in Zapotec)

Few syntactic theories are constructed in such a way as to allow reference to all the relevant factors. Theories based on a small number of semantic roles are particularly poor at accommodating these facts.

References

Bresnan, Joan and Jonni Kanerva. 1989. Locative inversion in Chichewa. *Linguistic Inquiry.* 20.1-50.

Broadwell, George A. 2001. Optimal order and pied-piping in San Dionicio Zapotec. Formal and empirical issues in Optimality Theoretic syntax, ed. by Peter Sells, 197-223. Stanford: CSLI.

Broadwell, George A. 2012. Clausal negation as raising in San Dionisio Ocotepec Zapotec. Proceedings of the Conference on Indigenous Languages of Latin America-V. Austin, TX: University of Texas, Austin. http://www.ailla.utexas.org/site/cilla5/Broadwell_CILLA_V.pdf

Culicover, Peter W. & Ray Jackendoff. 2005. Simpler syntax. Oxford: Oxford University Press.

Davis, Anthony. 2001. Linking by type in the hierarchical lexicon. Stanford: CSLI.

Dowty, David. 1991. Thematic proto-roles and argument structure. *Language* 67.547-619.

Kaufman,Terrence and John Justeson. 2009. Historical linguistics and Pre-Columbian Mesoameria. *Ancient Mesoamerica* 20.221-31.

Accessibility and word order: The case of ditransitive constructions in Persian

Pegah Faghiri
Université Sorbonne Nouvelle

Pollet Samvelian
Université Sorbonne Nouvelle

Barbara Hemforth
CNRS/Laboratoire de Linguistique Formelle (LLF)

Proceedings of the 21st International Conference on
Head-Driven Phrase Structure Grammar

University at Buffalo

Stefan Müller (Editor)

2014

Stanford, CA: CSLI Publications

pages 217–237

Faghiri, Pegah, Pollet Samvelian & Barbara Hemforth. 2014. Accessibility and word order: The case of ditransitive constructions in Persian. In Stefan Müller (ed.), *Proceedings of the 21st International Conference on Head-Driven Phrase Structure Grammar, University at Buffalo*, 217–237. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2014.12.
CC-BY 4.0

Abstract

In a most recent corpus study on Persian, Faghiri & Samvelian (2014) found a significant effect of relative length in the ordering preferences between the direct and indirect objects in the preverbal domain corresponding to "long-before-short". They furthermore showed that the position of the direct object mainly depends on its degree of determination, and put into question the broadly accepted dual view based solely on differential object marking. In this paper, we provide experimental evidence in support of these corpus findings and further propose a unified account of ordering preferences between the two objects on the basis of conceptual accessibility.

1 Introduction

Preferences in constituent ordering have often been explained by the widely accepted accessibility-based incremental model of sentence production. In this view, the linear order of constituents reflects the order in which they become available for production, as long as grammar rules do not intervene (e.g. Garrett, 1980; Bock & Levelt, 1994; Kempen & Harbusch, 2003). Constituents that become available at an earlier point in time, can occupy an earlier linear position than constituents emerging later. This view is fully compatible with the "short-before-long" principle (e.g. Wasow, 1997; Stallings et al., 1998; Arnold et al., 2000; Wasow, 2002). Short simple constituents can be processed and formulated faster and thus become available for production sooner than long and/or complex ones. However, the predictions of this model have been shown to be incompatible with the long-before-short preference observed in the preverbal domain in head-final languages such as Japanese (Hawkins, 1994; Yamashita & Chang, 2001) and Korean (Choi, 2007)

The mirror-image preference in head-initial and head-final languages was first observed by Hawkins (1994, 2004) who proposed a dependency-based distance-minimizing principle in terms of a theory of parsing efficiency. He proposed the *Early Immediate Constituent* (EIC) principle[1] to account for these seemingly contradictory preferences in head-final and head-initial languages. Yamashita & Chang (2001) provide experimental evidence for "long-before-short" preference in sentence production in Japanese. They further proposed a production-oriented account of these conflicting ordering preferences in the framework of the theory of grammatical coding (Bock & Levelt, 1994; Garrett, 1980).

[†]We would like to thank the audience at the Workshop on Understudied Languages and Syntactic Theory - HPSG21 (University at Buffalo) for their insightful comments, as well as Stefan Müller the editor of this volume. Parts of the present paper have also been presented at the Workshop on Ditransitive Constructions in a Cross-linguistic Perspective, adjacent to the SWL6 Conference at the University of Pavia in September, 2014. We would also like to thank the audience of this venue. This work is supported by a public grant funded by the French National Research Agency (ANR) as part of the "Investissements d'Avenir" program (reference: ANR-10-LABX-0083).

[1]And its more recent version, *Minimize Domains* (MiD).

Their argumentation is mainly based on the fact that in the theory of grammatical encoding decisions about word order depend on the properties of different levels in the production system. Some decisions are more conceptually driven while others depend more on form (Bock, 1982). In the conceptual arena, there is a tendency to place salient elements earlier in sentences while in the form arena, short elements are placed earlier. Long constituents have competing properties. They are semantically richer, due to the extra lexical material which makes them more salient and increases their overall accessibility in the conceptual arena. At the same time, in the form arena, the extra lexical content makes them slower to process and hence less accessible.

The authors suggest that acknowledging language-specific differences in sentence production is the key to a uniform account of word order preferences (also see Chang, 2009). They suppose that the sensitivity of sentence production system to conceptual vs. formal factors can be seen as being language-specific. The production system of Japanese, they argue, is more sensitive to conceptual factors than to form-related ones, contrary to English. This is because Japanese is a far less "rigid" language than English. Japanese has a fairly free word order and allows null pronouns. English, in contrast, has a fairly strict word order that requires all arguments to be overtly present (Yamashita & Chang, 2001, p.54). Moreover, in English Heavy-NP shift happens in the postverbal domain, where it is shown that the verb exerts strong influence, contrary to the preverbal domain (Stallings et al., 1998). These syntactic constraints presumably increase the effect of form-related factors over more conceptual ones. The authors consequently conclude that the Japanese sentence production system, more sensitive to conceptual factors, favors placing long constituents before shorter ones, while in English, more sensitive to form-related factors, placing short constituents before longer ones is favored.

As Japanese, Persian is an SOV language with a fairly free word order and null pronouns. Hence, this language share all properties singled out by Yamashita & Chang (2001) to motivate opposite length-based shifts in Japanese and English. In line with their prediction, in a corpus study on the ordering preferences between the direct (DO) and indirect (IO) objects in the preverbal domain in Persian, Faghiri & Samvelian (2014) have found a significant effect of relative length corresponding to the "long-before-short" preference. Moreover, they have shown that the relative order of the two objects depends mainly on the degree of determination of the DO, which is closely related to discourse status of the latter and hence reflects its conceptual accessibility.

The second aspect of this paper is that it undermines the broadly accepted view of the relative order of the DO and the IO in Persian, see section 3. It is generally assumed that differential object marking determines whether the DO follows or precedes the IO. Yet, the corpus data do not reflect a dichotomous behavior based on differential object marking. The study rather suggests that the position of the DO depends on its degree of determination.

In this paper, we provide experimental evidence in support of these corpus findings and propose a unified account of the relative order between the DO and the

IO in Persian on the basis of conceptual accessibility. The remainder of this paper is organized as follows. In the next section, we present an overview of Persian focusing on properties relevant for this study and in section 3, the prevailing view on the position of direct object. The corpus data of (Faghiri & Samvelian, 2014) will be summarized in section 4. In section 5, we present our experimental study, and in section 6, our unified account.

2 A Brief Overview of Persian

2.1 Word Order

The unmarked (neutral or canonical) word order in Persian is uncontroversially SOV (except for sentential complements which are strictly postverbal). Meanwhile, all phrasal categories (other than the VP), namely, NP, PP, and CP are head-initial, as illustrated in (1). Also, note in the same example that Persian does not require all arguments to be overtly realized.

(1) be ān doxtar=e javān ke diruz did-im goft (ke) emruz
to that girl=EZ[2][3] young that yesterday saw-1PL said.3SG that today
nay-āy-ad
NEG-come-3SG
'S/he said to that young girl we saw yesterday not to come today.'

While SOV is the canonical order, all other variations are possible. Although the written language is conservative with regards to the canonical SOV order, the colloquial register exhibits a fair amount of variation. In their corpus study, Faghiri & Samvelian (2014) focus on verb-final constructions. Given that the experimental data presented in this paper is a follow up on their conclusions, we have also kept the focus of this study on verb-final constructions.

2.2 Persian NPs

The relative order of objects in Persian has generally been linked to the differential object marking (DOM) (see section 2.3), which in turn is related to definiteness and/or specificity. This section provides an overview of Persian NPs in this respect.

In formal Persian there is no overt marker for definiteness; only indefiniteness is marked. Furthermore, Persian has what Corbett (2000) calls a *general number*, expressed by the singular form. This means that in Persian the number is not specified for a bare singular noun. These properties have some bearings on the

[2]Glosses follow the Leipzig Glossing Rules (www.eva.mpg.de/lingua/resources/glossing-rules.php). The following non-standard abbreviations are used for clarity: DOM = differential object marking; EZ = Ezafe.

[3]The *Ezafe*, realized as an enclitic, links the head noun to its modifiers and to the possessor NP (see Samvelian 2007).

readings of NPs. In the remainder of this section, we will discuss the following NP types: bare[4]and bare-modified nouns, indefinite/quantified NPs, and definite NPs.

2.2.1 Bare and Bare-modified Nouns

Bare nouns are non-specified for number and have a nonspecific reading, which can be generic as well as existential:

(2) gorg yek heyvān=e vahši va darande ast
wolf a animal=EZ wild and predator is
'The wolf is a wild and predator animal.'

(3) Maryam ketāb xarid
Maryam book bought
'Maryam bought a book/some books.'

Bare-modified nouns only differ from bare nouns by the presence of a (restrictive) modifier, as in (4).

(4) Maryam ketāb=e akkāsi xarid
Maryam book=EZ photography bought
'Maryam bought a photography book/some photography books.'

2.2.2 Indefinite NPs

Indefiniteness is overtly marked in Persian. It can be realized by the enclitic *=i*, as in (5-a), by the cardinal *ye(k)*[5] 'one', as in (5-b), or by the combination of these two determiners, as in (5-c).[6] Indefinite NPs can have either a specific or a nonspecific existential reading. As we will see, in the DO position the two readings will be differentiated by DOM. These NPs, contrary to bare nouns, are always specified for number.

Indefinite NPs are also formed by numerals or other indefinite quantifiers, as in (6). In this case, the noun remains in the singular form, even when the NP denotes more than one entity, and it cannot take *=i*.

(5) a. gorg=i zuze mi-kešid
wolf=INDF howl IPFV-pulled

[4]It should be noted that since definiteness is not overtly marked, bare singular nouns, that is, nouns occurring alone in their bare singular form with no (overt) determiner or quantifier, may correspond either to a definite and/or anaphoric NP, as in (i), or to a noun without any determination or quantification. By "bare noun" we only refer to the latter.

(i) xoršid dar āsemān mi-deraxš-ad
sun in sky IPFV-shine-3SG
'The sun shines in the sky.'

[5]Pronounced *ye* in colloquial speech. We will use the formal form throughout this article.

[6]The use of the enclitic alone is restricted to the formal language.

b. yek gorg zuze mi-kešid
a wolf howl IPFV-pulled

c. yek gorg=i zuze mi-kešid
a wolf=INDF howl IPFV-pulled
'A (any/certain) wolf was howling.'

(6) čand(=tā)/se(=tā) gorg zuze mi-kešid-and
few(=CLF)/three(=CLF) wolf howl IPFV-pulled-3PL
'A few/three wolves were howling.'

2.2.3 Definite NPs

Definite NPs can either be formed by different definite determiners, like demonstratives, or by no overt determiner, as in (7).[7] Furthermore, bare plural nouns[8] generally trigger a definite reading, as in (8). Note, however, that the plural marking is not incompatible with the indefinite determination *=i* or *yek*, as in (9) (for a discussion of plural marking and definiteness, see Ghomeshi 2003).

(7) (in) šiše emruz šekast
(this) glass today broke
'This/the glass broke today.'

(8) šiše-hā emruz šekast-and
glass-PL today broke-3PL
'The (*Some) glasses broke today.'

(9) yek ketāb-hā=i heyn=e asbābkeši gom šod-and
a book-PL=INDF during=EZ move lost became-3PL
'Some (of the) books get lost during the move.'

2.3 Differential Object Marking

Persian displays differential object marking (DOM),[9] realized by the enclitic *=rā*.[10] Definite and/or specific direct objects are necessarily *rā*-marked. Consequently, non-*rā*-marked direct objects receive an indefinite nonspecific reading, as in (10). DOM is not incompatible with the indefinite determination, as in (11). An indefinite NP like *ketāb=i* when *rā*-marked will receive a specific reading.

[7] It should be noted that colloquial speech displays a definite suffix, realized as *-(h)e*, which marks a noun as being discourse-given or anaphoric, for example, *gorbe-he* 'the cat'.

[8] Persian disposes of several nominal plural suffixes, among them the suffix *-(h)ā* is universal and can systematically be added to any noun to form a plural (for a review of the nominal plural marking see Lazard et al. 2006 and Faghiri 2010, among others).

[9] This designation coined by Bossong (1985) denotes the property of some languages with overt case-marking of direct objects to mark some objects, but not others, depending on semantic and pragmatic features of the object; see also Aissen (2003).

[10] Realized as *=(r)o* in colloquial speech. We use the formal form throughout this paper for the ease of reading.

(10) Maryam ketāb=rā xarid *vs.* Maryam ketāb xarid
Maryam book=DOM bought Maryam book bought
'Maryam bought the book.' *vs.* 'Maryam bought a book/some books.'

(11) Maryam ketāb=i=rā xarid *vs.* Maryam ketāb=i xarid
Maryam book=INDEF=DOM bought Maryam book=INDEF bought
'Maryam bought a (specific) book.' *vs.* 'Maryam bought a book.'

Nevertheless, *rā*-marking cannot be accounted for on the basis of definiteness and specificity only, as illustrated by (12). Furthermore, the use of the enclitic *=rā* is not limited to DOM. It is also used to mark discourse prominence for other non-subject functions, as in (13). Meanwhile, a more detailed discussion is beyond the scope of the present study (for further discussions see Lazard 1982; Meunier & Samvelian 1997; Dabir-Moghaddam 1992; Roberts et al. 2009, among others).

(12) ketāb=rā mi-xān-and
book=DOM IPFV-read-3PL
'A book, one reads (it).' or 'A book is meant to be read.'

(13) emruz=rā dars mi-xān-am
today=DOM lesson IPFV-read-1SG
'As for today, I (will) study.'

2.4 Complex Predicates

Persian has a limited number of simplex verbs, around 250, half of which are currently used by the speech community. The verbal lexicon mainly consists of syntactic combinations, called "complex predicates", also known as Compound Verbs or Light Verb Constructions, including a verb and a non-verbal element, for example, a noun, as in *bāzi kardan* 'to play' (lit. 'play do'), an adjective, as in *derāz kešidan* 'to lay down' (lit. 'long pull'), a particle, as in *bar dāštan* 'to take' (lit. 'PARTICLE have'), or a prepositional phrase, as in *az dast dādan* 'to loose' (lit. 'of hand give'). New "verbal concepts" are regularly coined as complex predicates rather than simplex verbs (see Samvelian 2012; Samvelian & Faghiri 2013, 2014, among many others).

Although, Persian complex predicates are multiword expressions and thus display some lexical properties such as lexicalization, they display all properties of syntactic combinations, including some degree of semantic compositionality. Hence, as Samvelian (2001, 2012) extensively argues, it is impossible to establish a clearcut distinction between (prep-)noun-verb complex predicates and "ordinary" object-verb combinations. In other words, the differentiation is better reflected by a continuum from highly lexicalized complex predicates to ordinary complement-verb combinations rather than a categorical distinction. Following this observation and given the impossibility of the task, Faghiri & Samvelian (2014) did not attempt to apply any filter to exclude complex predicates from their dataset. In our experimental study, we only included combinations that could hardly be qualified as complex

predicates and would safely be located on the other extremity of the continuum.

3 The Position of the Direct Object

The relative order of the DO and the IO in Persian is generally assumed to depend on *rā*-marking. It is broadly admitted that in a neutral word order *rā*-marked DOs precede the IO while non-*rā*-marked DOs are adjacent to the verb (Mahootian, 1997; **?**; Roberts et al., 2009, among others). Hereafter, we refer to this hypothesis as the DOM criterion.

Several theoretical studies, mainly in the generative framework, further argue for the existence of two different syntactic positions for the DO depending on its markedness or more precisely its specificity (Ghomeshi, 1997; Karimi, 2003; Ganjavi, 2007, among others). To give an example, (14) illustrates the two positions assumed by Karimi (2003, p.105), one of the most frequently cited paper among the above-mentioned. She, furthermore, assumes that a nonspecific or in other words non-*rā*-marked DO can be separated from the verb, that is, can undergo scrambling, only if it has a contrastive focus. The scrambling of specific objects is less constrained, since they can additionally be topicalized.[11] The examples provided by (Karimi, 2003, pp.91–92) to illustrate these claims are given in (15).

(14) a. $[_{VP}\ DP_{[+Specific]}\ [_{V'}\ PP\ V]]$
b. $[_{VP}\ [_{V'}\ PP\ [_{V'}\ DP_{[-Specific]}\ V]]]$

(15) a. Kimea aqlab barā mā še'r mi-xun-e
Kimea often for us poem IPFV-read-3SG
'It is often the case that Kimea reads poetry for us.'
b. Kimea aqlab barā mā ye še'r az Hafez mi-xun-e
Kimea often for us a poem from Hafez IPFV-read-3SG
'It is often the case that Kimea reads a poem by Hafez for us.'
c. Kimea aqlab hame=ye še'r-ā=ye tāza=š=**ro** barā mā
Kimea often all=EZ poem-PL=EZ new=3SG=DOM for us
mi-xun-e
IPFV-read-3SG
'It is often the case that Kimea reads all her new poems for us.'
d. Kimea aqlab ye še'r az Hafez=**ro** barā mā mi-xun-e
Kimea often a poem from Hafez=DOM for us IPFV-read-3SG
'It is often the case that Kimea reads a (particular) poem by Hafez for us.'
e. Kimea aqlab (ye) ketāb=e dāstān barā bačče-hām mi-xun-e
Kimea often a book=EZ story for child-PL IPFV-read-3SG
'Kimea often reads (a) STORY-BOOK for children (rather than a poetry book).'

[11] Karimi (2003, pp.106–111) assumes that discourse functions trigger movement in Persian and the landing site of a scrambled object is the specifier of a functional head, such as Topic or Focus.

Grammarians have also formulated generalizations about the canonical position of the DO, which are mostly in accordance with the DOM criterion. However, some additionally establish a distinction between unmarked DOs, depending upon the presence of the indefinite marked *-i*. Givi Ahmadi & Anvari (1995, p.305), for instance, state that *rā*-marked DOs should precede the IO, non-*rā*marked DOs should follow the IO, and *i*-marked (non *rā*-marked) indefinite DOs can either follow or precede the IO.

Faghiri & Samvelian (2014) have conducted the first corpus-based study to investigate the ordering preferences between the DO and the IO. Their study undermines the DOM criterion. Namely, in their corpus data, indefinite (non-*rā*-marked) DOs are in majority non adjacent to the verb, and hence group with *rā*-marked DOs and not with bare and bare-modified DOs.

The experimental study we present in this paper, see section 5, is a follow up on the corpus findings of Faghiri & Samvelian (2014). In the next section, we briefly present their data and results.

4 Summary of Corpus Results

4.1 The Dataset

The study is conducted on the Bijankhan corpus, a corpus collected from daily news and common texts, in particular, the newspaper *Hamshahri*, of about 2.6 million tokens, manually tagged for part-of-speech information.[12]

To constitute their dataset, the authors have selected the potentially ditransitive verbs of the corpus (122 verb types), corresponding to 42,550 tokens and have identified relevant sentences, that is, sentences matching either of NP PP V or PP NP V patterns (without take into consideration the preceding constituents of the sentence) in two separate samples : 1) a random sample of 2000 tokens out of this subset, and 2) a sample including all occurrences of two typically ditransitive low frequency verbs of the corpus (*rixtan* 'to pour' and *ferestādan* 'to send'; 219 and 254 tokens, respectively), as well as a random sample out of all occurrences of two high frequency typically ditransitive verbs (*gereftan* 'to give' and *dādan* 'to take'; 10494 and 6849 tokens, respectively). This second dataset contains 905 tokens.

4.2 The DOM Criterion

The data is annotated for the DO type according to a fine-grained classification based on the degree of determination of the NP. Marked and bare DOs correspond, respectively, to the highest and the lowest degree of determination for an NP in the DO position in Persian. For intermediate cases, that is, non-bare non-*rā*-marked DOs, the authors have separated determined NPs, that is, quantified or indefinite NPs, from non-determined NPs, that is, bare-modified NPs. Recall that the latter

[12]http://ece.ut.ac.ir/dbrg/bijankhan/

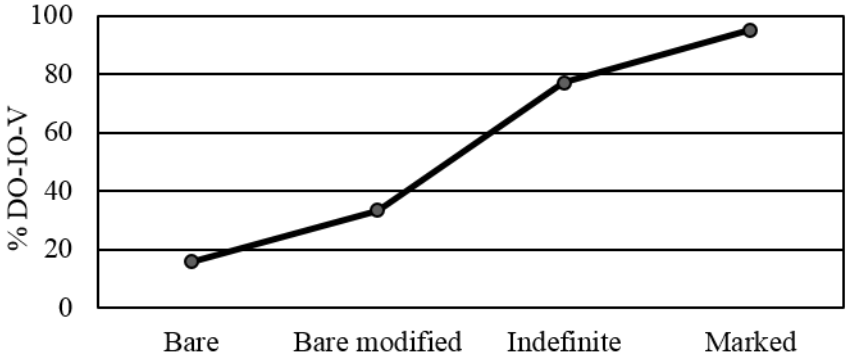


Figure 1: Number of DO-IO-V orders in % by DO-type

only differ from bare nouns by the presence of a modifier. Four DO types are hence defined: Bare, Bare-modified, Indefinite, and Marked.

The ordering preferences observed for marked DOs, on the one hand, and bare and bare-modified DOs, on the other hand, conform to the DOM criterion. Marked DOs prefer the DO-IO-V word order. Bare and bare-modified DOs prefer the IO-DO-V word order. Indefinite DOs, however, contrary to what is expected from the DOM criterion, prefer the DO-IO-V word order. Indeed, being non-*rā*-marked, it is expected for these DOs to group with bare and bare-modified DOs and prefer the IO-DO-V word order. Upon this observation, Faghiri & Samvelian (2014) call into question the DOM criterion and propose an account that subordinates the position of the DO to its degree of determination rather than to its markedness. This account, they claim, has furthermore the benefit of capturing the variation in the strength of the preference.

The degree of determination is, indeed, a continuum, and consequently, this account does not predict ordering preferences on a dichotomous basis as it is the case with the DOM criterion. Faghiri & Samvelian (2014) formulate their account as following: The more a DO is determined, the more it is likely to be placed leftward in the sentence and separated from the verb. Or, vice versa, the less a DO is determined, the more likely it is to be placed adjacent to the verb, see Figure 1. Put this way, it is expectable for DOs located in the middle of the continuum to show more variability than the ones located on the two extremities.

4.3 The Relative Length

Faghiri & Samvelian (2014) have also investigated the question of the relative length via their corpus data. They argue that this factor is not relevant for all DOs. On the one hand, marked DOs prefer the DO-IO order regardless of the relative length, and on the other hand, relative length is meaningless for bare DOs, since these DOs are by definition smaller than the IO in number of words. However,

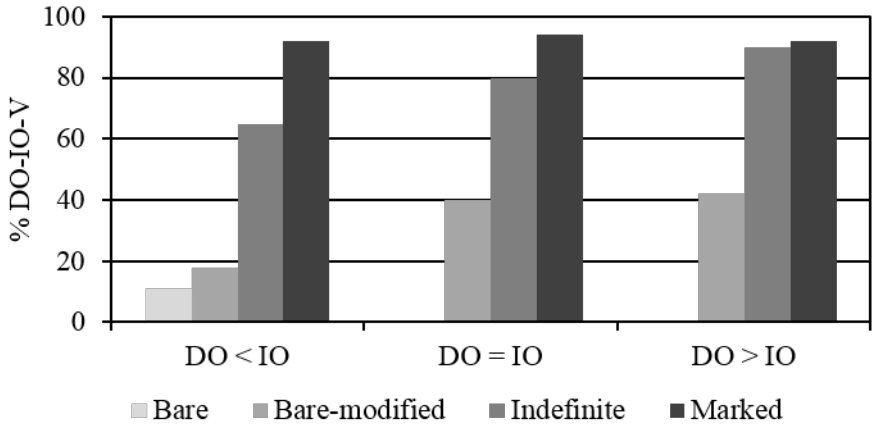


Figure 2: Number of DO-IO-V orders in % by DO-type and Relative Length

the relative length is a relevant factor in the case of indefinite and bare-modified DOs, see Figure 2. Accordingly, the authors have performed mixed-effect logistic regression modeling, on a subset of their dataset excluding marked and bare DOs.

As expected, DO-TYPE has a significant effect ($p < 0.001$). Bare-modified type favors the IO-DO order while Indefinite type favors the inverse, confirming thus that indefinite DOs do not behave like other non-*rā*-marked DOs. In line with our earlier assumptions, REL-LEN also has a significant effect ($p < 0.01$) corresponding to the "long-before-short" preference. The authors, hence, claim that Persian is another verb-final language that like Japanese displays an effect of relative length corresponding to "long-before-short". Apart from being verb-final, Persian shares all other properties singled out by Yamashita & Chang (2001) in their account of "long-before-short" in Japanese. Persian, like Japanese and contrary to English, displays a fairly free word order and does not require all arguments to be overtly realized. Consequently, following Yamashita & Chang (2001), assuming that longer constituents are lexically richer and hence more salient and conceptually accessible, Faghiri & Samvelian (2014) attribute this ordering preference in Persian to the more important influence of conceptual factors, comparing to form-related ones, in ordering preferences in the preverbal domain in this language.[13]

They note that this preference can be integrated in the continuum established on the basis of the degree of determination of the DO, given that it allows to capture some of the variation observed for the DOs in the middle of the hierarchy. In the case of intermediate DOs, lexical richness contributes to the accessibility of the DO and hence a relatively more salient DO would be located higher in the continuum and therefore is more likely to be separated from the verb, whereas on the two extremities, that is, in the case of marked and bare DOs, the nature of the DO

[13]Note that since Persian is not a head-final language like Japanese, EIC/MiD (Hawkins, 1994, 2004) fails to provide adequate predictions for Persian, as illustrated by Faghiri & Samvelian (2014).

determines its preferred position regardless of relative length.

5 Experimental study

The conclusions of the corpus study conducted by Faghiri & Samvelian (2014) contradicts the broadly accepted view of the relative order between the DO and the IO in Persian, that is, the DOM criterion. According to this view in an unmarked (neutral) word order, *rā*-marked DOs and non-*rā*-marked DOs precede the IO. Lambrecht (1996) highlights that unmarked word orders are not specified for a particular discourse function and can be used in any information structure. Therefore, having more distributional freedom, a neutral word order is the word order that has a greater overall frequency of occurrence. Yet, in the case of indefinite (non-*rā*-marked) DOs, for which the neutral word order is suppose to be IO-DO-V, the inverse order is significantly more frequent in Faghiri & Samvelian's (2014) corpus data.

One could argue that this discrepancy may be of stylistic nature due to the fact that the data is extracted from a journalistic corpus. Therefore, we have run a web-based questionnaire to study the ordering preference of indefinite non-*rā*-marked DOs in a controlled experiment. We designed a sentence completion experiment in order to obtain the preference of speakers for alternative word orders. In this experiment, besides the choice of the order between the two arguments of the verb, the task also required to make a choice between two given possibilities - formally identical but lexically different - for the theme argument. The idea was to bring the attention of the participants to the meaning of the sentence rather than to its form to avoid strategic responses. As mentioned previously, this experiment only included indefinite DOs. The relative length and givenness of the two objects were manipulated following a 2x2 design.

Givenness (or newness) in discourse, that is, the information status, is one highly discussed factor in constituent ordering preferences (e.g. Gundel, 1988; Arnold et al., 2000; Bresnan et al., 2007). In the corpus study of Faghiri & Samvelian (2014) the data was not annotated for the information status and thus the effect of the information structure could not be tested properly.[14] Indefinite DOs are by definition discourse new. As we have mentioned in section 2.3, a discourse given NP in the DO position in Persian is always *rā*-marked. Therefore, it is not possible to manipulate the givenness of an indefinite DO (on a dual discourse-given vs. discourse-new basis). Thus, in this experiment, we manipulated the relative givenness by manipulating the information status of the IO, with two conditions : discourse-new vs. discourse given. In the discourse-given condition, the IO is mentioned in the preamble, whereas in the discourse-new condition there is no

[14]Indeed, DO types are defined on the basis of the degree of determination of the NP and reflect the information status of the DO, more precisely, its referential givenness (see Gundel et al., 1993), to some extent. Yet, the corpus study did not allow to investigate the proper effect of the information structure, that is, independent of grammatical roles.

mention of the IO previously.

With regards to the relative length between the DO and the IO two conditions were defined : DO $>$ IO and DO $<$ IO. We manipulated the length by attaching a modifier ranging from 5 to 10 syllables, with an average of 7 syllables. In the case of the DO, we added adjectives, and in the case of the IO, we added a relative clause. See table 5.1.3 an example of a target item used in the experiment.

5.1 Method

5.1.1 Material

We constructed twenty experimental items in four conditions each. Every item contained a preamble and a target sentence constructed with a ditransitive verb following a $DO_{[-animate]}$ - $IO_{[+human]}$ pattern. We used seven semantically different ditransitive verbs (*baxšidan* 'to donate', *dādan* 'to give', *ferestādan* 'to send', *foruxtan* 'to sell', *gereftan* 'to take', *xaridan* 'to buy', and *xorāndan* 'to feed') implying a variety of prepositions : *be* 'to', *az* 'from', and *barāye* 'for'.[15] For each target sentence two formally identical versions, that is, having the same length and construction but different lexically, of the theme argument were prepared. The experimental items were combined with thirty fillers. Four lists were created according to a Latin Square design.

5.1.2 Procedure

The questionnaire was conducted via the Internet, on the *Ibex-Farm* platform. The participants were asked to take part in the questionnaire only if they had 15 minutes to spare, without doing anything else in the meantime. They were instructed to read the preamble and the three phrases which followed, and construct a sentence, as natural as possible, and fill in the blanks accordingly using drag-and-drop or copy-paste. The instructions indicated that their reaction time was counted in order to put them under some time pressure. Two training items followed the instructions. Figure 3 provides an example of an item on screen. Note that the three phrases were presented in a randomized order.

5.1.3 Participants

33 native speakers of Persian volunteered to complete the web-based (anonymous) questionnaire. They were reached through social networks.

[15] Note that we were limited in our choice of the verb, given the limited number of simplex verbs in Persian, see section 2.4.

Preamble	**Given IO**	
kuler az kār oftāde bud va moštarihā ebrāze nārezāyati mikardand. ābdārči vaqti e'terāzhā be owj resid... 'The air-conditioner had stopped working and the customers were complaining. When protests reached a peak, the janitor...'		
Preamble	**New IO**	
kuler az kār oftāde bud va havā hamintor garmtar mišod. ābdārči vaqti garmā be owj resid... 'The air-conditioner had stopped working and the weather kept getting warmer. When the heat reached a peak, the janitor...'		
Target sentence : .... *dād.* 'gave'		
	DO < IO	
IO	**DO1**	**DO2**
be moštarihā ke az garmā kalāfe budand 'to customers frustrated from the heat'	*yek livān šarbat* 'a glass of syrup'	*yek qāč hendevāne* 'a slice of watermelon'
	DO > IO	
IO	**DO1**	**DO2**
be moštarihā 'to customers'	*yek livān šarbate sekanjebine tagari* 'a glass of icy mint syrup'	*yek qāč hendevāneye bedune hasteye xonak* 'a slice of fresh seedless watermelon'

Table 1: Example of an item of the experiment

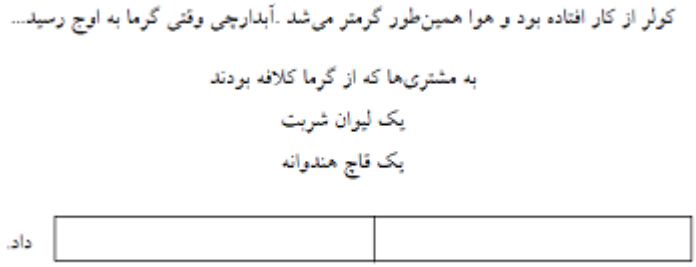


Figure 3: Example of an item on the screen

5.2 Results

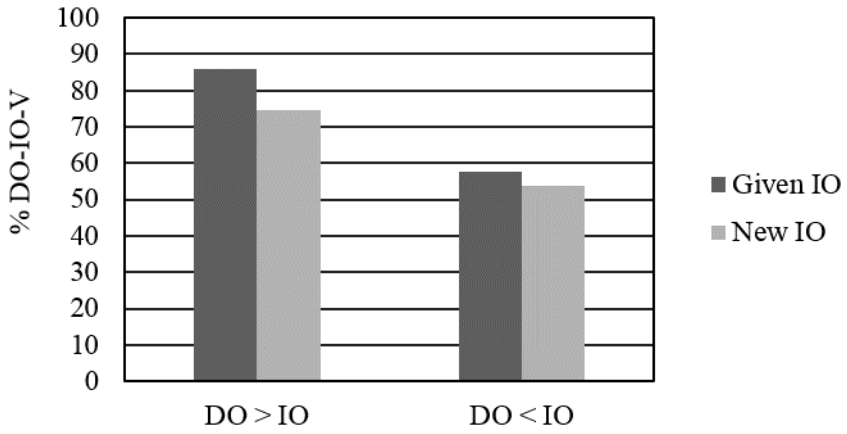


Figure 4: Number of DO-IO-V orders in % by Relative Length and Givenness

The mean proportion of the DO-IO-V order in participants sentences by relative length and givenness is presented in Figure 4. The data confirm an overall preference (68%; χ^2=85.8242, df=1, p<2.2e-16) for the DO-IO-V order. We observe that when the DO is longer that the IO, the preference for the DO-IO-V order is much more stronger than when the DO is smaller than the DO (80.3% vs. 55.7%; χ^2=44.5857, df=1, p=2.435e-11). Surprisingly, when the IO is given, the preference for the DO-IO-V order is stronger than when the IO is new; the difference however is relatively small (71.8% vs. 64.2%; χ^2=4.0127, df=1, p=0.04516). Note that the DO-IO-V order remains above average in all conditions.

To analyses the results statistically, we fitted a mixed-effect logistic model (Agresti, 2007), predicting the relative order between the DO and the IO by the two experimental factors, givenness of the IO and relative length, as fixed effects.

Random effects:				
	Variance	Std. Dev.		
SUBJ (Intercept)	0.7509	0.8666		
ITEM (Intercept)	0.2390	0.4889		
VERB (Intercept)	0.1418	0.3766		
Number of obs: 660, groups: SUBJ, 33; ITEM, 20; VERB, 7				
Fixed effects:				
	Estimate	Std. Error	z value	Pr($>\|z\|$)
(Intercept)	0.3909	0.2909	1.344	0.1790
GIVENNESS=IO-NEW	-0.1585	0.24205	-0.655	0.5124
REL-LEN=DO>IO	1.7499	0.2991	5.851	4.9e-09
IO-NEW:DO>IO	-0.7441	0.3916	-1.900	0.0574
Correlation of Fixed Effects:				
	(Intercept)	IO-NEW	DO>IO	
IO-NEW	-0.415			
DO>IO	-0.338	0.402		
IO-NEW:DO>IO	0.256	-0.616	-0.751	
N.B. Success corresponds to ORDER=DO-IO				

Table 2: Results of logistic mixed-effect model

We included subject item and verb as random intercepts,[16] in order to account for inter-subject and inter-item variation in the data, as well as the lexical bias of the verbal lemma. The results of the model are given in Table 5.2.[17]

The analysis shows that relative length has a significant effect ($p < 0.001$) corresponding to the "long-before-short" preference. Givenness of the IO, however, does not turn out to have a significant main effect ($p > 0.5$). There is, nevertheless, a marginal interaction between the two variables ($p < 0.1$). More experiments will be needed in order to pin down this interaction. Note that it may not be surprising that the givenness of the IO does not play a significant role in determining the relative order between the two objects, given the fact that on the hierarchy of the grammatical roles Keenan & Comrie (1977) the IO occupy a lower position than the DO.

[16]The maximal model also included main and interaction slopes for all random effects, but the analysis of variance (ANOVA) showed no significant advantage comparing to the simpler model we present here (χ^2=24.88, df=27, p=0.5812).

[17]These results are fully consistent with (actually nearly identical to) a former experiment run with 60 subjects (no overlaps), in which, due to a script error, each subject had completed a selection of 16 items out of 20.

6 A Unified Account

We propose a unified account of the relative order between the DO and the IO in Persian, on the basis of the conceptual accessibility, grounded in insights provided by Yamashita & Chang (2001) in their account of the "long-before-short" preference in Japanese. Recall that according to incremental models of sentence production, the linear order of constituents is related to their accessibility, in the formal as well as the conceptual arena. In the conceptual arena there is a tendency to place more conceptually accessible constituents, that is, discourse-given, salient, animate, etc., earlier in the sentence (e.g. Bock, 1982; Kempen & Harbusch, 2003)

The empirical evidence presented in this paper shows that the relative order between the DO and the IO in the preverbal domain in Persian depends on two factors. In the first place, the degree of determination of the DO and, in the second place, its length. Indeed, these two independent factors contribute both to the conceptual accessibility of the DO.

1. For an NP in the DO position in Persian, one can safely assume that *rā*-markedness, which corresponds to the highest degree of determination, corresponds also to the highest degree of discourse givenness (and/or prominence). The lowest degree can also be safely assumed to correspond to bare (and bare-modified for that matter) DOs. Indefinite DOs occupy an intermediate position. Hence, the continuum established on the basis of the degree of determination corresponds to a hierarchy of discourse givenness and/or prominence. A factor that contributes to the conceptual accessibility of a constituent.

2. As argued by Yamashita & Chang (2001), longer constituents, containing extra lexical material, are semantically richer, and hence are more salient, that is, more (conceptually) accessible, than shorter ones.

Now let us take a closer look into the ordering preferences for different types of DO with respect to length:

- Marked DOs, uncontroversially, strongly prefer to the DO-IO order, and bare DOs strongly prefer the IO-DO order, regardless of length.

- Indefinite DOs, our study has confirmed, present a moderate preference for the DO-IO order, which increases significantly for longer DOs.

- Bare-modified DOs can be viewed as longer counterparts of bare DOs. They are lexically richer and therefore, even though they display the same degree of discourse givenness as bare DOs, are more salient than the latter and hence conceptually more accessible. The corpus data have showed a rather moderate preference of these DOs for the IO-DO order (comparing to bare DOs), which decreases for longer DOs, see Figure 2 above.

Consequently, to account for these ordering preference as a whole, we suggest to establish a continuum on the basis of the increasing degree of conceptual accessibility - combining discourse givenness/prominence and lexical salience - of the DO, from the strong preference of bare DOs for the IO-DO-V order to the strong preference of *rā*-marked DOs for the DO-IO-V order.

7 Conclusion

In this paper, we have presented experimental data on the relative order between the DO and the IO in Persian to follow up on the corpus study conducted by Faghiri & Samvelian (2014). These findings have a twofold interest, one *vis-à-vis* the existing hypothesis for Persian, and second, with respect to the effect of the relative length cross-linguistically.

1. The position of the DO does not exclusively depend on its markedness. Ordering preferences of verbal complements in ditransitive constructions reflect a continuum on the basis of the degree of determination of the DO rather than a categorical behavior depending on its markedness. This fact contradicts a dual syntactic position hypothesis for the DO, as claimed by some theoretical studies (e.g. Karimi, 2003).

2. Ordering preferences of verbal complements in ditransitive constructions show a significant effect of relative length corresponding to the "long-before-short" principle. Supporting mirror-image preferences in OV and VO languages. Note that, as Faghiri & Samvelian (2014) argued, Persian data is not predicted by the EIC/MiD principle (Hawkins, 1994, 2004).

Reinforcing the hypotheses formulated in Faghiri & Samvelian (2014), we have proposed a unified account of the position of the DO based on its conceptual accessibility. We have combined discourse givenness (and/or prominence), on the one hand, and lexical richness on the other hand, both assumed to contribute to the conceptual accessibility of a constituent. Thus, we have provided an empirically valid account that not only predicts the preferred position of different types of DOs, but also accounts for the variation in the strength of these preferences.

The experimental study we presented in this paper is the first of a series of experiments we are undertaking to study ordering preferences between the DO and the IO in the preverbal domain in Persian. Namely, similar experiments for bare-modified DOs and experiments to test the gradual nature of ordering preferences depending on the DO type, are underway.

One open issue remaining is the role of the subject. As a matter of fact, in the literature that discuss word order variations in the preverbal domain in Persian, the position of the subject is rarely discussed, most probably because the neutral SOV word order is uncontroversial. The relative order of objects, however, is a long lasting debate, mainly because of the DOM in Persian. Nevertheless, it is

crucial to include the subject in studies on ordering preferences in the preverbal domain in Persian, as it is also the case with Yamashita & Chang (2001). More precisely, it is interesting to see to what extent a highly (conceptually) accessible DO, that is, for example, an animate *rā*-marked DO, is likely to win the competition for the initial position of the sentence over the subject. In future research, we are taking on ordering preferences between the subject and the direct object in transitive constructions, as well as between all the three constituents in ditransitive constructions.

References

Agresti, Alan. 2007. *An introduction to categorical data analysis*. Wiley-Interscience.

Aissen, Judith. 2003. Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory* 21(3). 435–483.

Arnold, Jennifer E., Thomas Wasow, Anthony Losongco & Ryan Ginstrom. 2000. Heaviness vs. newness: The effects of complexity and information structure on constituent ordering. *Language* 76(1). 28–55.

Bock, J. Kathryn. 1982. Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review* 89(1). 1–47.

Bock, J. Kathryn & Willem Levelt. 1994. Language production: Grammatical encoding. In Morton A. Gernsbacher (ed.), *Handbook of psycholinguistics*, 945–984. New York: Academic Press.

Bossong, Georg. 1985. *Empirische Universalienforschung: differentielle Objektmarkierung in den neuiranischen Sprachen*. Tübingen : Gunter Narr Verlag.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina & R. Harald Baayen. 2007. Predicting the dative alternation. In Boume Gerlof, Irene Kraemer & Joost Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94.

Chang, Franklin. 2009. Learning to order words: A connectionist model of heavy {NP} shift and accessibility effects in japanese and english. *Journal of Memory and Language* 61(3). 374 – 397.

Choi, Hye-Won. 2007. Length and order: A corpus study of Korean dative-accusative construction. *Discourse and Cognition* 14(3). 207–227.

Corbett, Greville G. 2000. *Number*. Cambridge University Press.

Dabir-Moghaddam, Mohammad. 1992. On the (in) dependence of syntax and pragmatics: Evidence from the postposition *-ra* in Persian. In *Cooperating with written texts: The pragmatics and comprehension of written texts*, 549–573. Mouton de Gruyter.

Faghiri, Pegah. 2010. *La morphologie du pluriel nominal du persan d'après la théorie whole word morphology*. Université de Montréal MA thesis.

Faghiri, Pegah & Pollet Samvelian. 2014. Constituent ordering in Persian and the weight factor. In Christopher Pinon (ed.), *Empirical issues in syntax and semantics 10 (EISS10)*, In press.

Ganjavi, Shadi. 2007. *Direct objects in persian*. University of Southern California.

Garrett, Merrill F. 1980. Levels of processing in sentence production. *Language production* 1. 177–220.

Ghomeshi, Jila. 1997. Topics in persian vps. *Lingua* 102(2—3). 133–167.

Ghomeshi, Jila. 2003. Plural marking, indefiniteness, and the noun phrase. *Studia linguistica* 57(2). 47–74.

Givi Ahmadi, Hassan & Hassan Anvari. 1995. *Dastur zabāne fārsi [Persian grammar]*. Mo'assese farhangi Fātemi.

Gundel, Jeanette K. 1988. Universals of topic-comment structure. In M. Hammond, E.A. Moravcsik & J.R. Wirth (eds.), *Studies in syntactic typology*, 209–239. John Benjamins.

Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69(2). 274–307.

Hawkins, John A. 1994. *A performance theory of order and constituency*. Cambridge University Press.

Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford University Press.

Karimi, Simin. 2003. On object positions, specificity, and scrambling in Persian. In Simin Karimi (ed.), *Word order and scrambling*, 91–124. Blackwell Publishing.

Keenan, Edward L & Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic inquiry* 8(1). 63–99.

Kempen, Gerard & Karin Harbusch. 2003. Word order scrambling as a consequence of incremental sentence production. In Holden Härtl & Heike Tappe (eds.), *Mediating between concepts and grammar*, 141–64. Mouton De Gruyter.

Lambrecht, Knud. 1996. *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge University Press.

Lazard, Gilbert. 1982. Le morphème *rā* en persan et les relations actancielles. *Bulletin de la Société de Linguistique de Paris* 77(1). 177–208.

Lazard, Gilbert, Yann Richard, Rokhsareh Hechmati & Pollet Samvelian. 2006. *Grammaire du persan contemporain*. Institut français de recherche en Iran.

Mahootian, Shahrzad. 1997. *Persian*. New York: Routledge.

Meunier, Annie & Pollet Samvelian. 1997. La postposition *rā* en persan : son rôle dans la détermination et sa fonction discursive. *Cahiers de Grammaire* 25. 187–232.

Roberts, John R, Behrooz Barjasteh Delforooz & Carina Jahani. 2009. *A study of persian discourse structure*. Acta Universitatis Upsaliensis.

Samvelian, Pollet. 2001. Le statut syntaxique des objets nus en persan. *Bulletin de la Société de Linguistique de Paris* 96(1). 349–388.

Samvelian, Pollet. 2007. A (phrasal) affix analysis of the Persian ezafe. *Journal of Linguistics* 43(3). 605–645.

Samvelian, Pollet. 2012. *Grammaire des prédicats complexes : les constructions nom-verbe*. Hermès–Lavoisier.

Samvelian, Pollet & Pegah Faghiri. 2013. Introducing PersPred, a syntactic and semantic database for Persian complex predicates. In *Proceedings of the workshop on multiword expressions*, 11–20. Atlanta, Georgia, USA.

Samvelian, Pollet & Pegah Faghiri. 2014. Persian complex predicates: How compositional are they? *Semantics-Syntax Interface* 1(1). 43–74.

Stallings, Lynne M., Padraig G. O'seaghdha & Maryellen C. MacDonald. 1998. Phrasal ordering constraints in sentence production: Phrase length and verb disposition in heavy-NP shift. *Journtal of Memory and Language* 39(3). 392 – 417.

Wasow, Thomas. 1997. Remarks on grammatical weight. *Language Variation and Change* 9(01). 81–105.

Wasow, Thomas. 2002. *Postverbal behavior*. CSLI.

Yamashita, Hiroko & Franklin Chang. 2001. "Long before short" preference in the production of a head-final language. *Cognition* 81(2). B45–B55.

Predication and NP structure in an omnipredicative language: The case of Khoekhoe

Michael Hahn

University of Tübingen

Proceedings of the 21st International Conference on
Head-Driven Phrase Structure Grammar

University at Buffalo

Stefan Müller (Editor)

2014

Stanford, CA: CSLI Publications

pages 238–258

Hahn, Michael. 2014. Predication and NP structure in an omnipredicative language: The case of Khoekhoe. In Stefan Müller (ed.), *Proceedings of the 21st International Conference on Head-Driven Phrase Structure Grammar, University at Buffalo*, 238–258. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2014.13.
 CC-BY 4.0

Abstract

We examine noun phrases and predication in Khoekhoe, a Central Khoisan language, arguing that members of all open word classes can function equally and without derivation as predicates, and that predicative use is primary and referential use is derived syntactically by relativization. We then present a formal HPSG analysis, in which members of all open word classes enter the syntax as predicates and in which all argument NPs are derived in a uniform manner as projections of pronominal elements, modified by relative clauses, building on Sag's (1997) analysis of English relative clauses. We will then argue that, additionally, DPs may project directly to clauses, yielding a second predication structure.

1 The Data

Launey (1994, 2002) has proposed the concept of *omnipredicativity*, describing languages where members of all major open word classes can function equally and without derivation as predicates, and in which the predicative use is primary and referential use is derived syntactically by relativization. Omnipredicativity is different from the lack of word classes: in Classical Nahuatl, the language studied by Launey, nouns and verbs are clearly distinguished by their morphological properties, but on the syntactic level, nouns and verbs can both be used predicatively and referentially in the same ways.

In this paper, we examine noun phrases and predication in Khoekhoe, a Khoisan language spoken in Namibia and South Africa. We show that it exhibits features typical of 'omnipredicative' languages and present a formal HPSG analysis, in which members of all open word classes enter the syntax as predicates and in which all argument NPs are derived in a uniform manner as projections of pronominal elements, modified by relative clauses. Despite the radical differences between Khoekhoe and European languages in the relevant areas, our analysis will crucially build on standard components of HPSG analyses. No special rule licensing predicative use of nouns is required, and referential use will be derived based on Sag's (1997) analysis of English relative clauses.

In Khoekhoe, there are three open word classes: verbs, nouns, and adjectives. They are clearly distinguished in morphology by the derivation morphemes applicable to them: only verbs and adjectives allow valencey-changing suffixes (passive, reflexive, reciprocal, applicative, pronominal object markers). In the realm of syntax, adjectives and nouns have a fixed order within NPs: adjectives can modify

[†]I want to thank Stefan Müller, Jean-Pierre Koenig, Ray Jackendoff, Rui Chaves, Berthold Crysmann, two anonymous reviewers, and the audience of HPSG 2014 for their helpful and inspiring comments. I am also indebted to Wilfrid Haacke for kindly providing access to two of his studies and for inspiring comments and e-mail discussion. Above all, I am indebted to Gerson Topnaar, who patiently shared his language with me. Of course, I alone am responsible for the opinions, errors, and shortcomings in this paper. In particular, it should be noted that, due to time constraints, not all examples from printed sources appearing here could be checked with G.T.

nouns, but not adjectives, and nouns cannot modify adjectives. Nonetheless, the three classes show striking similarities in their syntactic behavior.

Khoekhoe is an SOV language. The V slot may be occupied by a word from any of the three open word classes: a verb (1a), an adjective (1b), or a noun (1c-d). Both commons nouns (1c) and proper nouns (1d) can be used. Even deictic elements (1e-f), numerals (1g), and possessives (1h) can act as predicates. While the choice of the TAM marker depends on the predicate, the syntactic behaviour of the different predicates is entirely parallel:

(1) a. saa=ts ge ra |khii
you=2MS DECL TAM come
'You (m.) are coming.'

b. om=s ge (a) kai
house=3FS DECL TAM big
'The house[1] is big.'

c. saa=ts ge (a) gao-ao
you=2MS DECL TAM king
'You (m.) are a king.'

d. saa=ts ge (a) Petru
you=2MS DECL TAM Peter
'You (m.) are Peter.'

e. om=s ge a nee
house=3FS DECL TAM this
'The house is this one.'

f. tii=ta ge (a) saa
I=1s DECL TAM you
'I am you.'

g. tara=di ge a !nona
woman=3FS DECL TAM three
'The women are three.'

h. om=s ge a tii
house=3FS DECL TAM mine
'The house is mine.'

Most verbs use the 'dynamic' TAM marking pattern, while nonverbal predicates show the same 'static' TAM marking pattern as stative verbs like 'know', 'believe', 'be able', 'die', etc:

(2) a. go |khii
'came'
nî |khii
'will come'

b. go gao-ao ii
'was a king/queen'
nî gao-ao
'will be a king/queen'

c. go ǂan ii
'knew'
nî ǂan
'will know'

Unlike languages like Russian and Arabic, it is not possible to simply analyze these clauses as copulative structures without overt copula. The crucial point is that the predicative element in (1c-f) is not an NP as it would occur in an argument position. In Khoekhoe, argument NPs generally end with a person-gender-number (PGN) morpheme (=*ts*, =*ta* and =*s* in 1), which is not found when a noun is used as a predicate. Compare the NP *gao-ao=b* in (3a-b) with the predicative noun *gao-ao* in (1c), and the NPs in (3c) with the predicative adjectives and deictics in (1):

(3) a. [gao-ao=b] ge ra |khii
king=3MS DECL TAM come
'The king is coming.'

[1] Khoekhoe NPs are not marked for definiteness and the choice of definiteness in the translations is arbitrary.

b. tii=ta ge [gao-ao=b] !oa ra mîî
1S=1S DECL king=3MS to TAM speak
'I am speaking to the king.'

c. [kai=b]/[‖naa=b]/[saa=ts] ge ra |khii
big=3MS/those=3MS/you=2MS DECL TAM come
'the big one/those one/you are coming'

Predication and NPs structure in Khoekhoe has been studied by Wilfrid Haacke in several of his many publications on Khoekhoe, in particular Haacke (1976, 1977, 1978, 1980). In this line of work, he argued in detail that Khoekhoe NPs should be analyzed as being derived from clauses – by transformations in then common incarnations of derivational syntax, or at least diachronically. For instance, the NP *gao-ao=b* 'the king' is derived from the sentence

(4) ...=b a gao-ao
...=3ms TAM king
'he is a king'

in which the subject is realized only by a PGN clitic, which is argued to be related by some synchronic or diachronic 'transformation' to the final PGN marker of the argument NP *gao-ao=b* 'the king'. In argument position, NPs could interpreted as arising from parentheticals. For instance, (3a) could be interpreted – at least diachronically – as 'he – he is the king – is coming', and (3b) as 'I am speaking to him – he is the king'. In Haacke (1992a, 194), he asks to what extent the analysis has synchronic significance and whether it can 'be reconciled with the universals of core grammar as currently perceived'.

While we do not work in a framework assuming grammatical universals and therefore cannot presume to give an answer to the second question, we will address the first question and argue that a clausal analysis of Khoekhoe NPs is indeed very appropriate in a synchronic non-transformational grammar of Khoekhoe. More precisely, we will argue that nouns are primarily genuine predicates, just like verbs, and that argument NPs are free relative clauses. This differs somewhat from Haacke's transformational analysis, but agrees entirely with his general claim that Khoekhoe nouns as appearing in argument positions are not primary, but are derived from clauses.

1.1 Free Relative Clauses

Our point of departure is the observation that argument NPs formally resemble free relative clauses. Khoekhoe free relative clauses consist of a clause containing a gap or resumptive pronoun, followed by a PGN marker indicating the index features of the referent of the free relative clause. The filler can be in any depth, subject to island effects when there is no resumptive. Simple examples for subject and complement relativization are provided in (5).

(5) a. [|khii ra]=b
come TAM=3MS
'the one (m.) who is coming'

b. [mûû=ta ra]=b
see=1S TAM=3MS
'the one (m.) I am seeing'

Hagman (1977) notes the similarity to argument NPs with noun head, but claims that the crucial difference is that relative clauses require TAM marking:

(6) a. gao-ao a=b
king TAM=3MS
'the king'

b. kai a=n
big TAM=3P
'the big ones'

c. ‖naa a=n
those TAM=3P
'those ones'

d. saa a=ts
you TAM=2MS
'you'

However, as nominal predicates can be used with a zero TAM marker (1), free relative clauses of the shape of a lexical argument NP, like *gao-ao=b* 'one who is a king', would be expected to exist. This suggests that argument NPs with noun heads might in fact be identical to these expected relative clauses lacking a TAM marker. We will in the following argue that this is indeed the case.

NPs looking like relative clauses without a TAM marker do not seem to show unexpected restrictions in their distribution. They can be marked for negation in a way completely parallel to predicates and relative clauses with TAM marker:

(7) a. ‖îî=b ge (a) Petru tama
3=3MS DECL TAM Peter NEG
'He is not Peter.'

b. Petru tama (a)=b
Peter NEG TAM=3MS
'one who is not Peter'

c. ‖îî=b ge ǂgae tama
3=3MS DECL smoke NEG
'He does not smoke.'

d. ǂgae tama=b
smoke NEG=3MS
'one who does not smoke'

Word Order Alternation Nonverbal predicates are also subject to the same word order alternations as verbal predicates. In particular, both may be fronted to the position immediately in front of the subject, which then can only be expressed by an enclitic PGN marker:

(8) a. |khii=ts ge ra
come=2MS DECL TAM
'You are coming.'

b. gao-ao=ta ge (a)
king=2MS DECL TAM
'I am a king.'

c. gao-ao=b ge (a)
king=3MS DECL TAM
'He is a king.'

Since the TAM marker *a* and to some extent also the declarative clause type marker *ge* are optional in clauses like (8b-c), this has the consequence that expressions that look like noun phrases, such as *gao-ao=b* in (8c), may constitute clauses – which is one of Haacke's main arguments for his analysis (Haacke, 1980). This is reminiscent of the situation in Nahuatl, where a noun phrase may constitute an utterance, which Launey considers typical of 'omnipredicative' languages.

Nominalization Nonverbal predicates also behave like verbal predicates with respect to nominalization. Any clause can be nominalized by adding a 3FS PGN marker:

(9) a. ǁîî=s go mûû [|khii=b go]=s=a
3=3FS TAM see come=3MS TAM=3FS=OBL
'she saw that he came'

b. ǁîî=s go mûû [|ui=b a]=s=a
3=3FS TAM see stone=3MS TAM=3FS=OBL
'she saw that it was a stone'

1.2 Complex Nominal Predicates

NPs can be modified by elements preceding the noun, such as adjectives, relative clauses (that is, clauses containing a coreferent gap or a resumptive), numerals, deictics, and possessives, which all precede the noun:

(10) a. khoexa khoe=s ge go |khii
friendly person=3FS DECL TAM come
'The friendly woman came.'

b. [[ǂan tama=ta hâã] khoe=s] ge go |khii
know NEG=1S TAM person=3FS DECL TAM come
'The woman that I don't know came.'

The parallelism between NPs and free relative clauses extends to such modified NPs, since the part of the NP preceding the PGN marker may occupy the V slot:

(11) a. Petru=b ge a [khoexa khoe]
Peter=3MS DECL TAM friendly person
'Peter is a friendly person.'

b. Petru=b ge a [ǂan tama=ta hâã] khoe
Peter=3MS DECL TAM know NEG=1S TAM person
'Peter is a person that I don't know.'

We claim that these elements are complex predicates. It is impossible for the modifier to be separated from the noun:

(12) a. * Petru=b ge khoexa a khoe
Peter=3MS DECL friendly TAM person
'Peter is a friendly person.'

b. * Petru=b ge [ǂan tama=ta hâã] a khoe
Peter=3MS DECL know NEG=1S TAM TAM person
'Peter is a person that I don't know.'

In Khoekhoe, TAM markers like *a* can generally be placed anywhere between the subject and the verbal complex, but not within the verbal complex. Thus, (12) suggests that noun and modifier form a constituent. They can be embedded under at least one control verb, namely *kai* 'to become' (13), which supports the claim that they are predicates. Again, the modifier cannot be separated from the noun.

(13) a. tsîî=b ge ǂkham=ma ge [ǂkham ǂkhoa] kai
and=3MS DECL young=3MS+OBL TAM young elephant become
'and the young one became a young elephant' (Schaar, 1917, 83)

b. ǁîî=s ge go [gao=b di |hoo] kai
3=3FS DECL TAM king=3MS POSS friend become
'she became a friend of the king'

1.3 Conclusion

We conclude with the hypothesis that Khoekhoe argument NPs are syntactically indistinguishable from relative or nominalized clauses. As nouns by themselves can only be used predicatively, this suggests that nouns are essentially predicates and are not referential. The only lexical elements that can function referentially but not as predicates are the PGN markers themselves, which are also used as enclitic subject pronouns (8). Almost any more complex NP can be analyzed as consisting of a clause or a predicate and a PGN marker. As this is exactly the structure of relative clauses and nominalized clauses in Khoekhoe, we claim that all NPs – except for the bare PGN markers – are relative clauses or nominalized clauses – as argued for Classical Nahuatl by Launey (Launey, 2002, 117). As Khoekhoe nouns are essentially predicates and phrases only become referential by the addition of PGN markers, we assume that the PGN marker always is the head, which is compatible with the general head-final word-order of Khoekhoe. We will henceforth refer to Khoekhoe argument 'NPs' as DPs.

Khoekhoe does not appear to have type $\langle 1, 1 \rangle$ generalized quantifiers and no determiners representing generalized quantifiers at all, similar to what has been argued, among others, by Jelinek (1995) for Straits Salish and Koenig & Michelson (2012) for Oneida Iroquian.

There are two exceptions to the claim that all Khoekhoe DPs consist of a predicate or clause plus a PGN marker. First, two or more coreferent DPs with identical PGN marking can be serialized, yielding a single DP. Second, DPs may have a possessive marker consisting of âã and the PGN marker adequate for the possessor, which follow the PGN marker of the DP. Both cases can be accounted for easily by adding two phrasal types, the first one recursively licensing DPs consisting of two coreferent NPs, and the second one licensing DPs consisting of an DP and a possessive marker.

2 Analysis

In this section, we will make the ideas from the previous section precise in the context of a formal HPSG analysis of predicates, NPs, and DPs in Khoekhoe. Our point of departure will be the analysis of Khoekhoe clause structure developed in Hahn (2013). There, it was argued that Khoekhoe clauses can be analyzed assuming a relatively uniform right-branching constituent structure illustrated by the bracketing in (14). Word order variation is derived to a large extent by allowing constituents to be discontinuous. While the sequence of multimoraic dependents – which includes DPs, PPs, and adverbs – is determined entirely by their syntactic positions, monomoraic elements – that is, subject PGN markers, clause type markers, and most TAM markers – may appear within the VP, making it a discontinuous constituent. Similarly, the head may be fronted to the position in front of a subject realized by a PGN marker only, accounting for (8). In this paper, we will not need to consider linearization or the syntax of TAM markers, clause type markers and the subject, as they are essentially covered by the linearization component developed previously.

(14) [$_S$ tara=s [$_{TP}$ go [$_{VP}$ ao=ba [ǂkhani=sa [maa]]]]
woman=3FS TAM man=3MS+OBL book=3FS+OBL give
'the woman gave the man a book'

2.1 Lexical Predicates

We assume that not only verbs, but also nouns, adjectives, and some other words including deictics enter the syntax as predicates, with a non-empty SUBJ list, and that any phrase with an empty COMPS list and a nonempty SUBJ list may combine with a subject DP in a *head-subj-phrase* to form a clause. For instance, we assume the following entry for the noun *khoe* 'person':

(15)
$$\left[\begin{array}{ll} \text{CAT} & \left[\begin{array}{ll} \text{SUBJ} & \left\langle \text{DP}_i \right\rangle \\ \text{COMPS} & \langle\rangle \\ \text{HEAD} & \textit{noun} \end{array}\right] \\ \text{CONT} & \left[\begin{array}{ll} \text{INCONT} & \textit{person'} \\ \text{PARTS} & \left\langle \boxed{\begin{array}{c} \ldots\ s\ \ldots \\ \hline \ldots \\ \text{person}(s, x_i) \\ \ldots \end{array}}, \text{person}(s, x_i), s \right\rangle \\ \text{INDEX} & s \end{array}\right] \end{array}\right]$$

where DP is an abbreviation for a saturated structure with HEAD *pgn-marker*. The entries for verbs, deictics, etc. are analogous.

The semantic contribution is formalized in Discourse Representation Theory (DRT, Kamp & Reyle, 1993), which is embedded in HPSG using the constraint-based framework of Lexical Resource Semantics (LRS, Richter & Sailer, 2003).[2] INCONT is the core semantic contribution, while PARTS contains all subterms of the overall semantics that are contributed by the constituent. There is also an EXCONT attribute containing the semantic contribution of the maximal projection. The PARTS list of a phrase is the concatenation of the PARTS lists of the daughters. The LRS *Semantics Principle* (Richter & Sailer, 2003) says, inter alia, that, in an utterance, the PARTS list contains exactly the subexpressions of the EXCONT value. In the entry above, the first element of the PARTS list is a DRT box, which binds the event variable. Its content is not fully specified in the lexical entry – formally, the description says that its variables contain *s* and its content contains the entry $person(s, x_i)$. Other elements of the box can be contributed by other words in a sentence, and the Semantics Principle ensures that only elements that are contributed by some word can appear in the box.

In the context of our simple fragment, it does not seem to be necessary to use this particular approach to semantics, but the DRT representations showcase the essentials of the resulting expressions, and LRS offers the advantage that at the same time it allows semantic underspecification in the grammar and the feature structures yet contain the resolved semantic expressions as in the semantic framework of Pollard & Sag (1994), rather than some underspecified representations whose resolution is thought to take place outside of the linguistic representations, as in MRS (Copestake et al., 2005). A similar approach integrating DRT-style analyses in HPSG using LRS is pursued by Michelson & Koenig (2014).

The structure is similar to those resulting from a lexical rule licensing predicative nouns in English assumed, for instance, by Ginzburg & Sag (2001), but nouns do not have corresponding non-predicative lexical entries in Khoekhoe. As modifiers can always access the index of the referent via the noun's SUBJ list, as there only are predicative nouns in Khoekhoe, our analysis does not suffer from the semantic problems with a lexical rule analysis in English noted by (Gerbl, 2009, 241, Müller, 2009).

We assume that Khoekhoe nouns and adjectives always come with event variables. This is in line with much work on the semantics of nonverbal predicates (e.g., Roy, 2013). We leave open whether it is of the same type as Davidsonian event variables, or of a different one, like Maienborn's (2005, 2007) *Kimian states*.

Any predicate can combine with its subject in a *subj-head-phrase*, forming a clause. Assuming the our previous analysis of clause type and TAM markers and

[2]LRS was originally defined for Ty2, but the framework is more generally applicable to other kinds of logical languages that are defined inductively in the usual fashion. A DRS is represented by a feature structure with two list-valued attributes. The subexpression relation is defined straightforwardly. We assume that, within a feature structure representing a logical expression, two identical complex subterms occuring in different positions must not be token-identical, and that no variable can be bound by more than one box and that, in an utterance, a free variable cannot also occur bound. This is intended to prevent terms from occurring within several boxes.

of fronting developed, this licenses the basic structures in (1) and (8).

2.2 Relative Clauses

Modifying relative clauses In the analysis of English relative clauses by Sag (1997), relative clauses are treated as clauses that modify a noun. In Khoekhoe, modifying relative clauses are clauses that modify a noun and form with it a complex nominal predicate:

(16)
$$\begin{bmatrix} \textit{mod-rel-cl} & \\ \textsc{slash} & \left\langle \boxed{1}\ \textit{rel-pro}_i, \ldots \right\rangle \\ \textsc{to-bind} & \left\langle \boxed{1} \right\rangle \\ \textsc{mod} & \begin{bmatrix} \textsc{head} & \textit{noun} \\ \textsc{subj} & \left\langle \text{DP}_i \right\rangle \\ \textsc{lex} & + \end{bmatrix} \end{bmatrix}$$

LEX enforces predicate complex formation, see Section A. For nonlocal feature percolation, we assume an adaptation of the mechanism of Pollard & Sag (1994) with the modification that a mother node's SLASH value is the concatenation of those of the daughters minus those elements contained in the TO-BIND value of some daughter (not just the head daughter). Subjects are extracted by a unary construction or traces rather than lexically (Bouma et al., 2001) to allow extraction of subjects of complex nominal predicates. This analysis is not the only possible choice, but it makes our analysis of relative clauses simple and uniform and is consistent with the data that we are aware of. However, further research on Khoekhoe unbounded dependencies is needed to assess to what extent arguments for more recent analyses of unbounded dependency constructions apply to Khoekhoe.

Example Analysis: Complex Nominal Predicate We obtain the analysis in Figure 1 for the predicate *ǂan tama=ta hââ khoe* 'a person that I do not know'. We assume that, in a relative clause, the EXCONT of the relative clause must consist entirely of terms on its PARTS list, so in this case the EXCONT value is fully determined. Assuming that the PARTS list of the DP *Petru=b* 'Peter' has the elements

(17) $\left\langle \begin{array}{|c|}\hline \ldots\ x_i\ \ldots\ t\ \ldots \\ \hline \ldots \\ \text{peter}(t, x_i) \\ \ldots \\ \hline \end{array}, \text{peter}(t, x_i), t, x_i \right\rangle$

there is – up to ordering of variables and predicates – a single resolution for the semantics of the clause (11b) *'Peter is a person that I don't know'*, namely:

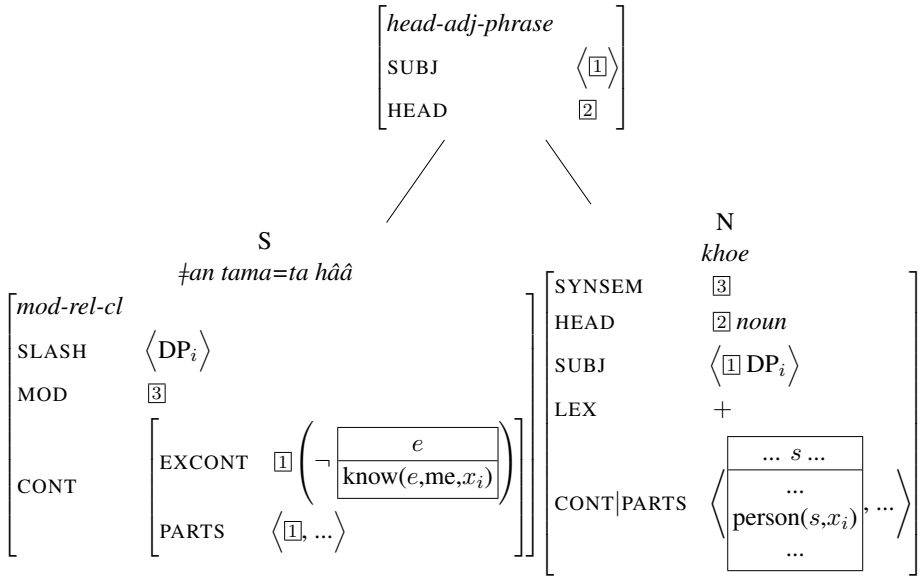


Figure 1: Analysis for the predicate in (11b)

(18)

x_i, s, t

peter(t, x_i)

person(s,x_i)

¬ [e : know(e,me,x_i)]

DP formation We analyze the PGN marker as a pronoun. For the second-person PGN marker *=ts* as in (1a), we assume the following entry:

(19)

HEAD *pgn-marker*

LEX +

CONT: INDEX i [PERSON *second*, NUMBER *singular*]; PARTS $\langle x_i \rangle$

Independent relative clauses modify a PGN marker and form with it a DP:

(20)
$$\begin{bmatrix} \textit{indep-rel-cl} & \\ \text{SLASH} & \left\langle \boxed{1}\ \textit{rel-pro}_i, \ldots \right\rangle \\ \text{TO-BIND} & \left\langle \boxed{1} \right\rangle \\ \text{MOD} & \begin{bmatrix} \text{HEAD} & \textit{pgn-marker} \\ \text{INDEX} & i \\ \text{LEX} & + \end{bmatrix} \end{bmatrix}$$

Thus, DPs are constructed from clauses or predicates by adjunction to a PGN marker via the ordinary *head-adjunct-phrase* type. Spurious ambiguities resulting from recursive application are prevented by the LEX feature. As an alternative, one might stipulate that PGN markers may select as their complement a relative clause, which however would be incompatible with (29.2). A similar phrasal type can be defined for nominalized clauses as in (9), which identify a variable representing their own DRS with the index of the modified PGN marker (cf. Asher, 1993).

Example Analysis: DP Figure 2 shows the analysis of the DP *khoe=s* 'a woman'. A unary projection removes the subject of the noun *khoe* from its SUBJ list and puts it on the SLASH list. We assume that subjects are extracted not lexically, but by a unary projection, because the subject still needs to be available on SUBJ when a noun combines with a modifier to form a complex nominal predicate. The projection applies to a LEX + predicate and forms a LEX − phrase, which can after the saturation of its COMPS requirements become a subject relative clause. In the example, there are no such requirements and the *subj-slash-intro* phrase is also an instance of *indep-rel-clause*, which binds off the nonlocal dependency and modifies a PGN marker, identifying its index with the index of the unrealized subject. In this simple example, the analysis may appear overly complicated, but the point is that DPs of this type are only special cases of a far more general pattern and our analysis captures this observation, generalizing immediately to more complex DPs.

As there are PGN markers for all persons, the analysis correctly predicts the availability of non-first-person DPs, which Launey considers typical of the 'omnipredicative' type. An example is *sa ǁnao=ta* 'I, your uncle' in (21):

(21) [saa ǁnao=ta]=s ta ǂûû?
your uncle=1S=2FS TAM eat
'Are you (trying to) eat me, your uncle?' (Schaar, 1917, 83)

As adjunction is in principle optional, PGN markers may also form complete DPs by themselves, but independently required constraints on the LEX feature enforce that this is only possible in the subject position in sentences like those in (8), as shown in Hahn (2013).

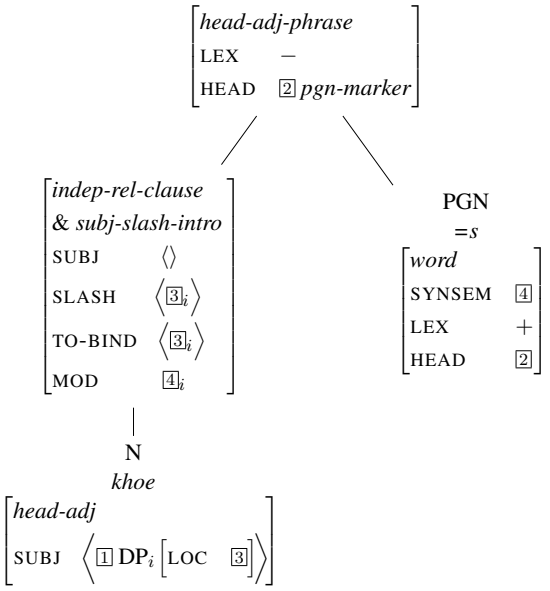


Figure 2: Analysis for the DP *khoe=s* 'a woman'

Other Modifiers Most NP modifiers are relative clauses or predicates by themselves, but some are not. This clearly includes unmarked possessives, which we assume are licensed by a construction combining a DP with an NP, phrasally introducing the possessive semantics. It also presumably includes adjectival modifiers, which need not be intersective and for which we assume special lexical entries derived together with the predicative versions from underlying lexemes. One way of handling their semantics resulting in Montague-style representations (Montague, 1970) could use a Curried version of DRT, where DRT terms are typed as in Montague's IL, and then underspecify which term is applied to the variable of the subject, allowing terms like $(alleged(murderer))(s, x)$ to appear in a DRT box.

3 A Second Predication Structure

We have seen above that word order alternations give rise to clauses that look like DPs, but which can be analyzed as word order alternations of ordinary predicative structures. However, this analysis cannot be maintained for all such structures. Consider

(22) a. |khii go=ts ge
come TAM=2MS DECL
'You came.'

b. |khii go=tsa?
come TAM=2MS+OBL
'Did you come?'

c. * |khii=tsa go?

(23) a. kai a=ts ge
big TAM=2MS DECL
'You are big.'

b. kai a=tsa?
big TAM=2MS+OBL
'Are you big?'

c. * kai=tsa a?

(24) a. ǁââ-tani-ao=b kai=b ge
warrior=3MS old=3MS DECL
'he is an old warrior' (adapted from Krönlein, 1889, 21)

b. ama !gâi-ba=te ǂuu=b ââ=ts ge
really good-APPL=1SG character=3MS POSS=2S DECL
'your character is really pleasing to me' (Krönlein, 1889, 320)

Haacke (1980) argues that these structures are to be analyzed the same way as (8). While this may be possible in his transformational analysis, this causes difficulties in our analysis, and, we believe, more generally in surface-oriented syntactic approaches. (22a), where the elements preceding *ge* together look like a free relative clause, is not licensed as a word order variant of (8a), as the TAM marker would be expected to follow the clause type marker in fronting. This may simply be a defect of our analysis in Hahn (2013), but there are more problems. If such structures are turned into interrogatives, the PGN marker receives the oblique marker *-a*, which is incompatible with the presence of a TAM marker following the PGN marker, as shown in (22c, 23c). Haacke derives the oblique marker from the TAM marker *a*, but it is not obvious how then the presence of two TAM markers in (22b, 23a) should be explained within our analysis. More strikingly, in (24a), there seems to be an element resembling a composite DP consisting of two DPs before the clause type marker. It cannot be analyzed directly as a word order variant as in (8), as the elements preceding the second PGN marker cannot form a predicate. Finally, in (24b), the element preceding *ge* clearly looks like a DP with a possessive marker, i.e., the final PGN marker is not the subject of the predicate, unlike the structures in (8).

Rather than stipulating special rules relating these patterns to structures like (8), it seems much more economical to describe them as a second structure of predication, in which an element that syntactically looks like a DP occupies the clause-initial position, followed by the clause-type marker. Indeed, it appears that any DP, preferably followed by a clause type marker when declarative, can be used in this structure – basic DPs consisting of a predicate or clause plus a PGN marker, multiple DPs, and DPs with possessive marker. Its analysis is very simple: a DP projects to a clause in a non-headed structure (25). We assume that the semantics of the clause is essentially the semantics of the DP. As the interpretation of a Khoekhhoe DP is not a generalized quantifier, but simply a DRS, it is of the right semantic type. Only the index has to be changed – we assume that the index of the clause is an object denoting the DRS of the DP:[3]

[3]We assume that INCONT is inherited, but the INCONT value of a *dp-predicate-clause* and its projections does not play a role in our current grammar fragment and we do not know whether it is needed at all.

(25)
$$\left[\begin{array}{l} \textit{dp-predicate-clause} \\ \textsc{cat} \quad \left[\begin{array}{ll} \textsc{subj} & \langle\rangle \\ \textsc{comps} & \langle\rangle \\ \textsc{head} & \textit{dp-predicate} \end{array}\right] \\ \textsc{cont} \quad \left[\begin{array}{ll} \textsc{index} & z \\ \textsc{excont} & \boxed{1} \end{array}\right] \\ \textsc{dtrs} \quad \left\langle \textsc{dp}\left[\textsc{cont} \quad \left[\textsc{excont} \quad \boxed{1}\right]\right]\right\rangle \end{array}\right] \wedge \; z \approx \boxed{1}$$

where $z \approx \boxed{1}$ means that z represents the DRS $\boxed{1}$ (Asher, 1993, Maienborn, 2005). Constraining the FRONTED value from Hahn (2013) to be $+$, the linearization after CLLD-ed DPs and before the clause type marker is obtained. The appearance of *-a* in questions is stipulated by a constraint referring to initial, LEX $-$, DP daughters of *dp-predicate-clause* and *subj-head-phrase*, as it also occurs in DP subjects in interrogatives.[4]

As argued in detail by Haacke (1980), it is the interaction of this construction with an independently observed phenomenon adjoining an oblique DP to a clause that gives rise to the structures containing two juxtaposed DPs which have traditionally been regarded as copulative structures (Hagman, 1977, 58):[5]

(26) saa=ts ge ǂgae-ǂgui-ao=tsa
you=2S DECL leader=2S+OBL
'You are the leader.' (Hagman, 1977, 59)

Applying the 'finalization' phenomenon described by Hagman (1977, 113) to arguments and adjuncts of the matrix predicate in the DP, this structure also yields the second 'initialization' pattern discussed by Hagman (1977, 111).[6] We should note that, even though we believe Haacke's analysis to be difficult to maintain on a

[4]CLLD-ed subjects show the same pattern as DP subjects, so this might also be stated for constructions realizing CLLD-ed DPs. We leave open how case marking on CLLD-ed DPs is analyzed.

[5]Haacke interprets the second DP as a *deposed subject* in the sense of Haacke (1978), which would be inconsistent with our analysis developed here and in Hahn (2013), as the PGN marker appearing on the first DP is not a subject under our analysis and empty subjects in relativization apparently cannot at the same time give rise to a 'deposed subject'. However, it would in any case be expected that the second DP can also be analyzed as a 'dislocated NP' in the sense of Haacke (1992b), which can be described as a DP adjoined to a clause that is coreferent with some DP occuring within the clause. If both the deposed subject analysis and the dislocated NP analysis were available, a spurious ambiguity would arise. Thus, it seems that the relevant mechanism is the adjunction of a DP.

The fact that the assertive marker *kom...o* may wrap around the second DP is not a problem, as it generally may wrap around 'disloacted NPs'. Since *=o* is optional and can appear on several items, we assume that *=o* may simply attach to any projection marked by the clause type marker *kom*.

[6]In Hahn (2013), it was argued that 'finalization' should be analyzed as extraposition. The idea that elements can be extraposed from within a relative clause may seem dubious, but at least PP modifiers presumably can also be extraposed from DPs in other contexts, cf.

synchronic level, the pattern may certainly be derived *diachronically* from ordinary predication the way described by him.

4 Discussion and crosslinguistic aspects

Crosslinguistic Aspects TAM marking on nouns has received attention in the linguistic literature, cf. the crosslinguistic survey by Nordlinger & Sadler (2004). It has been argued that, at least in some languages where it occurs, it is different from tense marked on verbs or does not mark tense at all (e.g., Tonhauser, 2008). To what extent the semantics of TAM markers for nouns and verbs differ in Khoekhoe and what this means for the analysis, will have to be left to future research. The fact that different predicates have different paradigms of TAM markers might suggest that there is also a semantic difference, but our analysis does not seem to make significant predictions in this regard. It does, though, make the prediction that the temporal semantics of argument DPs matches the temporal semantics of NP predicates. Further research is needed on this issue. Related to this is the observation that, in our analysis, there is no shared event variable for complex nominal predicates with relative clause modifiers, as illustrated by (18).

Haacke's interpretation of DPs as parentheticals is reminiscent of the analysis of DPs in certain omnipredicative languages, where they do not occur in argument positions, but are adjoined to the clause (e.g., Jelinek & Demers, 1994). The situation is clearly less radical in Khoekhoe, where verbs, adjectives and postpositions have obligatory complement positions, which have to be filled *either* by a pronominal clitic or by a DP.

Adverbs and PPs cannot be used by themselves as predicates in Khoekhoe, but require a copula-like verb, which cannot be used with other predicates. This matches a remark by Launey (1994), who observes the same phenomenon in Classical Nahuatl and considers it to be typical for the omnipredicative type, which contrasts with languages like Arabic, that have clauses without overt copula, but are not omnipredicative.

Referential Indices for Nouns? We claimed that nouns are inherently predicative and do not come with an individual variable. A less radical alternative might treat nouns as referential and make NPs predicative on the phrasal level, using the unary phrasal construction proposed by Müller (2009). It might make categorial differentiation between nouns, verbs, and adjectives redundant outside of morphology, as NPs could now be distinguished from verbs and adjectives by having an individual rather than an event as their index. It might also allow a unification of

(i) [tari='i] nî ǁkhuu [saa=ge xa] xam='e kha?
who=3MS TAM attack we=1P of lion=3MS+OBL QUE
'Who of us will attack the lion?' (Krönlein, 1889, 224)

modifying and free relative clauses as relative clauses modifying something coindexed with the SLASH element. Under such an analysis, Khoekhoe DPs would be obtained by making a referential element predicative and then referential again, which seems less economic than our treatment. Making it possible for referential NPs to directly combine with a PGN marker does not solve the problem, as there would then be expected to be a homophonous relative clause which would have to be ruled out artificially to avoid a spurious ambiguity. However, the presence of a individual index in a predicative NP seems to be suggested by the fact that there are examples where a predicative NPs seems to be taken up by a DP adjoined to the clause (Haacke, 1992b, 153):

(27) nee=s ge [ǂguuro !naa] ii xui-ao [‖îî=b ta ‖gûû=n ai
this=3FS TAM first time TAM because 3=3MS TAM parent=3P at
!gammeba ǂgan]=sa
marry ask=3FS+OBL
'because this was the first time he requested consent for marriage from the parents' (|Uriseb, 1993, 1)

However, the DP is also coreferent with the subject and it is not clear that the sentence should not be understood analogous to 'This was the first time', with the interpretation of 'first' reconstructed through discourse factors or semantic underspecification. We therefore see no compelling reasons to treat Khoekhoe nouns as referential and consider it more economical to treat them as primarily predicative.

A related idea is that the event variable for nominal predicates might come not from the noun, but from the TAM marker. However, as the same TAM markers are used for nouns and for stative verbs, and as present TAM markers are optional with nouns, this does not seem to be an attractive idea in the absence of supporting evidence.

5 Conclusion

Building on work by Haacke (1976, 1977, 1978, 1980), we have argued that Khoekhoe is an 'omnipredicative' language in which members of all open word classes function primarily and without derivation as predicates and in which argument DPs are derived in a uniform manner as projections of pronominal elements, modified by relative clauses. We have shown how the basic syntax and semantics of predication and NP structure in Khoekhoe can be described building on standard components of HPSG analyses. We then examined a second predication structure, which seems to be diachronically related to ordinary nominal predication, but which we showed to be a structure of its own, in which a DP directly projects to a clause.

A The LEX feature

Predicate complex formation is enforced by the LEX feature, following Müller (2002). In Hahn (2013), where it is used for complex verbal predicates, projections of the head are LEX + in the complex, outside −:

(28) saa=ts ge [‖îî=ba go [[mû$_{+}$ kai$_{+}$]$_{+}$=si$_{+}$]$_{+}$]$_{-}$
you=2MS DECL 3=3MS+OBL TAM see make=3FS
'you made him see her'

The LEX value is used to determine the behaviour in linearization: Using the terminology of Donohue & Sag (1999), LEX + constructions and projections of most non-verbal elements are 'compacting', while LEX − projections of verbs are 'liberating', which means that LEX + constituents are always continuous, while higher projections of verbs can be discontinuous. The constraints on LEX are very simple: phrases are LEX + if and only if all daughters are LEX +, and it is assumed that PGN markers are LEX +, while the mechanism building DPs generates LEX − phrases. While this works well for the verbal complexes considered in the previous paper, this leads to trouble with complex nominal predicates, as they can contain complex elements like DPs and clauses, and, under our analysis, not all DPs would be constrained to be LEX −. We therefore need a different analysis if we want to use the same LEX feature for enforcing the formation of verbal and nonverbal complex predicates. The problems are solved by two changes in the analysis: (1) the grammatical constraint linking the LEX values of phrases and daughters is restricted to complement-head-phrases, (2) the complex itself is LEX − and only its strict subcontituents are LEX +.

The first change has the effect that the constraints do not affect NP modifiers any more. The second change has the effect that now adjuncts can decide whether they are realized within the complex or not, which allows us to model the fact that complex nominal predicates can contain complex adjuncts, while otherwise complex adjuncts are to be realized outside of the verbal complex, and also the fact that only some adjuncts are realized inside complex nominal predicates, while others cannot.

We have the following principles, replacing those of Hahn (2013):

(29)
1. Phrases are 'compacting' if and only if their head-daughter is LEX +.
   This predicts without the further stipulations needed in the previous analysis that DPs and PPs are compacting, even though they are LEX −.
2. In a complement-head-phrase, the LEX values of the head daughter and the complement daughter are identical.
   This predicts that the complements realized within the predicate complex are exactly the LEX + complements, which agrees with the previous analysis.
3. If a headed phrase is LEX +, its head daughter must be LEX +.

4. Structures whose head is (the projection of) a non-predicate element (i.e., none of *verb*, *adjective*, *noun*, for which we can introduce a common supertype) are LEX −.
   This stipulation is assumed in Hahn (2013) for DPs and PPs.

Unlike the previous analysis, these principles imply no correlation between the LEX values of non-head daughters other than complements and the LEX value of the head daughter. For adjuncts, such correlations can be enforced by the MOD value. A modifier modifying the projection of a predicate has MOD|LEX + if and only if it is realized within a predicate complex.

The analysis makes the prediction that predicates cannot take LEX − complements in their predicative complex. It seems that only nouns can have LEX + dependents in their complex. There are some non-heads in noun phrases that may qualify as complements. These mainly are clauses attaching to words like *!khai* 'fact'. It seems, however, more reasonable to analyze these as adjunct clauses that differ from modifying relative clauses (16) in that they identify the index of the subject of the modified noun with a variable representing their own semantics. The prediction therefore seems not to conflict with the data.

References

Asher, Nicholas. 1993. *Reference to Abstract Objects in Discourse*. Dordrecht: Kluwer.

Bouma, Gosse, Rob Malouf & Ivan A Sag. 2001. Satisfying Constraints on Extraction and Adjunction. *Natural Language and Linguistic Theory* 1(19). 1–65.

Copestake, Ann, Dan Flickinger, Carl Pollard & Ivan Sag. 2005. Minimal Recursion Semantics: An Introduction. *Research on Language & Computation* 3(4). 281–332.

Donohue, Cathryn & Ivan A Sag. 1999. Domains in Warlpiri. In *Sixth international conference on HPSG – Abstracts*, 101–106. Edinburgh. `http://www-csli.stanford.edu/~sag/papers/warlpiri.ps`.

Gerbl, Niko. 2009. *An Analysis of Pseudoclefts and Specificational Clauses in Head-driven Phrase Structure Grammar*: Universität Göttingen dissertation.

Ginzburg, Jonathan & Ivan A Sag. 2001. *Interrogative Investigations: the form, meaning, and use of English Interrogatives*. Stanford, California: CSLI Publications.

Haacke, Wilfrid H G. 1976. *A Nama grammar: the noun-phrase*. University of Cape Town MA thesis.

Haacke, Wilfrid H G. 1977. The so-called 'Personal Pronoun' in Nama. *Khoisan Linguistic Studies* 3(3). 43–62.

Haacke, Wilfrid H G. 1978. *Subject deposition in Nama*. Colchester: University of Essex dissertation.

Haacke, Wilfrid H G. 1980. Nama coreferential copulative sentences re-assessed. In Jannie Winston Snyman (ed.), *Bushman and hottentot linguistic studies*, 80–106. Pretoria: University of South Africa.

Haacke, Wilfrid H G. 1992a. Compound noun phrases in Nama. In Derek F. Gowlett (ed.), *African linguistic contributions*, 189–194. Hatfield.

Haacke, Wilfrid H G. 1992b. Dislocated noun phrases in Khoekhoe (Nama/Damara): further evidence for the sentential hypothesis. *Afrikanistische Arbeitspapiere* 29. 149–162.

Hagman, Roy S. 1977. *Nama Hottentot grammar*, vol. 15. Indiana University Bloomington.

Hahn, Michael. 2013. Word Order Variation in Khoekhoe. In Stefan Müller (ed.), *Proceedings of the 20th International Conference on Head-Driven Phrase Structure Grammar*, 48–68.

Jelinek, Eloisa & Richard A. Demers. 1994. Predicates and Pronominal Arguments in Straits Salish. *Language* 70(4). 697–736.

Jelinek, Eloise. 1995. Quantification in Straits Salish. In E. Bach, E. Jelinek & B. H. Partee (eds.), *Quantification in natural language*, vol. 54 Studies in Linguistics and Philosophy, 487–540. Springer.

Kamp, Hans & Uwe Reyle. 1993. *From Discourse to Logic*. Dordrecht: Kluwer.

Koenig, Jean-Pierre & Karin Michelson. 2012. The (non)universality of syntactic selection and functional application. In Christopher Piñón (ed.), *Empirical studies in syntax and semantics* 1987, 185–205. Paris: Centre National de la Recherche Scientifique.

Krönlein, Johann Georg. 1889. *Wortschatz der Khoi-Khoin*. Berlin: Deutsche Kolonialgesellschaft.

Launey, Michel. 1994. *Une grammaire omnipredicative: Essai sur la morphosyntaxe du nahuatl classique*. Paris: CNRS Editions.

Launey, Michel. 2002. Compound nouns vs. incorporation in Classical Nahuatl. *Studia typologica* 4. 113–134.

Maienborn, Claudia. 2005. On the limits of the Davidsonian approach: the case of copula sentences. *Theoretical Linguistics* 31(3). 275–316.

Maienborn, Claudia. 2007. On Davidsonian and Kimian States. In Ileana Comorovski & Klaus von Heusinger (eds.), *Existence: Semantics and syntax*, 107–130. Springer.

Michelson, Karin & Jean-Pierre Koenig. 2014. Deconstructing SYNtax. In Stefan Müller (ed.), *Proceedings of the 21st International Conference on Head-Driven Phrase Structure Grammar*, Stanford: CSLI Publications.

Montague, Richard. 1970. English as a Formal Language. In Bruno Visenti (ed.), *Linguaggi nella Società e nella Tecnica*, 189–224. Milan: Edizioni di Comunità.

Müller, Stefan. 2002. *Complex Predicates: Verbal Complexes, Resultative Constructions, and Particle Verbs in German* (Studies in Constraint-Based Lexicalism 13). Stanford: CSLI Publications.

Müller, Stefan. 2009. On Predication. In *Proceedings of the 16th International Conference on Head-Driven Phrase Structure Grammar*, 213–233. Stanford: CSLI Publications.

Nordlinger, Rachel & Louisa Sadler. 2004. Nominal tense in crosslinguistic perspective. *Language* 80(4). 776–806.

Pollard, Carl & Ivan A Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Richter, Frank & Manfred Sailer. 2003. Basic Concepts of Lexical Resource Semantics. In Arnold Beckmann & Norbert Preining (eds.), *ESSLLI 2003 – Course Material I*, vol. 5 Collegium Logicum, 87–143. Wien: Kurt Gödel Society.

Roy, Isabelle. 2013. *Nonverbal Predication: Copular Sentences at the Syntax-Semantics Interface*. Oxford: Oxford Univerxsity Press.

Sag, Ivan A. 1997. English Relative Clause Constructions. *Journal of Linguistics* 431–484.

Schaar, W. 1917. Nama-Fabeln. *Zeitschrift für Kolonialsprachen* 8. 81–109.

Tonhauser, Judith. 2008. Nominal Tense? The Meaning of Guaraní Nominal Temporal Markers. *Language* 83(4). 831–869.

|Uriseb, A. 1993. *‖Hûisa !nubu ‖gaeǂhôadi*. Windhoek: Gamsberg Macmillan.

Degree adverbs in Mauritian

Shrita Hassamal
LLF, CNRS, Université Paris Diderot

Anne. Abeillé
IUF, LLF, CNRS, Université Paris Diderot

Proceedings of the 21st International Conference on
Head-Driven Phrase Structure Grammar

University at Buffalo

Stefan Müller (Editor)

2014

Stanford, CA: CSLI Publications

pages 259–279

Hassamal, Shrita & Anne. Abeillé. 2014. Degree adverbs in Mauritian. In Stefan Müller (ed.), *Proceedings of the 21st International Conference on Head-Driven Phrase Structure Grammar, University at Buffalo*, 259–279. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2014.14.
CC-BY 4.0

1 Introduction[1]

From a semantic point of view, degree words (*very, a lot*) can be analysed as modifying a value on a scale, which can be an intensity scale for properties (1a), or a quantity scale for objects (1b) or events (1c) (Kennedy & McNally 2005). Some degree words are specialised for a certain type of scale (e.g. an intensity scale for *extremely*) and some are underspecified (e.g. *a lot* is compatible with all scales) (1b, c, d).

(1) a. John is very / extremely intelligent.
b. A lot of books.
c. John goes to the movies a lot.
d. John likes the book a lot.

From a syntactic point of view, degree words usually differ with respect to their category or function: in English, *very, very much* are adverbs, the former modifying adjectives and adverbs (2a), the latter modifying verbs (2b); *many, much* are determiners (specifying count and mass nouns)(2c).

(2) a. John works very hard / drives very quickly.
b. John likes the book very much.
c. He does not have many books / much luck.

Comparative words, *more* and *less*, which compare two values on one (or more) degree scales, seem more polymorphic and combine with all the major syntactic categories:

(3) a. John is more intelligent/ laughs more often.
b. John goes more to the movies/ likes the book more.
c. He has more luck / more books.

Mauritian, a French-based Creole spoken in Mauritius, with no official status, is an understudied language (Baker 1970, Henri 2010, Allessaib 2012…) It has a very recent standardised spelling (Hookoomsing 2004), one dictionary (Carpooran 2009) and very few written corpora: some literary works of Dev Virahsawmy (www.boukiebanane.orange.mu) and the journal of a political party (www.lalitmauritius.org). We conducted several surveys with informants, in situ and on line; the first author is a native speaker.

[1] This work is part of the program Investissements d'Avenir, overseen by the French National Research Agency, ANR-10-LABX-0083, (Labex EFL).
We thank, for their comments F. Henri, D. Godard, A. Kihm, B. Crysmann, O. Bonami, G. Fon Sing, J-P. Koenig, S. Kriegel, the audience of GRGC (Paris) and of the 21st HPSG conference (Buffalo), as well as our Mauritian informants.

Interestingly, Mauritian degree words exhibit an extreme syntactic polymorphism: the same form *tro* 'too' may combine with all major categories: nouns (4a), verbs (4b), adjectives and adverbs (4c).

(4) a. Tro liv / tro per.
'Too many books, too much fear'
b. Paul tro travay.
'Paul works too much'
c. Tro long / tro vit
'Too long / too fast'

When there is more than one form (*mari* / *boukou* 'very, a lot'), they seem to be distinguished on semantic more than syntactic criteria, *mari* requiring an intensity scale (5a,b) and *boukou* a quantity scale (5a,c).

(5) a. Mari per / boukou liv.
very fear / a-lot books
' a lot of fear / a lot of books.'
b. Mo mari kontan liv.
I very like book
'I like the book(s) a lot'
c. Mo al sinema boukou.
'I go to the movies a lot'

On the other hand, Mauritian has different comparative forms: *pli / plis* ('more') (6).

(6) a. Pli vit / pli gran.
more fast /more tall
'faster / taller'
b. Paul travay plis. / ena plis liv.
Paul works more / has more books
'Paul works more / has more books.'

We will focus on the syntax and semantics of four degree adverbs in Mauritian: *mari* ('very'), *boukou* ('a lot'), *pli/plis* ('more') and *mwin/mwins* ('less'), considering their category (adverb? determiner?) and their syntactic function (specifier? adjunct? complement?), as well as their combination with a scalar predicate: do they select it on syntactic or semantic criteria?

2 High degree adverbs in Mauritian

2.1 High degree adverbs in French

French also has some polymorphic degree adverbs, like *trop* ('too, too much, too many'), which may combine all major categories (7).

(7) a. Trop grand / trop vite.
'too big / too fast'

b. Paul travaille trop.
'Paul works too much'
c. Trop peur/ trop de livres.
'too much fear/ too many books'

(Abeillé & Godard 2003) propose that French degree adverbs have a double life: as complements when they follow the verb, and as adjuncts when they precede it or combine with other categories.
French also has a high degree adverb pair: *beaucoup* 'a lot' / *très* 'very'. (Abeillé et al. 2004) analyse *beaucoup* as an adverb combining with any verb and with non-verbal categories that have a quantity scale. *Très*[2] on the other hand, combines only with non-verbal categories that have an intensity scale (table 1)

with verbs		with noun		with adjectives		with adverbs	
intensity scale	quantity scale	intensity scale	quantity scale	intensity scale	quantity scale	intensity scale	comparatives
Il l'aime trop 'he likes it too much'	*Il sort trop* 'he goes out too much'	trop peur 'too much fear'	*trop de livres* 'too many books'	*trop grand* 'too big'	*trop absent* 'too absent'	*trop vite* 'too fast'	-
Il l'admire beaucoup 'he admires it a lot'	*Il sort beaucoup*'he goes out a lot'	-	*beaucoup de livres* 'a lot of books'	-	*beaucoup absent* 'absent a lot'	-	*beaucoup mieux* 'a lot better'
-	-	*très faim* 'much hunger'	-	*très grand* 'very tall'	-	*très vite 'very fast'*	

Table 1: distribution of *trop* ('too'), *beaucoup* ('a lot') and *très* ('very') in French

2.1 Boukou ('a lot') in Mauritian

In Mauritian, *boukou* ('a lot') combines with nouns and verbs, but only if they have a quantity scale (table 2). It precedes nouns and follows verbs.

[2] *Très* ('very') also combines with participles – a mixed category with adjectival properties- with an intensity scale: *Ça m'a très étonné* ('This has a-lot surprised me').

With nouns

B*oukou* combines both with count (8a) and mass nouns (8b), but is infelicitous with predicative nouns (8c).[3]

(8) a. Mo ena boukou liv.
'I have a lot of books'

b. Peyna boukou letan reste. (Virahsawmy 1991, *Toufann*)
Not-have a-lot time left
'There is not much time left.'

c. ??Mo gagn boukou per/ fin.
I get- SF a-lot fear/ hunger
(Intended meaning) 'I am very afraid / hungry.'

In Mauritian, all nouns can be used bare (9a). Bare nouns may combine with a possessive (9b) or a demonstrative determiner (9c).

(9) a. Mo'nn vann liv.
I PERF sell- SF book
'I sold a book /books.'

b. Mo'nn vann mo liv.
I PERF sell-SF POSS.1sg book
'I sold my book(s).'

c. Mo'nn vann sa liv la.
I PERF sell-SF DEM book DEF
'I sold this book / these books.'

A plural marker – *bann* combines with bare nouns (10a) and with possessive (10b) and demonstrative (10c) determiners (Alleesaib 2012). However, degree words are not compatible with *bann* (10d*).* We therefore analyse *tro, boukou* as optional adverbs and not as determiners.

(10) a. Mo'nn vann bann liv.
I PERF sell- SF PLU book
'I sold the books.'

b. Mo'nn vann mo bann liv.
I PERF sell- SF POSS.1SG book
'I sold my books.'

c. Mo'nn vann sa bann liv la.
I PERF sell- SF DEM PLU book DEF
'I sold these books.'

d. *tro bann liv / *boukou bann liv.
too PLU book / a-lot PLU liv

[3] We gloss the TAM markers as PAST (past), PERF (perfective), PROG (progressive), FUT (future). SF stands for short verbal form and LF long verbal form.

With verbs

Boukou modifies verbs with a quantity scale, object-quantity (11a) or event-quantity (11b). In (11c), the meaning is ambiguous, as with verbs with an incremental theme, since *boukou* may modify the quantity of events (of writing) or the quantity of written objects. It is usually judged infelicitous with verbs associated with an intensity scale (11d), but there is variation among speakers – probably due to the influence of French *beaucoup* which is not so constrained (table 1).

(11) a. Mo manz boukou.
'I eat a lot'

b. Mo al sinema boukou.
'I go to the movies a lot'

c. Dev finn ekrir boukou lor Kreol Morisien. (Virahsawmy 1991, *Toufann*)
Dev PERF write a-lot on Creole Mauritian
'Dev has written a lot on Mauritian Creole'

d. %Mo admir li boukou.
'I admire him/her a lot'

Interestingly, Mauritian verbs have a conjugation with a short form (SF) and a long form (LF)[4], the choice of form being determined by syntactic and discourse factors: the verb is in SF when followed by a non-clausal complement and not focalised (12b); the LF is used otherwise (12a,c) (Henri & Abeillé (2008) and Henri (2010)).

(12) a. Paul inn manze. / *manz.
Paul PERF eat-LF / *eat-SF
'Paul has eaten'

b. Paul inn manz pom.
Paul PERF eat-SF apple
'Paul has eaten apples.'

c. Mo'nn danse. /*dans.
I PERF dance-LF/*dance-SF
'I danced.'

The verb is usually in LF when followed by an adverb (13a) or another adjunct (13b)

(13) a. *Paul inn danse yer / *dans yer.
Paul PERF danse- LF yesterday
'Paul has dansed yesterday.'

[4] Baker (1972) shows that 70% of the verbal lexicon in Mauritian have two forms, the others being syncretic, like *travay* (work), *ena* (have), *vo* (be worth), *ekrir* (write).

b. Paul inn manze gramatin. / *manz gramatin.
'Paul has eaten-LF in the morning.'

Interestingly, with *boukou* the verb is in SF, and the adverb thus behaves as a complement.

(14) a. Mo'nn manz boukou.
I PERF eat- SF boukou
'I have eaten a-lot.'
b. Mo'nn dans boukou.
I PERF dance-SF a-lot
'I have danced a lot.'

With other categories

Boukou modifies adjectives with an event quantity scale (15a), but not adjectives with an intensity scale (15b).

(15) a. Paul absan boukou.
'Paul is absent a lot.'
b. *Rita so zip boukou long.

Boukou does not combine with adverbs (16a), except with inequality comparatives (16b).

(16) a. * boukou vit / *boukou souvan
a-lot fast/ a-lot often
b. Vre lamitie vo boukou plis ki lor.
True friendship be-worth a-lot more than gold
'True friendship is worth much more than gold.' (Virahsawmy, *Tizistwarlasinn*)

with verbs		with nouns		with adjectives		with adverbs	
intensity scale	quantity scale	intensity scale	quantity scale	intensity scale	quantity scale	intensity scale	compa ratives
Mo tro admir li	*Mo al sinema tro*	tro per	*tro liv*	tro long	*tro absan*	tro vit	
% Mo admir li boukou 'I admire him a lot'	*Mo al sinema boukou* 'I go to the movies a lot'	**boukou per*	*boukou liv* 'a lot of books'	**boukou long*	*absan boukou* 'absent a lot'	**boukou vit*	*bouko u pli vit* 'a lot faster'
Mo mari admir li 'I admire him very much'	**Mo mari al sinema*	*mari per* 'very scared'	* *mari liv*	*mari long* 'very long'	**mari absan*	*mari vit* 'very fast'	*mari pli vit

Table 2: *tro* ('too'), *boukou* ('a lot') and *mari* ('very') in Mauritian

2.2 *Mari* (‘very’) in Mauritian

Mauritian has developed another form *mari*[5] (‘very’), which also combines with nouns and verbs, but selects predicates with an intensity scale, and always precedes them. *Boukou* and *mari* are thus in complementary distribution (table 2).

With nouns
With nouns, *mari* is optional and modifies predicative nouns with an intensity scale (17a), not count or mass nouns (17b).

(17) a. Konntou ti pe gagn mari traka.
Konntou PAST PROG have- SF very worry
‘Konntou was very worried.’ (Virahsawmy, *Proz Literer*)
b. * mari liv / *mari lafarinn
very books / very flour

With verbs
Mari combines with verbs associated with an intensity scale (18a) but not with count event ones (18b).

(18) a. Li mari amerd li. (Lalit 2012, *Magazine 104*)
‘(S)he annoys him/her a lot’
b. *Paul mari al sinema.
Paul very go-SF movies

Unlike *boukou, mari* always precedes the verb and never triggers the SF (19). It thus behaves as an adjunct.

(19) a. Mo soulie pe mari briye / *briy.
POSS1SG shoes PROG very shine-LF/ shine-SF
‘My shoes are shining a lot.’
b. *Mo soulie pe briye / briy mari.
POSS1SG shoes PROG shine-LF/ shine-SF very

Certain verbs can be associated with both types of scales. For example *plore* (‘to cry’) can have a quantity scale (object scale: quantity of tears or event-scale: frequency) or an intensity scale. The two degree adverbs thus trigger a different meaning, and a different verb form: the SF appears with *boukou* (20b), LF with *mari* (20a).

[5] *Mari* belongs to an informal register, and comes from the French noun *mari* (‘husband’), which also gave an adjective (‘superior’) in Mauritian. Mauritian has not retained French *très*. *Extra*, with the same distribution as *mari*, is also an innovation (French *extra*, from Latin, is a prefix or an adjective).

(20) a. Rita inn mari plore.
Rita PERF very cry-LF
'Rita cried a lot (intensively).'
b. Rita plor boukou.
Rita cry-SF a lot
'Rita cries a lot (often or a lot of tears).'

With other categories
Mari combines with adjectives associated with an intensity scale (21a) but not with count event adjectives (21b).

(21) a. Tifi la pa ti zoli me li mari riss.
Girl DEF NEG PAST pretty but she very rich
'The girl is not pretty but she is very rich.' (Virahsawmy, *Proz Literer*)
b. *Rita mari absan.
Rita very absent

Some adjectives, like *malad* (sick), may have both scales: *mari* triggers an intensive reading (22a) whereas *boukou* a count-event or frequency interpretation (22b).
(22) a. Paul mari malad.
'Paul is very sick.'
b. Paul malad boukou.
Paul sick a lot
'Paul is sick a lot (often).'

Unlike *boukou, mari* combines with adverbs (23a) but not with comparatives (23b).

(23) a. Me mari vit sitiasion ti sanze.
But very fast situation PAST change-LF
'But very fast the situation changed.' (Virahsawmy, *Proz Literer*)
b. * Mari pli vit.
very more fast 'a lot faster'

In Mauritian, both *boukou* and *mari* are adverbs with a complementary distribution. Syntactically, *boukou* follows the verb and triggers the verbal short form, as a complement, while *mari* precedes it as an adjunct. With non-verbal categories, both are adjuncts. Semantically, *boukou* modifies predicates with a quantity scale while *mari* modifies predicates with an intensity scale.

3 The inequality comparative adverbs in Mauritian

Mauritian has two forms: *pli* and *plis* for the superior comparative adverb ('more') and *mwin* and *mwins* for the inferior comparative adverb ('less').

3.1 *Pli* / *plis* ('more') in Mauritian

The two forms *pli / plis* [6], coming from French *plus* ('more')[7], are in complementary distribution (table 3), *plis* being the default form (used in isolation).

The distribution between the two forms is not phonologically determined since both forms can be followed by a vowel or a consonant (24).

(24) a. Paul pli intelizan / pli gran.
'Paul is more intelligent / more tall.'

b. Paul plis admir li. / plis travay.
'Paul admires her/him more / works more.'

with verbs		with nouns		with adjectives		with adverbs
intensity scale	quantity scale	intensity scale	quantity scale	intensity scale	quantity scale	
**Mo admir li pli*	**Mo al sinema pli*	*pli per* 'more scared'	**pli liv*	*pli gran* 'more tall'	**pli absan*	*pli vit* 'more fast'
Mo admir li plis 'I admire her more'	*Paul al sinema plis* 'I go to the movies more'	**plis per*	*plis liv* 'more books'	**plis gran*	*plis absan* 'more absent'	** plis vit*

Table 3: Distribution of *pli* and *plis* 'more' in Mauritian

With nouns

With nouns, *pli* only combines with predicative gradable nouns (25a) and *plis* is used with count and mass nouns (25b,c)

---

[6] Carpooran (2009)`s dictionary codes *plis* as a conjunction (because of the additive use: 2 plis 2 = 4) and *pli* as an adverb which may affect quantity or quality.

[7] French has one form of the superior comparative *plus* ('more') and some synthetic comparatives: *meilleur* ('better'), *pire* ('worse'). Spoken French has three forms for *plus*: /ply/, /plys/ and /plyz/ before vowel for the liaison form. A precise study of the distribution of ply/plys in French still has to be done. French also has a negative homonym *plus* ('no longer') which in Mauritian has become *nepli.*

(25) a. Paul inn gagn pli / %plis per (ki Rita).
Paul PERF have-SF more fear (than Rita)
'Paul is more scared (than Rita).'
b. Paul ena *pli / plis liv / lafarinn.
Paul have more book / flour.
c. Vedir dimunn pe manz buku plis legim.
Means people PROG eat-SF a lot more vegetables
'It means people are eating much more vegetables.'
(Lalit 2010, *Magazine 107*)

With verbs

With verbs, only *plis* is possible, whether they have an intensity scale (26a) or a quantity scale (26b).

(26) a. Paul admir li plis/*pli (ki Rita).
'Paul admires him/her more than Rita.'

b. Paul al sinema plis/*pli (ki Rita).
'Paul goes to the movies more than Rita.'

Plis can either precede (27a) or follow (27b) the verb. When it follows, it triggers the short verbal form and thus behaves as a complement (27b). It can also appear after a complement (27c).

(27) a. To soulie pe plis briye /*briy (ki pou mwa).
POSS.2SG shoes PROG more shine-LF / *shine-SF than POSS.1SG
b. To soulie pe briy plis (ki pou mwa).
POSS.2SG shoes PROG shine-SF more than POSS.1SG
'Your shoes are shining more than mine.'

c. Finn ler pu devlop mobilizasyon plis.
finish time to develop-SF mobilization more
'It is time to develop mobilization more.'
(Lalit 2012, *Magazine 107*)

When it precedes the verb, it occurs between the TAM markers and the verb, and can only be modified by a degree adverb (28a), not by a *ki* ('than') argument (28b), which must be postverbal.

(28) a. To soulie pe [boukou plis] briye.
'Your shoes are shining a lot more'
b. *To soulie pe [plis ki pou mwa] briye.
'Your shoes are shining more than mine.'

With other categories

With adjectives, both comparatives are possible, but *pli* selects adjectives with an intensity scale (29a) while *plis* combines with event count adjectives (29b).

(29) a. Paul pli / *plis gran (ki Rita).
'Paul is taller (than Rita).'
b. Paul inn plis/*pli absan ki Rita sa lane la.
Paul PERF more absent than Rita DEM year DEF
'Paul has been more often absent than Rita this year.'

Certain adjectives (*malad*, 'sick') can have both scales, *pli* triggering an intensity reading (30a) and *plis* a frequency reading (30b).

(30) a. Paul pli malad ki Rita. 'Paul is sicker than Rita.'
b. Paul malad plis ki Rita. 'Paul is sick more often than Rita.'

Adverbs, on the other hand, only combine with *pli.*

(31) a. Pli / *plis vit ; pli / *plis boukou.
More fast ; more a-lot
'Faster' 'much more'
b. (…) e nu amen li pli lwin. (Lalit 2012, *Magazine 107*)
(…) and 1PL bring-SF 3SG more far
'(…) and we bring it further'

In other contexts

In short answers (32a) and in elliptical clauses (32b), *plis* is always used.

(32) a. Spkr 1: – Paul pli gran ou mwin gran?
Paul more tall or less tall
'Is Paul taller or shorter?'
Spkr 2: – Plis. / *Pli ' More'

b. Rita inn kontan fim la enn tigit me Paul plis/ *pli.
Rita PERF like movie DEM a little but Paul more
'Rita liked the movie a little, but Paul more.'

In metacomparison contexts, when the adverb modifies not a degree scale but the truth value of the proposition, *plis* is also always used.

(33) a. Paul plis / *pli kamarad ki papa ar so bann zenfan.
Paul more friend than father with POSS.3sg PLU kids
'Paul is more of a friend than a father to his kids.'

b. Sa latab la plis / *pli kare ki rektang.
DEM table DEF more square than rectangle
'This table is more of a square than a rectangle.'

3.2 *Mwin / Mwins (less)* in Mauritian

Mwin and *mwins*[8] ('less') come from French *moins*[9], and have the same distribution as *pli* and *plis* respectively: they are in complementary distribution (table 4), *mwins* being the default form (used in isolation).
The distribution of the two forms is not phonologically determined since both can be followed by a vowel or a consonant (34).

(34) a. Paul mwin intelizan / mwin gran
Paul less intelligent / less tall
'Paul is less intelligent / less tall'
b. Paul mwins admir li / mwins travay
Paul less admires3SG / less works
'Paul admires her/him less / works less'

with verbs		with nouns		with adjectives		with adverbs
intensity scale	quantity scale	intensity scale	quantity scale	intensity scale	quantity scale	
**Mo admir li mwin*	**Mo al sinema mwin*	*mwin per* 'less scared'	**mwin liv*	*mwin gran* 'less tall'	**mwin absan*	*mwin vit* 'less fast'
Mo admir li mwins 'I admire her less'	*Mo al sinema mwins* 'I go to the movies less'	**mwins per*	*mwins liv* 'less books'	**mwins gran*	*mwins absan* 'less absent'	**mwins vit*

Table 4: distribution of *mwin* and *mwins* ('less') in Mauritian

With nouns
With nouns, *mwin* only combines with predicative gradable nouns (35a) and *mwins* is preferred with count and mass nouns (35b).

[8] Only *mwins* is recorded in Carpooran (2011)'s dictionary, and is coded as an adjective affecting quantity.
[9] French has one form for the inferior comparative *moins* ('less') which is pronounced /mwɛ̃/, and /mwɛ̃z/ in liaison contexts. A non standard realization /mwɛ̃s/ may appear in the South West of France, in conditions to be determined, but it is unlikely that it could have influenced Mauritian creole.

(35) a. Paul inn gagn mwin/ %mwins per (ki Rita).
Paul PERF have-SF less fear (than Rita)
'Paul is less scared (than Rita).'
b. Paul ena mwins/ * mwin liv / lafarinn.
Paul have less book / flour
'Paul has less books / flour.'

With verbs

With verbs, only *mwins* is allowed, whether they have an intensity (36a) or a quantity scale (36b).

(36) a. Paul admir li mwins/ *mwin (ki Rita).
Paul admire 3SG less (than Rita)
'Paul admires him/her less than Rita.'
b. Paul al sinema mwins/ *mwin (ki Rita).
Paul go-SF cinema less (than Rita)
'Paul goes to the movies less than Rita.'

Mwins can either precede (37a) or follow (37b) the verb. When it follows, it triggers the SF and is thus a complement.

(37) a. To soulie pe mwins briye /* briy.
POSS.2sg shoes PROG less shine-LF/ *shine-SF
b. To soulie pe briy mwins.
POSS.2sg shoes PROG shine-SF less
'Your shoes are shining less.'

With other categories

With adjectives, both comparatives are possible: *pli* selects adjectives with an intensity scale (38a)(38b) and *mwins* event count adjectives (38c).

(38) a. Paul mwin / *mwins gran (ki Rita).
Paul less tall (than Rita)
'Paul is less tall (than Rita).'

b. (…) li kapav mwin ot. (Lalit 2012, *Magazine 107*)
(…) 3SG can less high
'(…) it can be less high.'

c. Paul inn mwins/*****mwin absan ki Rita sa lane la.
Paul PERF less absent than Rita DEM year DEF
'Paul has been less often absent than Rita this year.'

With adjectives that can have both scales, *mwin* triggers an intensity reading (39a) and *mwins* a frequency one (39b).

(39) a. Paul mwin malad ki Rita.
Paul less sick than Rita
'Paul is less sick than Rita (intensively more sick).'
b. Paul malad mwins ki Rita.
Paul sick less than Rita
'Paul falls sick less often than Rita.'

Adverbs, on the other hand, only combine with *mwin.*

(40) a. Paul roul mwin / *mwins vit ki Rita.
Paul drive-SF less fast than Rita
'Paul drives less fast than Rita.'
b. Paul ena mwin / *mwins boukou.
Paul have less a lot
'Paul has less many.'

In other contexts

In short answers (41a) and in elliptical clauses (41b), *mwins* is always used.

(41) a. Spkr 1: – Paul pli gran ou mwin gran?
Paul more tall or less tall
'Is Paul taller or shorter?'
Spkr 2: – Mwins. / * Mwin
Less 'less tall'

b. Rita al sinema boukou me Paul mwins/ *mwin.
Rita goes-SF movies a-lot but Paul less
'Rita goes to the movies a lot, but Paul less.'

In metacomparison contexts, when what are compared are the truth values of two propositions, *mwins* is always used.

(42) a. Paul mwins/ *mwin papa ki kamarad ar so bann zenfan.
'Paul is less of a father than a friend to his kids.'

b. Sa latab la mwins/*mwin kare ki rektang.
DEM table DEF less square than rectangle
'This table is less of a square than a rectangle.'

The inequality comparatives *pli, plis, mwin* and *mwins* are all adverbs. In the non-verbal domain, they are adjuncts, preceding the category they modify. On the semantic level, *pli* and *mwin* only combine with predicates that have an intensity scale and *plis* and *mwins* with predicates that have a quantity scale.

In the verbal domain, only *plis* and *mwins* are possible, and they have a double life : they can precede the verb, as adjuncts, but may also follow it, as complements, triggering the short verbal form.

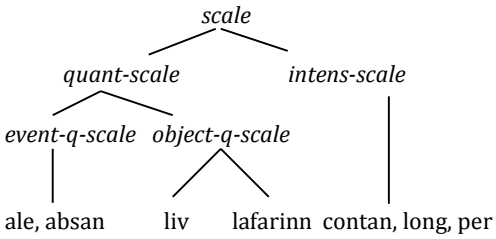
4 An HPSG analysis of Mautitian degree adverbs

In HPSG, degree adverbs have been analysed as specifiers (Pollard and Sag 1994), adjuncts or complements (Abeillé and Godard 2003) or as functors (Van Eynde 2007). Semantically, they have been analysed as modifiers (Abeillé and Godard 2003) or quantifiers (Kay and Sag 2012). We follow (Kennedy and McNally 2005) in analysing them as modifiers of scalar predicates and not as quantifiers.

4.1 Degree adverbs as polymorphic adjuncts

Lexical entries for degree adverbs

In Mauritian, *tro* ('too') can combine with any scalar predicate whereas *boukou* ('a lot') and *mari* ('very') specify a subtype of scale. We rely on the following hierarchy for types of scales, ignoring the possible distinction between open and closed scales (intensity scales are always open).



Syntactically, they may underspecify the category they combine with (Abeillé and Godard 2003)[10].

tro

$$\begin{bmatrix} SYN\ [HEAD\ \begin{bmatrix} adv \\ MOD\ \text{SEM}\ [SCALE\ s \\ VALUE\ d] \end{bmatrix}] \\ SEM\ \begin{bmatrix} excess-val-rel \\ ARG\ d \end{bmatrix} \end{bmatrix}$$

boukou

$$\begin{bmatrix} \begin{bmatrix} SYN\ \begin{bmatrix} HEAD\ \begin{bmatrix} adv \\ MOD\ SEM\ \begin{bmatrix} SCALE\ quant-s \\ VALUE\ d \end{bmatrix} \end{bmatrix} \\ LEX- \end{bmatrix} \\ SEM\ \begin{bmatrix} high-deg-rel \\ ARG\ d \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

[10] In SBCG, Kay and Sag 2012 analysed them as functors (Van Eyde 2007), with a SEL feature instead of a MOD feature, triggering a MARKING value *deg*. As far as we can see, nothing would be changed if we adopt this analysis, as long as they may also be complements in the verbal domain.

mari

$$\begin{bmatrix} SYN & \begin{bmatrix} HEAD & \begin{bmatrix} adv \\ MOD\ SEM & \begin{bmatrix} SCALE\ intens-s \\ VALUE\ d \end{bmatrix} \end{bmatrix} \\ & LEX + \end{bmatrix} \\ SEM & \begin{bmatrix} high-deg-rel \\ ARG\ d \end{bmatrix} \end{bmatrix}$$

pli

$$\begin{bmatrix} SYN & \begin{bmatrix} HEAD & \begin{bmatrix} adv \\ MOD & \begin{bmatrix} SYN & \begin{bmatrix} HEAD\ non-verbal \\ EXTRA\ \ L \end{bmatrix} \\ SEM & \begin{bmatrix} SCALE\ intens-s \\ VALUE\ d1 \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ & EXTRA\ \ L + \left\langle \begin{bmatrix} SYN[MARKING\ ki] \\ SEM\ [VALUE\ d2] \end{bmatrix} \right\rangle \end{bmatrix} \\ SEM & \begin{bmatrix} super-rel \\ ARG1\ d1 \\ ARG2\ d2 \end{bmatrix} \end{bmatrix}$$

Comparative adverbs

The comparative adverbs have two arguments: the degree of the predicate they modify and the degree of their *ki* ('than') marked dependent (using Kay and Sag's 2012's EXTRA feature). They also exhibit a double selection – both syntactic and semantic, where *pli* ('more') and *mwin* ('less') combine with non-verbal categories associated with an intensity scale, whereas *plis* and *mwins* combine with verbs, and non-verbal categories associated with a quantity scale.

plis

$$\begin{bmatrix} SYN & \begin{bmatrix} HEAD & \begin{bmatrix} adv \\ MOD & \begin{bmatrix} SYN & \begin{bmatrix} HEAD\ verb \\ EXTRA\ \ L \end{bmatrix} \\ SEM & \begin{bmatrix} SCALE\ s \\ VALUE\ d1 \end{bmatrix} \end{bmatrix} \end{bmatrix} V \begin{bmatrix} SYN\ [EXTRA\ L] \\ SEM & \begin{bmatrix} SCALE\ quant-s \\ VALUE\ d1 \end{bmatrix} \end{bmatrix} \\ & EXTRA\ \ L + \left\langle \begin{bmatrix} SYN\ [MARKING\ ki] \\ SEM\ [VALUE\ d2] \end{bmatrix} \right\rangle \end{bmatrix} \\ SEM & \begin{bmatrix} super-rel \\ ARG1\ d1 \\ ARG2\ d2 \end{bmatrix} \end{bmatrix}$$

4.2 Degree adverbs in the non verbal domain

In Mauritian, count and mass nouns can be used as bare NPs and are associated with a quantity scale. With nouns and other non verbal categories, we analyse degree adverbs as optional adjuncts, and not specifiers, as they do not have the same distribution as possessive or demonstrative determiners.

(43) a. Mo'nn aste lafarinn pou fer gato
'I bought flour to make a cake/ cakes'
b. boukou dilo 'a lot of water'
c. mwins rob 'less dresses'
d. plis lafarinn 'more flour'

Adverbial adjuncts precede nouns and non verbal categories (*pli vit* 'more quickly', *tro gran* 'too big', *mari long* 'very long', *boukou liv* 'a lot of books'):

Adjunct < Head [HEAD non-verbal]

4.3 Degree adverbs in the verbal domain

Mari precedes the verb, *boukou* follows it, and *plis* and *mwins* either precede or follow it. We use the feature LEX (Arnold and Sadler 1994) as for the English adjectives to distinguish *mari* from the others. Most degree adverbs are unspecified for LEX, while *mari* is [LEX +] and *boukou* [LEX -].

Degree adverbs as complements of verbs
The short verbal form implies the presence of a (non clausal) complement.[11]

Lexical constraint on Mauritian verbs (Henri 2010):
verb
[HEAD [VFORM short]] => [VAL [COMPS non-empty-list]]

We have seen that a postverbal degree adverb is enough to trigger the short verbal form, and we thus analyse it as a complement. As shown by Henri 2010, manner adverbs have the same property:

(44) a. Paul manz bien.
Paul eat-SF well 'Paul eats well'
b. Paul roul vit.
Paul drive-SF fast 'Paul drives fast'

Degree and manner adverbs can thus optionally be added to the COMPS list of verbs (Abeillé & Godard 2003), except *pli* ('more'), *mwin* ('less'), which cannot modify verbs, and *mari, which* is [LEX +].[12]

Lexical rule for Mauritian verbs

$$\begin{bmatrix} verb \\ SYN \begin{bmatrix} HEAD\ [0] \\ VAL \mid COMPS\ \ L \end{bmatrix} \\ SEM\ [1] \end{bmatrix} \Rightarrow \begin{bmatrix} SYN \mid VAL \mid COMPS\ \ L + \langle \begin{bmatrix} deg - man - adv \\ MOD\ \begin{bmatrix} SYN \mid HEAD\ [0] \\ SEM\ [1] \end{bmatrix} \\ LEX - \end{bmatrix} \rangle \end{bmatrix}$$

11 Henri 2010 analyses clausal complements as extraposed.

12 Degree adverbs can also head a verbless copular clause (Henri & Abeillé 2007) : *3000 roupis, li boukou* ('3000 roupies, it is a lot'), *Sa liv la plis* ('This book is more'). A lexical construction can be used to provide them with a SUBJ feature (Müller 2009).

As complements, they follow the verb and trigger the short verbal form: HEAD < COMPS

As oblique complements, they follow the direct complements (45a) and may permute with other oblique complements (45b)

(45) a. Paul zwenn so frer boukou.
Paul meet-SF POSS.3sg a-lot
'Paul meets his brother a lot.'

b. Paul koz [ar so frer] boukou / boukou [ar so frer].
Paul speak-SF with POSS.3sg brother a-lot/ a-lot with POSS.3sg brother
'Paul speaks with his brother a lot / a lot with his brother.'

Degree adverbs as adjuncts

They are adjuncts when they precede the verb. Like two manner adverbs (*bien* 'well', *mal* 'badly'), and a few other adverbs (*nek* 'only', *fek* 'just')), they occur between the TAM marker (*pe, inn, pou*…) and the lexical verb (Henri & Kihm forthcoming).

(46) a. Paul inn mal / mari travay.
Paul PERF badly/ very work
'Paul has worked badly / a lot'

b. Paul pe nek plore.
Paul PROG only cry- SF
'Paul is only crying'

In preverbal position, we consider them as [LEX+] (Arnold & Sadler 1994): they can be premodified (*boukou plis 'a lot more'*) but cannot head a full adverbial phrase (**plis ki pou twa* 'more than yours') (28). They adjoin to the lexical verb, except *pli* ('more'), *mwin* ('less'), which cannot modify verbs, and *boukou* ('a lot'), which is [LEX -] :

Adjunct [LEX +] < Head $\begin{bmatrix} verb \\ LEX + \end{bmatrix}$

Other adverbs adjoin to VP (follow the verb and its complements) or S (are sentence initial or sentence final), like *vit* ('quickly'), *yer* ('yesterday').

(47) a. *Paul inn vit / yer travay.
Paul PERF fast/ yesterday work

b. Paul inn travay so prezantasion vit.
Paul PERF travay POSS.3sg presentation fast
'Paul worked fast on his presentation.'

c. Yer Paul inn travay / Paul inn travay yer.
'Yesterday Paul worked / Paul worked yesterday'

5 Conclusion

In Mauritian, an understudied French-based creole, degree words like *tro* ('too') should be analysed as polymorphic adverbs, combining with all the major categories. From a syntactic point of view, they are complements after the verb and adjuncts when they precede it. With other categories, they are always adjuncts.

Mauritian also has a pair of high degree adverbs (*boukou* - *mari*) in complementary distribution: while *boukou* follows the verb as a complement, *mari* precedes it as an adjunct. From a semantic point of view, *boukou* selects predicates with a quantity scale, and *mari* predicates with an intensity scale.

Comparatives also come in pairs (*pli, plis,* 'more'; *mwin, mwins,* 'less'*)* in a complementary distribution: *plis, mwins* select verbs with any scale and predicates with a quantity scale while *pli, mwin* combine with non verbal predicates with an intensity scale. Like *tro, plis, mwins* are complements after the verb and adjuncts when they precede it.

We have provided new arguments for the adverb as complement analysis proposed in HPSG by Bouma et al. 2001, Kim and Sag 2003, and the polymorphic analysis proposed for French adverbs by Abeillé & Godard 2003, using HPSG underspecification.

References

Abeillé, A., Godard, D. 2003. The Syntactic Flexibility of French Degree Adverbs. In Stefan Müller (ed.) *Proceedings of the 10th international conference on Head-Driven Phrase Structure Grammar*. *Michigan State University,* 26–46. Stanford CA: CSLI Publications.

Abeillé, A, J. Doetjes, and H. de Swart. 2004. `Adverbs and quantification', in F. Corblin, H. de Swart (eds) *Handbook of French semantics*, CSLI, Stanford.

Alleesaib M. 2012. Bann en créole mauricien: nom de quantité et morphème de pluralité, *Proceedings CMLF, Lyon.*

Arnold D., L. Sadler 1994, Prenominal adjectives and the phrasal /lexical distinction, *Journal of Linguistics*, 30:1, 187-226.

Baker, Philip. 1972. Kreol: A description of Mauritian Creole. Ann Arbor: Karoma.

Bouma, G., R. Malouf, I. A. Sag. 2001. Satisfying constraints on extraction and adjunction, *Natural Language and Linguistic Theory* 19: 1-65.

A. Carpooran 2009, *Diksioner Morisien*, Koleksion Text Kreol.

Henri F., Abeillé A. 2007. The syntax of copular constructions in Mauritian. In Stefan Müller (ed.), *Proceedings of the 14th international*

conference on Head-Driven Phrase Structure Grammar, *Stanford Department of Linguistics and CSLI's LinGO Lab,* 130-149. Stanford, CA: CSLI Publications.

Henri F., Abeillé A. 2008. Verb Form alternations in Mauritian. In Stefan Müller (ed.). *Proceedings of the 15th international conference on Head-Driven Phrase Structure Grammar, National Institute of Information and Communications Technology, Keihanna,* 378-398. Stanford CA: CSLI Publications.

Henri F. 2010. A constraint-based approach to verbal constructions in Mauritian: morphological, syntactic and discourse-based aspects, PhD Thesis, University Paris Diderot.

Henri F., A. Kihm (to appear), The morphology of TMA marking in creole languages, a comparative study, *Word Structure*.

Hookoomsing, V. Y. 2004. *Grafi larmoni*. Ministry of Education, Mauritius.

Kay P., Sag I. 2012, Cleaning up the big mess, in H. Boas & I. Sag (eds) *Sign-based construction grammar*, CSLI Publications, 229-256.

Kennedy C. & McNally L. 2005. Scale Structure, degree modification, and the semantics of gradable predicates. *Language* 81, 345-381.

Kim J-B, I. A. Sag 2003. Negation without Head Movement. *Natural Language and Linguistic Theory* 20.2: 339-412

Müller S. 2009. On Predication. In S. Müller (ed.), *Proceedings of the 16th international conference on Head-Driven Phrase Structure Grammar, University of Göttingen, Germany,* 213-233. Stanford CA: CSLI Publications.

Van Eynde F. 2007. The Big Mess Construction. In Stefan Müller (ed), *Proceedings of the 14th international conference on Head-Driven Phrase Structure Grammar, Stanford Department of Linguistics and CSLI's LinGO Lab,* 415-433. Stanford, CA: CSLI Publications.

Obligatory control and event structure in Kavalan

Dong-yi Lin
Ghent University

Proceedings of the 21st International Conference on
Head-Driven Phrase Structure Grammar

University at Buffalo

Stefan Müller (Editor)

2014

Stanford, CA: CSLI Publications

pages 280–298

Lin, Dong-yi. 2014. Obligatory control and event structure in Kavalan. In Stefan Müller (ed.), *Proceedings of the 21st International Conference on Head-Driven Phrase Structure Grammar, University at Buffalo*, 280–298. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2014.15. CC-BY 4.0

Abstract

The embedded verb of so-called object-control verbs in Kavalan must be affixed with the causative marker *pa-*. It is argued that such control predicates in Kavalan like *pawRat* 'force' feature an internal Logophoric Center in its complement clause and this property of logophoricity is absent in other control predicates. Moreover, control predicates that do not take a causativized verb complement like *paska* 'try' and *tud* 'teach' are restructuring predicates and are thus devoid of a Fin head in their complement that can be linked to an internal Logophoric Center. In contrast, the TP and CP of the complement of *pawRat* 'force'-type predicates are still projected and active. The causativization of the embedded verb in a control sentence cannot be explained by a purely syntactic or semantic account of obligatory control. Instead, a comprehensive and satisfactory explanation for Kavalan obligatory control must take into account how event structure and Logophoric Center are encoded in Syntax.

1 Introduction

The present paper investigates the control structure of Kavalan, an Austronesian language in Taiwan. The structure of control sentences is one of the principal issues in the generative grammar approach to syntax. One central goal in the study of control sentences is to explain why a DP can semantically form two separate thematic relations with two different verbs. The standard account in the Government and Binding (GB) framework proposes a phonetically null nominal element, PRO, to explain this property of control sentences. For example, in both (1a) and (1b), it is the PRO that is assigned a theta role by the embedded verb. The co-reference between *Bill* and PRO in (1a) or between *Ryan* and PRO in (1b) results from the syntactic operation of binding.

(1) a. Bill_i tried [PRO_i to leave].
b. Michael persuaded Ryan_i [PRO_i to buy the car].

The postulation of PRO raises two important theoretical questions regarding the distribution of PRO and its interpretation. It has been suggested that PRO can only occur in the subject position of a non-finite clause. The GB account reduces this distributional constraint to the PRO Theorem, which states that a PRO can only occur in an ungoverned position. An alternative analysis argues that a PRO bears null Case and that it must occur in a position where its null Case can be checked (Chomsky and Lasnik 1993). As for the interpretation of a PRO, the Minimal Distance Principle (MDP) first postulated by Rosenbaum (1967) has been regarded as the underlying principle that determines the controller of a PRO. This principle states that a PRO is controlled by the closest c-commanding DP. Therefore, the PRO in (1b) must be controlled by *Ryan* instead of *Michael*.

† This research was funded by FWO: 2009-Odysseus-Haegeman-G091409.

In recent years, there have been attempts to eliminate the Control module from Universal Grammar and reduce the control mechanisms to movement operations (Hornstein 1999; Boeckx, Hornstein, and Nunes 2010). This line of research, i.e., the Movement Theory of Control (MTC), is motivated by the elimination of the D-Structure in the Minimalist Program. Without the D-Structure, the theoretical validity of the Theta-Criterion is cast in doubt and thus the theoretical motivation for PRO or the entire Control module is untenable.

Under the Movement Theory of Control, the co-reference between an argument of a control predicate and an argument in its complement clause is not mediated by a PRO and the Minimal Distance Principle. Instead, a control predicate is akin to a raising predicate in that they both involve movement of a DP in the embedded clause to the matrix clause. The crucial difference between them is that the target of movement in a control sentence is a Θ-position, whereas the moved DP in a raising sentence targets a Θ'-position. It is argued that this reductionist analysis can not only achieve theoretical parsimony but also allow for wider empirical coverage, e.g., backward control and copy control, as opposed to forward control (Polinsky and Potsdam 2006). Moreover, as any movement must obey the locality condition, this analysis can capture the same effects of the Minimal Distance Principle without any extra stipulations.

Another strand of syntactic analysis attributes obligatory control to the relationship between finiteness features and the ability to license a referentially (in)dependent person feature (+R or -R), e.g., Landau's (2004) R-assignment Rule in (2). On this analysis, the co-reference between a controller DP and a PRO is subject to Agree between the matrix F licensing the controller DP and the PRO or the embedded Agr head. Bianchi (2003) also argues that finiteness is the crucial conditioning factor for obligatory control. She links finiteness to logophoric anchoring, which can determine the licensing of [+R] or [-R] person feature. On her analysis, the Fin head in a complement clause can be linked to either an external LC, i.e., speech participants, or an internal LC, i.e., the participants of the matrix clause event. A [-finite] Fin head linked to an internal LC can only license a referentially dependent [-R] person feature.

(2) R-assignment Rule (Landau 2004: 842)
For $X^0_{[\alpha T, \beta Agr]} \in \{I^0, C^0 \ldots\}$:
$\varnothing \rightarrow [+R]/X^0_{[_]}$, if $\alpha = \beta =$ '+'
$\varnothing \rightarrow [-R]$/elsewhere

Still another primary analysis is to resort to semantic principles and lexical semantics of control predicates as a way of explanation (Jackendoff and Culicover 2003; Sag and Pollard 1991). The semantic analysis of control contends that a purely syntactic treatment of control sentences can never offer a satisfactory and comprehensive explanation for control configurations. On this analysis, the controller of the implicit argument in the embedded clause of a control predicate is determined by the semantics of the control predicate, e.g., its conceptual structure (Jackendoff and Culicover 2003).

According to Chang and Tsai (2001), the structure of control sentences in some Formosan languages is distinct from the English control structure in

that the verb in the subordinate clause of a control sentence must take a causative marker, as illustrated by the following Kavalan sentence.[1]

(3) pawRat-an-na ni buya$_i$ aiku$_k$ [pa-qibasi __$_i$ __$_k$ tu
force-PV-3ERG ERG Buya 1SG.ABS CAUS-wash OBL
qudus]
clothes
'Buya forces me to wash clothes.' (lit. 'Buya forces me, causing (me) to wash clothes.)

Chang and Tsai (2001) attribute this control configuration to a semantic constraint called the Actor-Sensitivity Constraint, which states that control operations are sensitive to the agent argument of a control predicate. The present paper provides and discusses new data that cannot be explained by the Actor-Sensitivity Constraint. It is argued that the syntactic structures of Kavalan control sentences reflect their event structures regarding the perspective from which the event is reported.

The present paper will present new empirical evidence from Kavalan that suggests that the Actor-Sensitivity Constraint is empirically inadequate. It will be argued that the syntactic control configurations of Kavalan correspond to event structures of control predicates and that the event structure of control predicates is indirectly encoded in Syntax through the linking of embedded Fin(iteness) head to a Logophoric Center (LC). The article is organized as follows. Section 2 discusses the empirical patterns of obligatory control sentences in Kavalan. It will be shown that not all control sentences in Kavalan observe the Actor-Sensitivity Constraint. How to explain the control configurations in Kavalan is discussed in sections 3, 4, and 5. It will be argued that a purely syntactic or semantic analysis cannot explain the control constructions in Kavalan. A comprehensive and satisfactory explanation for Kavalan obligatory control must take into account how event structure and Logophoric Center are encoded in Syntax. Section 6 concludes the study.

2 Obligatory Control in Kavalan

2.1 Controller of PRO

The interpretation of the PRO in control sentences in Kavalan is not associated with a DP that takes a specific case marker. As illustrated in (4), in a *try*-type control sentence in Kavalan, the PRO is always co-referential with the agent of the matrix verb, i.e., the only DP argument in the matrix clause, regardless of its grammatical role or case marking.

(4) a. m-paska ya sunis$_i$ [satzai PRO$_i$]
AV-try ABS child sing
'The child tries to sing.'

1 Glossing conventions in this paper follow the Leipzig Glossing Rules. Additional glossing conventions are as follows: AV – Agent voice; CONJ – Conjunction; LNK – Linker; NCM – Non-common noun marker; PV – Patient voice.

b. paska-an na sunis$_i$ [satzai PRO$_i$]
try-PV ERG child sing
'The child tries to sing.'

As discussed briefly in section 1, a *persuade*-type control sentence in Kavalan is typologically unique in that the embedded verb must take an overt causative prefix *pa-*. The causativization of the embedded verb in a *persuade*-type control sentence can also be observed in other Formosan languages, e.g., Budai Rukai, Puyuma, and Tsou (Chang and Tsai 2001). Chang and Tsai (2001) argue that this is because control verbs in these Formosan languages have to observe a constraint called Actor-Sensitivity, which stipulates that only an agent DP argument can control a PRO. The following Kavalan sentences are for illustration.

(5) a. m-linana aizipna$_i$ tu sunis [*(pa)-lusit PRO$_i$]
AV-persuade 3SG.ABS OBL child CAUS-leave
'He persuades a child to leave.' (lit. 'He persuades a child, causing (him) to leave.')
b. pawRat-an-na ni buya$_i$ aiku [*(pa)-qibasi PRO$_i$
force-PV-3ERG ERG Buya 1SG.ABS CAUS-wash
tu qudus]
OBL clothes
'Buya forces me to wash clothes.' (lit. 'Buya forces me, causing (me) to wash clothes.')

In both (5a) and (5b), the embedded verb has to be prefixed with the causative marker *pa-*. The PRO in the embedded clause does not correspond to the theme argument in the matrix clause, but is consistently co-referential with the agent DP, which functions semantically as the causer in the embedded clause. Grammatical roles and case marking do not determine the interpretation of PRO in a *persuade*-type control sentence. In (5a), it is the absolutive DP that controls the PRO; in (5b), it is the ergative DP that controls the PRO. In both cases, the controller is the agent argument of the matrix control verb.

The empirical facts that motivate Chang and Tsai's (2001) proposal of the Actor-Sensitivity Constraint, however, are not entirely correct. Not all control verbs in Kavalan obey the Actor-Sensitivity Constraint. Control verbs like *sulud* 'allow', *tabal* 'stop; prevent', *pangmu* 'help', and *tud* 'teach' do not require their embedded verbs to take the causative marker *pa-*.

(6) a. sulud-an-na ni abas ya sunis 'nay$_i$
allow-PV-3ERG ERG Abas ABS child that
[mawRat PRO$_i$]
AV.play
'Abas allows that child to play.'
b. t<m>abal=iku tu sunis$_i$ [q<m>an PRO$_i$
<AV>stop=1SG.ABS OBL child <AV>drink
tu Raq]
OBL alcohol
'I stop a child from drinking.'

c. pangmu-an-na ni abas aiku$_i$ [m-kyala PRO$_i$
help-PV-3ERG ERG Abas 1SG.ABS AV-pick.up
tu byabas]
OBL guava
'Abas helps me pick up guavas.'

d. tud-an-na=iku$_i$ na tina-ku
teach-PV-3ERG=1SG.ABS ERG mother-1SG.GEN
[s<m>udad PRO$_i$]
<AV>write
'My mother teaches me to write.'

In each sentence in (6), the PRO in the embedded clause is controlled by the theme DP argument in the matrix clause. The verb in the subordinate clause does not take the causative marker and no argument is co-referential with the matrix agent argument. If the embedded verb in this type of sentence takes the causative marker *pa-*, the matrix theme argument will still be construed as the controller, as illustrated below.

(7) sulud-an-ku ya sunis-ku$_i$ [pa-qawRat PRO$_i$]
allow-PV-1SG.ERG ABS child-1SG.GEN CAUS-play
'I allow my child to let (someone) play.'

In (7), the causative marker is attached to the embedded verb and the causer in the subordinate clause must be controlled by the matrix theme argument instead of the matrix agent argument.

What underlies the obligatory control pattern in Kavalan is thus not the Actor-Sensitivity Constraint. Not all *persuade*-type control verbs, or the so-called object-control verbs, in Kavalan, take a morphologically causativized verb phrase as a complement. Whether the agent or the theme argument can serve as the controller varies from a control verb to another. In order to provide a comprehensive and satisfactory account for the obligatory control pattern in Kavalan, it is thus imperative to identify the common feature(s) shared by the control verbs that require a morphologically causativized verb in their complement clause. The following examples illustrate other control verbs whose complement clause must contain a verb affixed with *pa-*.

(8) a. tezung-an-na ni utay ti-abas *(pa-)qibasi
instruct-PV-3ERG ERG Utay NCM-Abas CAUS-wash
tu qudus
OBL clothes
'Utay instructs Abas to do the laundry.'

b. tuluz-an-na ni buya aiku *(pa-)qapaR tu
send-PV-3ERG ERG Buya 1SG.ABS CAUS-catch OBL
mutun
mouse
'Buya sends me to catch a mouse.'

c. pupuk-an-ku ya sunis-ku *(pa-)taqsi
ask-PV-1SG.ERG ABS child-1SG.GEN CAUS-study
'I ask my child to study.'

d. qeRas-an-na ni imuy aiku *(pa-)tenun
require-PV-3ERG ERG Imuy 1SG.ABS CAUS-weave
'Imuy requires me to weave.'

The control verbs that require their embedded verbs to be affixed with the causative prefix all denote an event where the agent obligates the theme to perform some action. The attempt to compel someone to do something constitutes an indispensable part of the semantics of these verbs: *linana* 'persuade', *pawRat* 'force', *tezung* 'instruct', *tuluz* 'send', *pupuk* 'ask', and *qeRas* 'require'. They only differ in the degree of coercion and the way how the agent places the theme under an obligation to perform a task. Control verbs that do not take a causativized verb phrase as a complement do not encode an attempt to compel someone to do something. For example, *sulud* 'allow' denotes a scenario where the agent consents to the theme's action without obligating the theme to perform the action. The verb *tabal* 'stop; prevent' encodes the opposite meaning of *compel*, *force*, or *require*. Instead of placing the theme under an obligation to execute an action, the agent of *tabal* intends for the theme to not assume the obligation.

To summarize, contrary to what Chang and Tsai (2001) claim, the theme argument of some control verbs in Kavalan can control the PRO in the subordinate clause. The Actor-Sensitivity Constraint cannot account for the interpretation of PRO in Kavalan. Whether a PRO is controlled by the agent or the theme argument of the matrix control verb is contingent on the semantics of the control verb. Only verbs that encode an attempt to obligate someone to do something are required to take a *pa*-marked subordinate verb.

2.2 The Causative Marker *Pa-*

As the affixation of the causative marker *pa-* to the embedded verb of certain control predicates is obligatory, an investigation of the grammatical properties of *pa-* is essential to the structural analysis of obligatory control constructions in Kavalan.

The causative marker *pa-* in Kavalan is fully productive, being able to attach to almost any verb and thereby transform it into a causative verb. It can occur either in an agent voice construction (9b) or a patient voice construction (8d).

(9) a. qibasi tu qudus ya ti-imuy
wash OBL clothes ABS NCM-Imuy
'Imuy does the laundry.'

b. pa-qibasi tu qudus ya ti-abas ti-imuy-an
CAUS-wash OBL clothes ABS NCM-Abas NCM-Imuy-OBL
'Abas makes Imuy do the laundry.'

c. qaway-an-na ni utay ya beRas
carry-PV-3ERG ERG Utay ABS rice
'Utay carries rice.'

d. pa-qaway-an-na ni ipay ya ti-utay tu
CAUS-carry-PV-3ERG ERG Ipay ABS NCM-Utay OBL
beRas
rice
'Ipay makes Utay carry rice.'

The affixation of the causative marker introduces an additional causer argument. In an agent voice construction, the causer argument takes the absolutive case marker, e.g., *ya ti-abas* in (9b), while the causee takes the oblique case marker, e.g., *ti-imuy-an* in (9b). In a patient voice construction, the ergative DP is interpreted as the causer, e.g., *ni ipay* in (9d), while the absolutive DP is interpreted as the causee, e.g., *ya ti-utay* in (9d).

The causer argument introduced by *pa-* does not need to be an agentive DP. This is illustrated by the following example.

(10) pa-Rubatang ya iyu tu tazungan
CAUS-be.beautiful ABS medicine OBL woman
'The medicine makes women (become) beautiful.'

This suggests that the causative marker *pa-* thematically introduces a generic causer, not an agent argument.

Assuming a syntactic approach to causative affixes (Baker 1988; Harley 2008), I analyze *pa-* in Kavalan as the lexical realization of v_{CAUSE}, which assigns a generic causer role to the DP in its specifier position. Its complement is another vP with an independent argument structure. The complement vP of v_{CAUSE}, or *pa-*, serves as the event argument of the causative marker, i.e., the event that the causer brings about. The structure is represented below.

(11)[$_{vPcause}$ CAUSER [$_{v'cause}$ v_{CAUSE} *pa-* [$_{vP}$ DP [$_{v'}$ v [$_{VP}$ V DP]]]]]

3 Against Syntactic Treatments of Kavalan Obligatory Control

The causativization of the embedded verb in Kavalan control sentences presents a problem for analyses that attribute the distribution and interpretation of the phonetically null argument in a control complement to purely syntactic principles and operations, e.g., the standard PRO theory of obligatory control and the Movement Theory of Control (MTC) (Hornstein 1999). A purely syntactic analysis cannot explain why some "object-control" verbs take a causativized verbal complement but others don't. On the standard PRO analysis, the embedded verb in a Kavalan *persuade*-type sentence is causativized and thus there should be a causer PRO and a causee PRO in accordance with the Theta-Criterion. The fact that there are two PROs that need to be bound by two different DPs creates a problem for the Minimal Distance Principle (Rosenbaum 1967). The structurally closest DP that c-commands the two PROs is the theme DP in the matrix clause, but only the causee PRO is actually controlled by it.

Neither can the MTC provide a satisfactory account for the control patterns in Kavalan. On the MTC analysis, the unexpressed arguments in (3), (5), and (8) would be analyzed as the traces or copies of the two DPs in the matrix clause. To move both the causer DP and the causee DP to the matrix clause would incur a violation of the Minimal Link Condition regardless of the order of their movement. As the movement of both DPs is motivated by theta-feature checking on this analysis, the higher causer DP will always block the movement of the lower causee DP. Note that the MTC allows a DP

to acquire more than one theta role, so there is no limit on the number of theta roles the causer DP can receive.

4 Semantic Analysis

4.1 Types of Control Verbs and Their Conceptual Structures

Jackendoff and Culicover (2003) argue that the interpretation of the covert argument in the embedded clause of a control predicate is not determined by any syntactic principles or conditions like the Minimal Distance Principle or the locality condition on movement. The following sentences show that the syntactic position of the DP that is co-referential with the covert argument in a control sentence is irrelevant to its interpretation.

(12) Jackendoff and Culicover (2003: 520)
- a. John$_i$ persuaded Sarah$_j$ to $_{j/*i}$dance.
- b. John$_i$ promised Sarah$_j$ to $_{i/*j}$dance.
- c. John$_i$ talked about $_{i/gen}$dancing with Jeff.
- d. John$_i$ refrained from $_{i/*gen}$dancing with Jeff.
- e. Bill ordered Fred$_i$ to $_i$leave immediately.
- f. Fred$_i$'s order from Bill to $_i$leave immediately.
- g. the order from Bill to Fred$_i$ to $_i$leave immediately.
- h. Fred$_i$ received Bill's order to $_i$leave immediately.

The so-called control relationship between an overt DP and a covert DP is contingent on the meaning/semantics of the control predicate.

Jackendoff and Culicover (2003) thus argue that only a semantic analysis can fully account for obligatory control. On their semantic analysis, the obligatory control relation should be encoded in the conceptual structure (CS) of a verb instead of its syntactic structure. Within their framework, the conceptual structure is the level that structurally represents thematic roles and their relationships. For example, *intend* and *decide* exhibit the conceptual structure in (13). The DP argument that refers to the intender is always the controller of the actional complement. A verb that inherently denotes an obligation on the part of the theme argument is assigned either of the two conceptual structures in (14).

(13) Jackendoff and Culicover (2003: 537)
X^{α} INTEND *[α ACT]*

(14) Jackendoff and Culicover (2003: 537)
- a. X^{α} OBLIGATED *[α ACT]* TO Y
- b. X^{α} OBLIGATED *[α ACT]*$^{\beta}$
  β BENEF Y

In the two conceptual structures in (14), X bears an obligation to execute a certain action, as indicated by the α notation. English verbs that exhibit this type of conceptual structure include *order*, *instruct*, and *promise*. An obligation is imposed on the theme argument of both *order* and *instruct* and thus the one that receives the order or instruction (X^{α}) is always interpreted as

the controller of the implicit argument in the embedded clausal complement ([α ACT]), regardless of its syntactic position, as illustrated in (12e) – (12h). As for *promise*, it is the person making a promise that undertakes an obligation to execute some action and thus it is always the person making a promise that is interpreted as the controller. The syntactic position of the controller DP and the syntactic distance between the controller DP and the controlled implicit DP are irrelevant to obligatory control.

Another class of control predicates discussed by Jackendoff and Culicover (2003) involves force dynamics and includes verbs like *cause*, *force*, *prevent*, *enable*, and *help*. They all describe a scenario where one character attempts to influence another character's execution of an action and exhibit the following generic conceptual structure.

(15) Jackendoff and Culicover (2003: 538)
X CS Y^{α} *[α ACT]*

The character that is influenced is represented as Y^{α} and its action is represented as [α ACT]. The α notation signals the control relationship between the two. The verbs differ in the types of influence that is exerted, e.g., causing, forcing, preventing, helping, and enabling.

4.2 Syntax and Event Structure of Control Verbs in Kavalan

The semantic analysis based on Jackendoff and Culicover (2003) makes a wrong prediction about the control configuration of *order*-type or *force*-type control verbs in Kavalan. According to the conceptual structure in (14), the control relationship holds between the DP argument that bears an obligation and the actional complement, not between the DP argument that imposes such an obligation and the actional complement. However, a Kavalan control sentence headed by *tezung* 'instruct' manifests both control relationships.

Even if the conceptual structure in (14) can be modified so as to reflect the control relationship between the agent argument of the control predicate and the causer argument in the actional complement, the semantic analysis still fails to explain the different control configurations among the predicates that are identified as force dynamic predicates by Jackendoff and Culicover (2003). Verbs like *force*, *prevent*, *help*, and *permit* all exhibit the conceptual structure in (15) on Jackendoff and Culicover's (2003) analysis. It is thus expected that they should all manifest the same control configurations in Kavalan, contrary to fact. As discussed in section 2.1, *pawRat* 'force' in Kavalan requires its embedded verb to take the causative marker, whereas *sulud* 'allow', *tabal* 'prevent', and *pangmu* 'help' do not. The control configuration of *sulud* 'allow', *tabal* 'prevent', and *pangmu* 'help' follows from the conceptual structure in (15) in that the theme argument is identified as the controller. The control configuration of *pawRat* 'force' is not entirely compatible with this conceptual structure as the agent theta role of this predicate is also identified as a controller in syntax.

Owing to these problems, I argue that the syntactic control configurations of different Kavalan control predicates constitute the grammatical encoding of different control event structures.[2] The analysis lends support to theories

2 Rooryck's (2008) analysis of obligatory control resorts to the event structure of

that incorporate lexical semantics and event structure into the construction of syntactic structure (Rosen 2003; Travis 2000).

I concur with Jackendoff and Culicover (2003) and Rooryck (2008) in their contention that the semantics and event structure of control predicates plays a significant role in determining the control configurations. Their specific proposals, however, need to be modified in order to account for the Kavalan facts of obligatory control. As exemplified in section 2.1, Kavalan control verbs that exhibit the Actor-Sensitivity phenomenon all depict a scenario where the agent attempts to bring about an event by imposing an obligation on someone else to execute the action or simply by forcing someone else to carry out the action. The event described by such control verbs actually consists of two sub-events. The first sub-event involves the agent's act and the second sub-event is the execution of an action by someone else. Moreover, the two sub-events are connected by a cause-result relation, or a CAUSE/BRING ABOUT operator. The event structure of such control verbs in Kavalan is represented below.

(16) $[X^{\alpha}$ ACT (ON $Y^{\beta})]$ $^{\alpha}$BRING ABOUT $[Y^{\beta}$ ACT]

This event structure is mapped to the syntactic structure of *pawRat* 'force', *tezung* 'instruct', and other verbs that involve the same event structure. The cause-result relation between the two sub-events is mediated in Syntax by the specific v head, v_{CAUSE}, or *pa-*, which is the syntactic realization of the semantic CAUSE/BRING ABOUT operator.[3]

The addition of the causative marker in Kavalan *persuade*-type control sentences is thus tied to the event structure in (16) only. Control predicates that do not exhibit the event structure in (16) will not take vP_{CAUSE} as the complement and their embedded verb will not undergo morphological causativization. However, on this semantic analysis of Kavalan control predicates, it is not clear how the event structure of a control predicate is mapped to its syntactic structure. In other words, how is the event structure of a control predicate linked to Syntax? To complement the semantic analysis, section 5 will argue that the event structure of control predicates is encoded in Syntax through the linking of embedded Fin(iteness) head to a Logophoric Center (LC).

5 Control Predicates and Logophoricity

According to Bianchi (2003), the Fin head in a complement clause can be linked to either an external LC, i.e., speech participants, or an internal LC, i.e., the participants of the matrix clause event. She proposes that a Fin head linked to an internal LC can only license a referentially dependent [-R]

control predicates, but it cannot extend to Kavalan. The subevent structure he proposes for *force* cannot explain why the embedded verb of *pawRat* 'force' must be affixed with the causative marker *pa-*.

3 One problem of the event structure in (16) is that it seems to assume that the actional complement is always realized. Whether this is true in Kavalan requires more research.

person feature and argues that obligatory control is a concomitant consequence of this structural licensing.

(17) … V [$_{CP}$ [$_{FinP}$ -Fin [$_{TP}$ DP$_{-R}$ [$_{T}$ -T] …]]]
| |
E_i Int LC_i

5.1 *PawRat* 'force'-type Control Predicates and Logophoricity

Adopting Bianchi's (2003) conception of logophoricity, the present paper argues that a *pawRat* ('force')-type control predicate in Kavalan features an internal LC in its complement clause and this property of logophoricity is absent in other control predicates.

Firstly, when the complement of *pawRat* 'force' is negated, the imperative negator *naRin*, instead of the indicative negator *mai*, is used, as illustrated in (18). By contrast, the complement clause of *paska* 'try' and *tud* 'teach' cannot be negated by either negator, as shown in (19).

(18) a. pawRat-an-na=iku ni utay naRin
force-PV-3ERG=1SG.ABS ERG Utay NEG.IMP
m-qila tu sunis
AV-scold OBL child
'Utay forces me to not scold children.'
b. *pawRat-an-an=iku ni utay mai m-qila
force-PV-3ERG=1SG.ABS ERG Utay NEG AV-scold
tu sunis
OBL child

(19) a. *paska=pa=iku mai/naRin m-qila tu sunis
try=FUT=1SG.ABS NEG/NEG.IMP AV-scold OBL child
b. *tud-an-na ni utay ya sunis-na
teach-PV-3ERG ERG Utay ABS child-3GEN
mai/naRin m-tebu tu qudus
NEG/NEG.IMP AV-patch OBL clohes

Secondly, the complement of *pawRat* 'force' can be a direct quotation of imperative, as illustrated in (20). Other control predicates cannot take an imperative clause as their complement (21). This suggests that *pawRat* 'force'-type control predicates denote a speech event, a potential internal LC. Predicates denoting speech, thought, knowledge, and direct perception are common predicates that can license a logophoric clause (Culy 1994). The controller of *pawRat* 'force'-type control predicates is always the internal Speaker of the internal LC.

(20) pawRat-an-na ni utay ti-imuy, qibasi-ka tu
force-PV-3ERG ERG Utay NCM-Imuy wash-IMP OBL
qudus
clothes
'Utay forced Imuy, "Do the laundry!"'

(21) a. *pangmu-an-na ni utay ti-imuy, qibasi-ka tu
help-PV-3ERG ERG Utay NCM-Imuy wash-IMP OBL
qudus
clothes

b. *tud-an-na ni utay ti-imuy, satzai-ka
teach-PV-3ERG ERG Utay NCM-Imuy sing-IMP

Thirdly, the interpretation of a deictic in the complement of *pawRat* 'force' is ambiguous. The deictic center can be either the external LC or the internal LC. In (22), *tazian* 'here' can refer to a place near the speaker (the external LC) or near Utay (the internal LC). The shift of deictic center does not occur in other types of control sentences, e.g., (23).

(22) pawRat-an-na ni utay ti-imuy pa-qawtu tazian
force-PV-3ERG ERG Utay NCM-Imuy CAUS-come here
sasakay
play
'Utay forces Imuy to come here to play.' ('come here': near speaker or near Utay)

(23) t<m>abal ti-buya tu sunis mawtu tazian sasaqay
<AV>stop NCM-Buya OBL child AV.come here play
'Buya stops children from coming here to play.'

As the Fin head of the complement clause of a *pawRat* ('force')-type control predicate is linked to an internal LC, specifically the initiator whose point of view is reported, the [-R] person feature it licenses must be identified with the [+R] person feature of this initiator. This perspective shift from an external LC to an internal LC is the underlying reason why the complement o f *pawRat* 'force' and other similar control predicates must be morphologically causativized.

The analysis that attributes the causativization of the embedded verb to the internal LC on Fin can also explain why *pa-* affixation is absent when the embedded verb is a collective predicate, as illustrated in (24). (24) is an example of partial control where both the agent and the theme are the arguments of the embedded verb. The [-R] person feature in the complement can be partially identified with the [+R] person feature of the internal Logophoric Center and thus no causativization is necessary. Likewise, when the embedded verb is reciprocal, causativization is not necessary, as illustrated in (25a). Note that when the internal LC brings about a reciprocal event but is not a participant of the event, the embedded verb still needs to take the causative marker, as shown in (25b).

(24) pawRat-an-na ni utay ti-imuy masulun matiw
force-PV-3ERG ERG Utay NCM-Imuy AV.together AV.go
sa taypak
to Taipei
'Utay forces Imuy to go to Taipei together (with him, Utay).'

(25) a. pawRat-an-na ni utay ti-imuy sim-qa-qila
force-PV-3ERG ERG Utay NCM-Imuy RECP-QA-scold
'Utay forces Imuy to scold each other. (Utay and Imuy scold each other.)'
b. pawRat-an-na ni utay ti-imuy
force-PV-3ERG ERG Utay NCM-Imuy
pa-sim-qa-qila ti-buya-an
CAUS-RECP-QA-scold NCM-Buya-LOC
'Utay forces Imuy to have a dispute with Buya. (Imuy and Buya scold each other.)'

In a scenario where the agent and the theme of a control predicate are the same, i.e., a reflexive event, the analysis presented here predicts that the embedded verb should not take the causative prefix, as there is only one participant in the matrix clause, i.e., the internal LC or the internal Speaker. There is no need for the occurrence of *pa-* to signal which participant the [-R] person feature should be identified with. The prediction is borne out (26).

(26) a. pawRat-an-na ni utay ya izip-na q<m>an
force-PV-3ERG ERG Utay ABS body-3GEN <AV>eat
tu mutun
OBLmouse
'Utay forced himself to eat a mouse.'
b. *pawRat-an-na ni utay ya izip-na
force-PV-3ERG ERG Utay ABS body-3GEN
pa-qan tu mutun
CAUS-eat OBL mouse

5.2 Functional Projections of Control Complements

To corroborate the proposed analysis, this section will show that control predicates that do not take a causativized verb complement like *paska* 'try' and *tud* 'teach' are restructuring predicates and are thus devoid of a Fin head in their complement that can be linked to an internal LC. A *pawRat* ('force')-type control predicate, on the other hand, does not involve restructuring in that the functional heads of its non-finite complement clause are still projected and active.

5.2.1 Finiteness

The complement clause of a control verb in Kavalan is not introduced by any overt complementizer, coordinating conjunction, or linker. This is true of both *try*-type control verbs and *persuade*-type control verbs, as illustrated in (27). The fact that the coordinating conjunction cannot introduce the causativized verb phrase in (27b) suggests that this type of control structure cannot be analyzed as coordination between a control verb and a causativized verb. This fact also indicates that the lack of an overt DP argument in the causativized verb phrase cannot be attributed to discourse-induced pro-drop.

(27) a. *m-paska ya sunis [tu/sRi/a satzai]
AV-try ABS child COMP/CONJ/LNK sing
b. *pawRat-an-na ni buya aiku [tu/sRi/a
force-PV-3ERG ERG Buya 1SG.ABS COMP/CONJ/LNK
pa-qibasi tu qudus]
CAUS-wash OBL clothes

Moreover, tense and aspect markers, if any, must be attached to the matrix control predicate. Whether the embedded verb takes the causative marker or not, it is unable to host its own tense or aspect marker. The following sentences in (28) are for illustration.

(28) a. paska=pa=iku s<m>alaw tu babuy na
try=FUT=1SG.ABS <AV>hunt OBL pig GEN
na'ung
mountain
'I will try to hunt boars.'
b. *paska=iku salaw=pa tu babuy na na'ung
try=1SG.ABS hunt=FUT OBL pig GEN mountain
c. tezung-an-na=ti ni utay ti-abas pa-qibasi
instruct-PV-3ERG=PFV ERG Utay NCM-Abas CAUS-wash
tu qudus
OBL clothes
'Utai instructed Abas to do the laundry.'
d. *tezung-an-na ni utay ti-abas
instruct-PV-3ERG ERG Utay NCM-Abas
pa-qibasi=ti tu qudus
CAUS-wash=PFV OBL clothes

Another grammatical property that is indicative of the nonfinite feature of the embedded verb in a control sentence is that it does not show voice alternation. Whether the embedded verb in a control sentence is affixed with the causative marker or not, it is not allowed to take the patient voice suffix, but must occur in the agent voice form. This voice restriction is exemplified below in (29).

(29) a. *paska=pa=iku salaw-an tu/ya babuy na
try=FUT=1SG.ABS hunt-PV OBL/ABS pig GEN
na'ung
mountain
b. *tezung-an-na ni utay ti-abas pa-qibasi-an
instruct-PV-3ERG ERG Utay NCM-Abas CAUS-wash-PV
tu/ya qudus
OBL/ABS clothes

The restriction of a verbal form to the agent voice form is a unique morphosyntactic property of a verb in a non-finite subordinate clause in Kavalan. Therefore, the causativized verb phrase in a *persuade*-type control sentence in Kavalan cannot be identified as a conjunct of the matrix predicate and neither can it be analyzed as a separate independent clause with pro-drop. If it were an independent clause with pro-drop, it would be able to host its

own tense and aspect markers and occur in the patient voice form, contrary to fact.

5.2.2 Structural Differences Between *Paska* 'try'-type Control and *PawRat* 'require'-type Control

Although the complements of both *paska* 'try'-type and *pawRat* 'force'-type control predicates are non-finite, they still differ in the structural complexity of functional projections. While the complement of *paska* 'try' and *tud* 'teach' is fully reduced and shows properties of restructuring (Wurmbrand 2001). The TP and CP of the complement of *pawRat* 'force'-type predicates are still projected and active.

First of all, no temporal adverbs can occur in the complement clause of *paska* 'try' and *tud* 'teach', as illustrated by the ungrammaticality of (30). This suggests that their complement clause is untensed or the tense of their complement clause is anaphoric. By contrast, (31) shows that the complement of *qeRas* 'require' can take a temporal adverb distinct from the matrix temporal adverb. There is a tense operator in the complement clause of *qeRas* 'require' that is distinct from, but constrained by, the matrix tense operator.

(30) a. *siRab paska=ti=iku [temawaR q<m>apaR
yesterday try=PFV=1SG.ABS tomorrow <AV>catch
tu mutun]
OBL mouse

b. *siRab tud-an-na=iku na
yesterday teach-PV-3ERG=1SG.ABS ERG
tina-ku [temawaR m-tebu tu qudus]
mother-1SG.GEN tomorrow AV-patch OBL clothes

(31) siRab qeRas-an-na=iku ni utay [temawaR
yesterday require-PV-3ERG=1SG.ABS ERG Utay tomorrow
pa-qibasi tu qudus]
CAUS-wash OBL clothes
'Yesterday, Utay required me to do the laundry tomorrow/today.'

Secondly, the question particle *ni* can occur in the complement clause of *paska* 'try' and *tud* 'teach' with matrix clause interpretation, as illustrated by (32c) and (33c). This is suggestive of a mono-clausal structure without embedded CP. The addition of *ni* to the complement clause of *qeRas* 'require', however, induces ungrammaticality (34c). This is because CP is projected in the complement of *qeRas* 'require' and thus the question particle is unable to receive matrix clause interpretation.

(32) a. paska [q<m>apaR tu mutun] ti-buya ni?
try <AV>catch OBL mouse NCM-Buya Q
b. = paska ni [q<m>apaR tu mutun] ti-buya?
c. = paska [q<m>apaR ni tu mutun] ti-buya?
'Does Buya try to catch mice?'

(33) a. tud-an-na=isu na tina-su
teach-PV-3ERG=2SG.ABS ERG mother-2SG.GEN
[m-tebu tu qudus] ni?
AV-patch OBL clothes Q
b. = tud-an-na=isu na tina-su ni [m-tebu tu qudus]?
c. = tud-an-na=isu na tina-su [m-tebu ni tu qudus]?
'Does your mother teach you to patch clothes?'

(34) a. pawRat-an-na=isu na tina-su
force-PV-3ERG=2SG.ABS ERG mother-2SG.GEN
[pa-qibasi tu qudus] ni?
CAUS-wash OBL clothes Q
b. = pawRat-an-na=isu na tina-su ni [pa-qibasi tu qudus]?
c. *pawRat-an-na=isu na tina-su [pa-qibasi ni tu qudus]?
'Does your mother force you to do the laundry?'

Finally, as discussed in section 5.1, the complement clause of *paska* 'try' and *tud* 'teach' cannot be negated, whereas the complement clause *qeRas* 'require' can take the imperative negator *naRin*.

Table 1 summarizes the structural properties of different types of control predicates in Kavalan. Control predicates that do not take a causativized verb complement like *paska* 'try' and *tud* 'teach' are restructuring predicates and are thus devoid of a Fin head in their complement that can be linked to an internal LC. By contrast, the functional heads of the complement clause of *pawRat* 'force' and *qeRas* 'require' (extended projections of IP and CP) are still projected and active.

Table 1. Structural properties of different control predicates

	***paska* 'try'**	***tud* 'teach'**	***pawRat* 'force'**
subordinator	x	x	x
tense/aspect affix on the embedded verb	x	x	x
AV-restriction on the embedded verb	√	√	√
temporal adverb in the embedded clause	x	x	√
***ni* in embedded clause, with matrix interpretation**	√	√	x
negation in the embedded clause	x	x	√
imperative/direct quote as complement	x	x	√
shift of deictic center	x	x	√ (optional)

6 Conclusion

The research findings on Kavalan obligatory control make significant contributions to the study of control both empirically and theoretically. It is shown that some control predicates like *pawRat* 'force' in Kavalan utilize a distinct grammatical structure in which the embedded verb must take the causative marker *pa-*. The control verbs that require their embedded verbs to be affixed with the causative prefix all denote an event where the agent obligates the theme to execute some action. It is argued that the causativization of the embedded verb in a control sentence cannot be explained by a purely syntactic or semantic account of obligatory control. Instead, a comprehensive and satisfactory explanation for Kavalan obligatory control must take into account how event structure and Logophoric Center are encoded in Syntax.

References

Baker, Mark. 1988. *Incorporation: A theory of grammatical function changing*. Chicago: University of Chicago Press.

Bianchi, Valentina. 2003. On finiteness as logophoric anchoring. In Jacqueline Guéron & Liliane Tasmowski (eds.), *Temps et point de vue/Tense and Point of View*, 213–246. Nanterre: Université Paris X.

Boeckx, Cedric, Norber Hornstein & Jairo Nunes. 2010. *Control as movement*. Cambridge: Cambridge University Press.

Chang, Yung-li & Wei-tien Dylan Tsai. 2001. Actor-sensitivity and obligatory control in Kavalan and some other Formosan languages. *Language and Linguistics* 2. 1–20.

Chomsky, Noam & Howard Lasnik. 1993. The theory of principles and parameters. In Joachim Jacobs, Arnim von Stechow, Wolfgang & Theo Vennemann (eds.), *Syntax: An international handbook of contemporary research*, 506–569. Berlin: Walter de Gruyter.

Culy, Christopher. 1994. Aspects of logophoric marking. *Linguistics* 32. 1055–1094.

Harley, Heidi. 2008. On the causative construction. In Shigeru Miyagawa & Mamoru Saito (eds.), *Handbook of Japanese Linguistics*, 20–53. Oxford: Oxford University Press.

Hornstein, Norbert. 1999. Movement and control. *Linguistic Inquiry* 30. 69–96.

Jackendoff, Ray & Peter W. Culicover. 2003. The semantic basis of control in English. *Language* 79. 517–556.

Landau, Idan. 2004. The scale of finiteness and the calculus of control. *Natural Language and Linguistic Theory* 22. 811-877.

Polinsky, Maria & Eric Potsdam. 2006. Expanding the scope of control and raising. *Syntax* 9. 171–192.

Rooryck, Johan. 2008. Control via selection. In William D. Davies & Stanley Dubinsky (eds.), *New horizons in the analysis of control and raising*, 281–292. Dordrecht: Springer.

Rosen, Sara Thomas. 2003. The syntactic representation of linguistic events. In Lisa Cheng and Rint Sybesma (eds.), *The second Glot international*

state-of-the-article book: The latest in linguistics, 323–366. New York: Mouton de Gruyter.
Rosenbaum, Peter. 1967. *The grammar of English predicate complement constructions*. Cambridge, MA: MIT Press.
Sag, Ivan A. & Carl Pollard. 1991. An integrated theory of complement control. *Language* 67. 63–113.
Travis, Lisa. 2000. Event structure in syntax. In Carol Tenny & James Pustejovsky (eds.), *Event as grammatical objects: The converging perspectives of lexical semantics and syntax*, 145–185. Stanford: CSLI Publications.
Wurmbrand, Susanne. 2001. *Infinitives: Restructuring and clause structure*. Berlin: Mouton de Gruyter.

Focus feature percolation: Evidence from Tundra Nenets and Tundra Yukaghir

Dejan Matić
Max Planck Institute for Psycholinguistics, Nijmegen

Irina Nikolaeva
SOAS, University of London

Proceedings of the 21st International Conference on
Head-Driven Phrase Structure Grammar

University at Buffalo

Stefan Müller (Editor)

2014

Stanford, CA: CSLI Publications

pages 299–317

Matić, Dejan & Irina Nikolaeva. 2014. Focus feature percolation: Evidence from Tundra Nenets and Tundra Yukaghir. In Stefan Müller (ed.), *Proceedings of the 21st International Conference on Head-Driven Phrase Structure Grammar, University at Buffalo*, 299–317. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2014.16.
CC-BY 4.0

Abstract

Two Siberian languages, Tundra Nenets and Tundra Yukaghir, do not obey strong island constraints in questioning: any sub-constituent of a relative or adverbial clause can be questioned. We argue that this has to do with how focusing works in these languages. The focused sub-constituent remains in-situ, but there is abundant morphosyntactic evidence that the focus feature is passed up to the head of the clause. The result is the formation of a complex focus structure in which both the head and non-head daughter are overtly marked as focus, and they are interpreted as a pairwise list such that the focus background is applicable to this list, but not to other alternative lists.

1 Preliminaries

This paper intends to enhance the empirical basis for the typology of constituent questions and syntactic islands by presenting new data on systematic island constraints violations in two languages of the extreme north of Eurasia, Tundra Yukaghir (TY, north-eastern Siberia, isolate) and Tundra Nenets (TN, north-western Siberia, Uralic). The data were obtained through the authors' own fieldwork supported by the Max Planck Society (MPI EVA and MPI for Psycholinguistics) for TY, as well as an ELDP grant and a grant from the Academy of Finland (project number 125225) for TN. As explained below, we believe these data to be interesting because they contribute to our understanding of how the focusing operation works and can be representation in grammar.

Both languages display a total lack of strong island effects in questioning. Relative and adverbial clauses are headed by non-finite verbal forms such as participles, action nominals or converbs, and the *wh*-word remains in situ. Examples (1) and (2) illustrate the relative and adverbial clauses in TN and TY, respectively.

(1) a. [[xənʹana yilʹe-wi°] nʹenecʹ°] to-sa ?
where live-PF.PTCP man come-INTER.3SG
lit. 'The man who lived where arrived?'

b. Petʹa [Wera-h ŋəmke-m xada-qma-xəd°] to-sa ?
Petya Wera-GEN **what-ACC** kill-PF.AN-AB come-INTER.3SG
lit. 'Petya came after Wera killed what?'

(2) a. [[qaduŋudəŋ uː-nu-j] köde] ŋol-k?
whither go-IMPF-PTCP person be-NEUFOC.INTER.2SG
lit. 'You are a person who goes where?'

b. [neme lew-rəŋ] qudoːl-ŋu?
what eat-SS.CVB lie-NEUFOC.INTER.3PL
lit. ‘While eating what are they lying?’

While TN and TY behave identically with respect to questions, they diverge with respect to the other types of filler-gap dependencies. In TN relativization and topicalization obey island constraints, while in TY they do not, similar to questioning. This difference is illustrated below for topicalisation out of adverbial clauses. Example (3a) demonstrates that TY allows syntactic topicalization out of an adverbial clause but the parallel TN example (3b) is ungrammatical.

(3) a. čoɣojə-lə met mə=kewečəŋ [amaː-gi
knife-ACC 1SG EX=leave.NEUFOC.INTR.1SG father-3POSS
met-in __ kiː-də-ɣa]
1SG-DAT give-3-DS.CVB
‘Knife, I left after his father gave __ to me.’

b. *ti [nʹīsʹa-nta __ xada-qma-xəd°] Wera xəya-sʹ°
reindeer father-GEN.3SG kill-PF.AN-ABL Wera go-PST.3SG
‘The reindeer, Wera left after his father killed __.’

This suggests that islands violations in questions do not come from the same source as in other types of extractions. We will argue that they have to do with how focusing works in these languages, cf. Matić (2014).

Among numerous explanations for question islands violations even in ‘well behaved’ languages, it has been suggested that the issue may not be the nature of the filler-gap dependency itself, but the focusability of certain types of structures: only those structures that are focusable can be subject to inquiry (Erteschik-Shir 1973, 2007; Van Valin 1994, 2005). Syntactic islands such as relative and adverbial clauses are known to be inherently presupposed and therefore cannot normally function as the locus of focusing operations (Frege 1892; Lambrecht 1994; Erteschik-Shir 2007). However, if an island clause is embedded into the matrix clause which itself is presupposed or easily presupposable (e.g. an existential clause), the island is the only candidate for focusing (Erteschik-Shir 1973, 2007; Shimojo 2002). This reverses the focusability relationship and renders the question island the major focus domain in the sentence. As a result, any sub-constituent of the island clause can be focused, as in the following Danish example.

(4) Hvad for en slags is er der mange børn [der kan li‘ __]?
what kind of ice.cream are there many children who can like
‘What kind of ice cream are there many children [who can like __]?’

In languages like Danish, on which Erteschik-Shir's original proposal was based, focusability shift is observed when pragmatic/semantic factors conspire to render the island clause focusable. We propose below that TN and TY achieve the same effect through grammar, by formally treating phrases with focused sub-constituent as focused. The focused sub-constituent remains in-situ but the focus feature percolates up to the mother node to provide interpretation which, we show, is a pairwise-listing complex focus.

2 Focus meaning and focus marking

The definition of focus we employ in this paper relies on Alternative Semantics (Rooth 1992; Krifka & Musan 2012). Focus will be understood as an operator that triggers common ground update via invocation of alternatives. While the ordinary semantic value, ignoring the contribution of focus, is a standard proposition ($[\![\alpha]\!]^o$), the focus semantic value ($[\![\alpha]\!]^f$)is a set of propositions that differ from each other only in that the denotatum of the focused expression is replaced by another object of the same type. To constrain the focus semantic value to relevant alternative propositions, a context variable C is introduced. It refers to a contextually determined set of alternatives, along with a focus operator which induces the requirement that C be a subset of focus-induced alternatives (Q).

(5) *JOHN arrived.*
   a. $[\![\alpha]\!]^o$: arrive'(John)
   b. $[\![\alpha]\!]^f$: $Q = \lambda p \exists x[p = \text{arrive'}(x)] \sim C$, where $C \subseteq Q$

This is largely identical to the widely accepted semantics for questions, according to which the meaning of a question is a set of contextually relevant propositions corresponding to the answer (Hamblin 1973; Hagstrom 1998). For instance, the question *Who arrived?* and the answer with the focus on the subject, *JOHN arrived*, have an identical focus-semantic value, the set of propositions of the form *x arrived*, where x is a variable ranging over entities constrained by C. The difference between questions and the answers is the identification of one true alternative in the latter. Following Abusch (2010), we assume that wh-words are a subtype of focus with a semantic contribution of their own. Minimally, they are soft presupposition triggers; the presupposition induces existential quantification over the question word and thus creates the ordinary semantic value with specific indefinite interpretation. This results in the following semantics: the ordinary semantic value (6a) is that someone arrived; the focus semantic value comprises all contextually plausible identifications of the indefinite.

(6) *Who arrived?*
   a. $[\![\alpha]\!]^o$: $(\exists x)[arrive'(x)]$
   b. $[\![\alpha]\!]^f$: $Q = \lambda p \exists x[p = arrive'(x)] \sim C$, where $C \subseteq Q$

Focus can also be associated with certain items that are commonly interpreted as quantifying over alternatives and are therefore focus-sensitive. In this paper we only deal with *only* and assume the following meaning for it based on König (1991), Horn (1996), and Krifka (1998):

(7) *Only JOHN arrived.*
   a. $[\![\alpha]\!]^o$: arrive'(John)
   b. $[\![\alpha]\!]^f$: $Q = \lambda p \forall x[p = arrive'(x) \Rightarrow x = John\} \sim C$, where $C \subseteq Q$

The exhaustive effect of focus modified by *only* is explained via universal quantification: any element to which the focus background is applicable must correspond to the description provided by the focus phrase. Exclusive particles of the *only*-type thus correspond to a universal quantifier which scopes over alternatives generated by focus.

These three types of focus, i.e. wh-question focus, narrow 'argument focus' and focus generated by *only*, are all relevant here because they are encoded identically in both TN and TY. Consider TN first. This language exhibits obligatory subject agreement in person/number, while object agreement in number is 'optional' in the sense that transitive verbs agree only with a subset of objects. A non-focused object can trigger agreement depending on a number of semantic and information structure-related conditions (for detail see Nikolaeva 2014). In (8) the marker *-da* that indicates that the 3rd person singular subject is acting upon a singular object is optional on the verb. When this marker is absent, the verb only indexes the subject.

(8) Wera-h ti-m xadaə(-da)
   Wera-GEN reindeer-ACC kill.3SG(>**SG.OBJ)**
   'Wera killed a/the reindeer.'

However, a focused object never triggers agreement on the verb. This applies to all three relevant types of focus, i.e. questions and answers focus as in (9a) and the focus modified by *only* as in (9b). The latter example demonstrates that when the object word hosts the focus-sensitive particle *-r'i/-l'i* **'**only', the object is treated as focused for the purpose of agreement.

(9) a. ŋəmke-m xada-sa(*-da)?
   what-ACC kill-INTER.3SG(*>**SG.OBJ)**

ti-m xadaə(*-da).
reindeer-ACC kill.3SG(*>**SG.OBJ)**
‘– What did he kill? ─ He killed a REINDEER.’

b. te-r′i-m xadaə(*-da)
reindeer-**ONLY**-ACC kill.3SG(*>**SG.OBJ)**
‘He only killed a REINDEER.’

Thus, the impossibility for the transitive verb to exhibit object agreement is a reliable indicator of the focus status of its object.

In TY agreement in person/number is always with the subject, but its form depends on what non-verbal element is in focus (and partially on transitivity). The actual agreement exponence must be chosen from one of the three available paradigms, the so-called subject-focus paradigm (SFOC), the object-focus paradigm (OFOC) or the neutral-focus paradigm (NEUFOC), so the form of the verb is ultimately determined by the position of focus. For example, the 3rd person singular subject agreement suffix is *-l* for SFOC, *-mələ* for OFOC, *-č/-j* for intransitive verbs in NEUFOC, *-m* for transitive verbs in NEUFOC, and agreement is altogether absent when focus is on the transitive subject. In addition, intransitive subjects and transitive objects must bear a special focus marker *-(ə)k/-ləŋ* instead of the grammatical case marker when they are in focus. Focus marking and focus agreement facts are summarized in the table below, where S stand for an intransitive subject, O for a direct object, and A for transitive subject.

Table 1: Focus marking in Tundra Yukaghir

Focused element	**Focus marking**	**Focus agreement on the verb**
S	-(ə)k/-ləŋ	SFOC
O	-(ə)k/-ləŋ	OFOC
A	Ø	Ø
Oblique	Ø	NEUFOC

This distribution is shown in (10) for wh-questions and answers. (10a) exemplifies the focus on the intransitive subject: the subject is marked by the focus marker *-(ə)k* and the 3rd person singular verbal form is chosen from the SFOC paradigm. In (10b) the object is in focus; it bears the focus marker and the verb agrees with the 1st person singular subject but the agreement form is OFOC. In (10c) the focus is on the transitive subject, which results in the lack of focus marking and agreement. Finally, in (10d) the focus is on an oblique element. This element stands in the required grammatical case (the dative, in this instance) and bears no focus marker, whereas the form of the verb is NEUFOC.

(10)a. kin-ək ewrə-l? ma:rqə-n köde-k ewrə-l.
who-**FOC** walk-**SFOC.**3SG one-GEN man-**FOC** walk-**SFOC.**3SG
'– Who went (there)? ─ One MAN did.'

b. neme-ləŋ iŋe:-məŋ? labunmə-ləŋ iŋe:-məŋ.
what-**FOC** fear-**OFOC**.1/2SG ptarmigan-**FOC** fear-**OFOC**.1/2SG
'– What do you fear? ─ I fear PTARMIGANS.'

c. tet-qanə kin-Ø tite we:-Ø? əl=tet-Ø tite we:-Ø
you-ACC who-**Ø** so do-**Ø** NEG=you-**Ø** so do-**Ø**
met-ul?
1SG-ACC
'– Who treated you like that? ─ Didn't YOU treat me like that?'

d. qaduŋudəŋ kewej? Moskva-ŋiń keweč.
whither go.**NEUFOC.**INTER.3SG Moscow-DAT go.**NEUFOC**.3SG
'– Where did he go? ─ He went to MOSCOW.'

We only have limited data for *only*-type focus in TY, but example (11) demonstrates that when the free-standing focus-sensitive particle *mo:rqo:n* 'only' has object in its scope, this object must bear the focus marker and agreement on the verb must come from the OFOC paradigm. NEUFOC agreement on the verb in combination with the non-focus marked object is ungrammatical.

(11) mo:rqo:n lačiləŋ ičo:-mələ
only fire.**FOC** look-**OFOC**.3SG
/ *mo:rqo:n lačilə ičo:-m
/ only fire.**ACC** look-**NEUFOC**.3SG
'He only saw FIRE.'

We can see then that both TN and TY have complicated systems of focus marking on core arguments and focus-sensitive agreement on the main verb. This equally applies to *wh* and non-*wh* types of narrow focus.

3 Focusing sub-constituents

In both languages if a sub-constituent of a complex phrase is interpreted as focused, the whole phrase is treated as focus for the purpose of focus marking and agreement. In TN this can be most clearly seen on focused objects. If any non-head sub-constituent of the object NP is focused, object agreement is impossible on the verb, suggesting that the whole NP is marked as focus.

This is shown in (12) for the focused possessor, attributive modifier, and a complement of the head noun.

(12) a. [xībʹa-h ti-m] xada-sa(*-da) ?
who-GEN reindeer-ACC kill-INTER.3SG(*>**SG.OBJ)**
[Wera-h ti-m] xadaə(*-da).
Wera-GEN reindeer-ACC kill.3SG(*>**SG.OBJ)**
‘– Whose reindeer did he kill?’ – He killed WERA’s reindeer.’

b. [xurka ti-m] xada-sa(*-da) ?
what.kind reindeer-ACC kill-INTER.3SG(*>**SG.OBJ)**
[serako ti-m] xadaə(*-da).
white reindeer-ACC kill.3SG(*>**SG.OBJ)**
‘– What kind of reindeer did he kill?’ – He killed a WHITE reindeer.’

c. [ŋəmke-h nʹamna ləx°nako-m] wadʹe-ca-n° ?
what-GEN about tale-ACC tell-INTER-2SG
/ *wadʹe-ca-r° ? [[tʹonʹa-h nʹamna] ləx°nako-m]
/ tell-INTER-2SG>**SG.OBJ** fox- GEN about tale-ACC
wadʹeqŋa-d°m / *wadʹeqŋa-w°
tell-1SG / tell-1SG>**SG.OBJ**
‘– A tale about what did you tell? – I told a tale about a FOX.’

Sentences (13) exemplify the same distribution for the *only*-type focus.

(13) a. [Wera-rʹi-h ti-m] xadaə-d°m / *xadaə-w°
Wera-**ONLY**-GEN reindeer-ACC kill-1SG / kill-1SG>**SG.OBJ**
‘I only killed WERA’s reindeer.’

b. [parʹidʹenʹa-rʹi ti-m] xadaə(*-da)
black-**ONLY** reindeer- ACC kill.3SG(*>**SG.OBJ)**
‘He only killed a BLACK reindeer.’

c. [tʹonʹa-rʹi-h nʹamna ləx°nako-m] wadʹeqŋa / *wadʹeqŋa-da
fox-**ONLY**-GEN about tale-ACC tell-3SG / tell-3SG>**SG.OBJ**
‘He told a tale only about a FOX.’

Similarly, if a modifier or a possessor is in focus, the whole NP behaves like focus in TY. In (14a) the focus is associated with the possessor of the locative oblique, and the verbal agreement must come from the NEUFOC paradigm. In (14b) the modifier of the intransitive subject is in focus, which results in the focus marking of the head noun and SFOC agreement on the verb.

(14) a. [kin nime-də-ɣa] ewreː-nu-k? [wolʹbə
who house-3-LOC walk-IMPF-**NEUFOC.**INTER.2SG **friend**
nime-də-ɣa] ewreː-nu-jəŋ.
house-3-LOC walk-IMPF-**NEUFOC.**1SG
'– To whose house are you going? – I'm going to a FRIEND's house.'

b. pure-n [neme nime-k] oɣoːlə-l?
above-LOC **what** house-**FOC** stand-**SFOC.**3SG
[joqon nime-k] oɣoːlə-l.
Yakut house-**FOC** stand-**SFOC.**3SG
'– What kind of house stands on the top? – A YAKUT house does.'

Crucially, island clauses behave identically to simple NPs with respect to focus-sensitive agreement and focus-sensitive marking. In TN, if any sub-constituent of a relative clause is focused and the relative clause modifies the object of the main verb, this verb cannot be marked for object agreement. This is illustrated in (15): in (15a) the focus on the relative clause subject and in (15b) it is on the clausal adjunct. In both instances object agreement on the main verb is ungrammatical.

(15) a. [[xībʹa-h xada-wi°] ti-m] məne-ca-n° ?
who-GEN kill-PF.PTCP reindeer-ACC see-**INTER**-2SG
/ *məne-ca-r° ? [[Wera-h xada-wi°] ti-m]
/ see-INTER-2SG>**SG.OBJ** **Wera-GEN** kill-PF.PTCP reindeer-ACC
məneqŋa-dm° / *məneqŋa-w°.
see-1SG / see-1SG>**SG.OBJ**
'– You saw the reindeer killed by whom? – I saw the reindeer killed by WERA.'

b. [[Wera-h sʹax°h xo-wi°] noxa-m]
Wera-GEN **when** find-PF.PTCP polar.fox-ACC
xada-sa-n° / *xada-sa-r° ? [[Wera-h tʹenʹana
kill-**INTER**-2SG / kill-INTER-2SG>**OBJ.SG** Wera-GEN **yesterday**
xo-wi°] noxa-m] xadaə-d°m / *xadaə-w°
find-PF.PTCP polar.fox-ACC kill-1SG / kill-1SG>**OBJ.SG**
'– You killed the polar fox which Wera found when? – I killed the polar fox which Wera found YESTERDAY.''

Similar patterns obtain for the *only*-type focus. In (16) the sub-constituent in the scope of *only* hosts the bound particle *-rʹi/ -lʹi*, but the head noun must be marked as focused too, as is evidenced by the lack of object agreement.

(16) a. [[Wera-rʹi-h pedara-xəna xo-wi°] ti-m]
Wera-**ONLY**-GEN forest-LOC find-PF.PTCP reindeer-ACC

mənеqŋa-d°m / *mənеqŋa-w°
see-1SG / see-1SG>SG.OBJ
‘I saw the reindeer that only WERA found in the forest.’

b. [[Wera-h pedara-r´i-x°na xada-wi°] ti-m]
Wera-GEN forest-**ONLY**-LOC kill-PF.PTCP reindeer-ACC
mənеqŋa-d°m / *mənеqŋa-w°
see-1SG / see-1SG>SG.OBJ
‘I saw the reindeer which Wera killed only IN THE FOREST.’

In TY, if the relative clause with a focused sub-constituent modifies the intransitive subject, it is morphologically marked as focus and the verb bears SFOC agreement.

(17) [[kin pa:j-o:l] o:-k] o:riń a:-nu-l?
who beat-STAT.AN child-**FOC** weep-IMPF-**SFOC.**3SG
[[taŋ we:n kö:d’ədo: paj-o:l] rukun-ək]
that **other** **boy** beat-STAT.AN thing-**FOC**
o:riń a:-nu-l
weep-IMPF-**SFOC.**3SG
‘– The child beaten by whom is crying? – The one beaten by THAT OTHER BOY is crying.’

If the relative clause modifies the transitive subject, there is no agreement or focus marking.

(18) sespə-lə [qaduŋudəŋ kewej-l′əl-d′ə] köde-Ø oŋotej-Ø?
door-ACC **whither** leave-EV-IMPF.PTCP person-**Ø** open-**Ø**
lit. ‘The man who went where opened the door?’

Focusing a sub-constituent of the relative clause that modifies the object of the main verb requires focus marking on that object and object-focus agreement on the verb. As shown in (19b), alternative marking, e.g. a NEUFOC form of the verb and the regular accusative case marker on the object, would be ungrammatical in this instance.

(19) a. [[kin jaqta:-nu-l] jaqtə-k] möri:-məŋ?
who sing-IMPF-AN song-**FOC** hear-**OFOC.**1SG
[[ama: jaqta:-nu-l] jaqtə-k] möri:-məŋ.
father sing-IMPF-AN song-**FOC** hear-**OFOC.**2SG
lit. ‘– The song which who was singing did I hear? – You heard the song which FATHER was singing.’

b. *[kin jaqta:-nu-l] jaqtə-ɣanə möri:-mək?
who sing-IMPF-AN song-**ACC** hear-**NEUFOC.**2SG

Finally, questioning out of the relative clause that modifies the oblique element requires NEUFOC agreement on the verb with no special focus marking on the oblique. In (20a) the relativized nominal corresponds to the complement of the copula verb *be* and counts as oblique for the purpose of focus agreement. In (20b) we exemplify questioning out of the adverbial clause, which also requires NEUFOC agreement.

(20) a. [[kin pa:j-o:l] rukun] ŋol-k?
who hit-STAT.AN thing be-**NEUFOC.INTER**.2SG
'You are a person hit by whom?'

b. [kin kelu-də-ɣa] tet kewej-k?
who come-3-DS.CVB you leave-**NEUFOC.INTER**.2SG
'After whose arrival did you leave?' [lit. after who arrived did you leave]

So it is not the syntactic role of the focused element within the island clause that affects the patterns of agreement and focus marking in the main clause, but the syntactic role of its head. The noun modified by a relative clause or the dependent verb form in the adverbial clause are morphosyntactically treated as focused elements. For instance, in (20b) the main verb bears NEUFOC agreement even though the questioned/focused word corresponds to the intransitive subject, so technically SFOC agreement could have been expected.

In sum, the focus feature responsible for the marking of the phrase as focused and for the patterns of agreement it triggers on the verb must be associated with the head of that phrase in both languages in question, even though, at the first glance, the semantic operation of focusing appears to target one of its non-head daughters.

4 Complex focus structures

Based on the morphosyntactic evidence presented in the previous section, we propose that the grammar of TN and TY has to refer to two focus-related features: the feature [FOC], whose value is some semantic expression corresponding to the semantics of the focused word, and the [WH] feature. [WH] is a subtype of [FOC], so that wh-words are positively specified for both [FOC] and [WH], while the non-wh focus is only specified as [FOC]. We take this double specification to be a direct corollary of the meaning of questions briefly introduced in Section 2. Clearly, *wh*-words must carry [FOC] since they define the disjunctive set of alternatives which is the meaning of the

question proper. On the other hand, they must also be specified for [WH] in order to trigger the appropriate force. In other words, we are assuming that [FOC] is there to delimit the scope of alternatives, while [WH] signals illocution.

This is reflected in the morphosyntax of both TN and TY. As in many standard analyses of *wh*-questions, question formation involves the relationship between the [WH]-specified word and the clause (the main verb) which enforces its interrogative reading. This relationship has a morphosyntactic expression: in both languages some environments (past tense in TN and NEUFOC in TY) require the verb to take a special interrogative form, independently on whether the *wh*-word is located in the main or embedded clause. This can be seen from the comparison between TN and TY questions which contain the interrogative form of the verb in (15a) and (20b) and their ungrammatical counterparts in which the verb does not host the interrogative marker in (21).

(21) a. *[xībʹa-h xada-wi°] ti-m mənеqŋa-nə-sʹ°?
who-GEN kill-PF.PTCP reindeer-ACC see-2SG-**PST**
'You saw the reindeer killed by whom?'

b. *[kin kelu-də-ɣa] tet kewe**čək**?
who come-3-DS.CVB you leave.**NEUFOC**.2SG
'After whose arrival did you leave?' [=after who arrived did you leave].

The relationship between the wh-word and the main verb is direct in the sense that the [WH] feature is not morphosyntactically reflected on the head of the dependent clause and does not immediately contribute to its interpretation, as its main role is to signal the illocutionary force conveyed by the whole sentence.

In contrast to [WH], the [FOC] feature is crucially passed to the head of the dependent clause (the relativized nominal in relative clauses or the non-finite verbal form in adverbial clauses) first, and only then enters the relationship with the main verb. We will not provide the technical implementation of this idea here but believe that syntactically it may be expressed via some kind of percolation mechanism which targets [FOC] and resembles the theory of focus projection which accounts for the placement of focal accents in English (Selkirk 1995). On the standard focus projection view, only heads and arguments project focus. However, we are not aware of any structural restrictions in TN and TY that would permit the focus feature to be transmitted to the maximal projection from certain positions only, so the mechanism has to be freer for these languages. Any non-head sub-constituent of the phrase carrying [FOC] can pass it to the head, as was in fact reflected in Bürings' (2006) theory of 'Unrestricted Vertical Focus Projection'. We can

formulate this as two basic principles, which ensure that the head must be focus-marked no matter what non-head daughter is specified as focus:

(22) [FOC] on a non-head daughter licenses [FOC] on the head
[FOC] on the head licences [FOC] on the phrase.

Another important difference is semantic. Whereas the focus projection mechanism was originally intended to account for broad focus structures, the focus percolation to the head of the clause/phrase in TN and TY creates what Krifka (1991) refers to as 'complex focus' in which both the head of the phrase/clause and the original carrier of focus are foci, i.e. expressions whose denotations have alternatives in the context. These two foci are not interpreted independently, but rather as a pairwise list, such that the focus background is applicable to this list, but not to other alternative lists. The functioning of complex foci is especially clear if they are modified with focus-sensitive items with quantificational force. To shows this, we adopt Krifka's (1991) enrichment of the representational language with *lists* (marked with •) which function identically to simple arguments, so that they can be bound by a single operator. In the sentence *John only introduced PETER to STEVE*, with a nuclear accent on both *PETER* and *STEVE*, the exclusive particle does not only scope over one of these two arguments. The interpretations according to which John introduced Peter only to Steve (and to nobody else) or only Peter (and nobody else) to Steve, while he might have introduced other people to each other, do not capture what this sentence conveys, namely that the only introduction event in which John was involved was between Peter and Steve. This interpretation follows from the complex focus structure (indicated by two nuclear accents): the focus alternatives have the form 'John introduced (x•y)', and *only* introduces universal quantification over these alternatives, similar to (7). The way this works is represented in (23).

(23) $[\![\alpha]\!]^o$ = introduce' (j,p,s)
$[\![\alpha]\!]^f$ = introduce' (j, x•y) & $\forall$x•y [introduce'(j, x•y) $\Rightarrow$ x•y = p•s] ~C,
C $\subseteq$ Q

This principle of complex focus interpretation is also at work with foci that are buried in island clauses in TN and TY. We first need a general rule that connects the island clause to the matrix clause:

(24) For a pair x,y, such that P(x•y), it is true that R(x)
where x = head of a clause, y = focused word within the clause, • = list operator, P = $\lambda x \lambda y$.island clause(x•y), R = λx.matrix clause(x)

The way alternatives are computed in this context is represented in (24').

(24') $[\![\alpha]\!]^o$ = R(a) & P(a•b)
$[\![\alpha]\!]^f$ = R(x) & P(x•y)~C & C ⊆ Q
$[\![\alpha]\!]^f$ = {R(a) & P(a•b), R(a) & P(a•c), R(a) & P(a•d)...
R(i) & P(i•b), R(i) & P(i•c), R(i) & P(i•d)...
R(m) & P(m•b), R(m) & P(m•c), R(m) & P(m•d)...}

The identity of *x* in the matrix clause is dependent on the list it forms with *y* in the island clause, so that these two variables are co-dependent – which is the essence of the pairwise list reading. Now consider the answer in (15b) again.

(15b) [[Wera-h t′en′ana xo-wi°] noxa-m] xadaə-d°m
Wera-GEN yesterday find-PF.PTCP polar.fox-ACC kill-1SG
'I killed the polar fox which Wera found YESTERDAY.'

Its rough semantics (abstracting from quantification, time, deixis, reference, and other details) can be represented as follows:

(15b') $[\![\alpha]\!]^o$ = kill' (me, fox) & find' (Wera, fox•yesterday)
$[\![\alpha]\!]^f$ = kill' (me, x) & find' (Wera, x•y)~C & C ⊆ Q
$[\![\alpha]\!]^f$ = {kill' (me, fox) & find' (Wera, fox•today),
kill' (me,fox) & find' (Wera, fox•yesterday),
kill' (me,fox) & find' (Wera, fox•last year)...
kill' (me, bird) & find' (Wera, bird•today),
kill' (me,bird) & find' (Wera, bird•yesterday),
kill' (me,bird) & find' (Wera, bird•last year)...
kill' (me, elk) & find' (Wera, elk•today),
kill' (me,elk) & find' (Wera, elk•yesterday),
kill' (me,elk) & find' (Wera, elk•last year)...}

The ordinary semantic value of this sentence is trivial: I killed a polar fox and Wera had found that fox the day before. The focus value consists of propositions with the format *I killed x such that Wera killed x at time y* in which the focused variables are co-dependently replaced by contextually appropriate alternatives of the same type, e.g. (fox•today), (bird•yesterday), etc. Importantly, the identity of *x,* which corresponds to the head of the phrase, is defined via relationship with *y,* which is the primary carrier of [FOC]. The pairwise list reading induces co-dependent identification of variables. In this way it is ensured that the alternatives cover both such propositions in which I killed the fox found by Wera today (as opposed to the one which he found yesterday) and such in which I killed the bird or the elk which Wera found yesterday (as opposed to the polar fox he found at the same time).

Question islands are a special case of this more general semantic operation. The head noun denotes a set of entities defined in terms of the properties specified in the *wh*-word. As indicated in Section 2, we take it that questions with *wh*-words have an ordinary semantic value which comes about via default presuppositional interpretation; this semantic value can be enriched via further specifications of *wh*-words as to the type of the element they denote, such that *who* adds the feature [+human], *what* [-human], etc. The focus value of questions is the set of propositions that differ in the denotation of the *wh*-word slot, restricted by the *wh*-word's specifications and by the context variable. In case of pairwise list readings within island clauses, focus-induced alternatives differ in the denotations of both the head noun and the *wh*-word, which co-dependently vary and form a pairwise list. Informally, this can be represented as follows:

(25) For which pair x,y, such that P(x•y), does it hold true that R(x)
where x = head of the question island, y = question word,
P = λxλy.island clause(x•y), R = λx.matrix clause(x)

For instance, in (15a) the question word 'who' denotes a set of men who kill reindeer and the question ranges over the set of reindeer which have the property of having been killed by these men and are defined in terms of this property.

(15a)	[[xīb′a-h	xada-wi°]	ti-m]	məne-ca-n° ?
	who-GEN	kill-PF.PTCP	reindeer-ACC	see-INTER-2SG
	'You saw the reindeer killed by whom?'			

The resulting meaning can be formulated as follows: For which pair (reindeer, person), such that it is true that person killed the reindeer, is it true that you saw the reindeer?

(15a') $[\![\alpha]\!]^o$ = see' (you, reindeer) & kill' (person•reindeer))
$[\![\alpha]\!]^f$ = see' (you, x) & kill' (y•x)~C & C ⊆ Q
$[\![\alpha]\!]^f$ = {see' (you, reindeer1) & kill' (Petya•reindeer1),
see' (you,reindeer1) & kill' (Misha•reindeer1),
see' (you,reindeer1) & kill' (Vasya•reindeer1)...
see' (you, reindeer2) & kill' (Petya•reindeer2),
see' (you,reindeer2) & kill' (Misha•reindeer2),
see' (you,reindeer2) & kill' (Vasya•reindeer1)...
see' (you, reindeer3) & kill' (Petya•reindeer3),
see' (you,reindeer3) & kill' (Misha•reindeer3),
see' (you,reindeer3) & kill' (Vasya•reindeer3)...}

This creates the broadening of the object of inquiry formally expressed as the broadening of question focus (cf. Nishigauchi 1990; Jin 2013). Like in Japanese, Chinese and a number of other languages with *wh* in-situ, question islands inquire about the identity of the whole island, making a crucial use of the identity of the element represented by the question word: this is due to the list-reading induced by complex focus. This can be seen in answers to *wh*-questions in island clauses. A felicitous answer normally recapitulates the entire island with the specified question word variable. (26a) is the regular answer to the question in (15a), although speakers can occasionally produce (26b) too, so that certain degree of variation is observed here, possibly due to the interfering influence of Russian in which all speakers are bilingual.

(26) a. Wera-h xada-wi° ti-m
Wera-GEN kill-PF.PTCP reindeer-ACC
'the reindeer killed by Wera.'

b. ?/* Wera-h (xada-wi°-m)
Wera-GEN kill-PF.PTCP-ACC
'(killed) by Wera.'

TN provides an additional morphosyntactic indication that focus within island clauses triggers complex focus interpretation forming a pairwise list with the head. It comes from the semantics of *only* in relative clauses. The focus-sensitive item *-r'i/-l'i* 'only' can take different scope within a relative clause, but the head noun always has to be specified as [FOC] irrespective of its scope, as follows from agreement on the main verb. The important point is that different scopes of *only* result in different focus readings, as indicated in the translations of examples (16) above. (16a) roughly means 'For the pair (reindeer, Wera), such that it is true that Wera (and no-one else) killed the reindeer in the forest, it is true that I saw the reindeer':

(16a')
$[\![\alpha]\!]^o$ = see' (me, reindeer) & kill' (Wera•reindeer, in.forest)
$[\![\alpha]\!]^f$ = see' (me, x) & $\forall$x•y [kill'(x•y, in.forest) $\Rightarrow$ x=Wera]~C & C $\subseteq$ Q

In contrast, (16b) can be represented as follows: 'For the pair (reindeer, forest), such that it is true that Wera killed the reindeer in the forest (and not anywhere else), it is true that I saw the reindeer'.

(16b')
$[\![\alpha]\!]^o$ = see' (me, reindeer) & kill' (Wera, reindeer•in.forest))
$[\![\alpha]\!]^f$ = see' (me, x) & $\forall$x•y [kill'(Wera, x•y) $\Rightarrow$ y=in.forest]~C & C $\subseteq$ Q

What is important here is that no matter what the scope of *only* is, the head of the island clause must be included in the focused pairwise list as indicated by the ungrammaticality of object agreement on the main verb, which signals that its object must be in focus. Different readings are derivable from the interaction of the focus-sensitive particle and the focus expression in its scope, on the one hand, and the denotation of the head noun, on the other.

Examples like (16) are particularly interesting because they appear to challenge the view, which has become popular in the recent years, that focus is associated with an overt or covert operator that either adjoins to the whole phrase or takes it as its complement. For example, Cable (2007, 2010ab) and later Coon (2009) propose that *wh*-questioning is not directly triggered by any properties of the *wh*-word itself. Rather, the question operation targets the features of a distinct formal element termed Q(uestion)-particle, which c-commands the *wh*-phrase and is accessible to the larger clause. This analysis creates the effect of the whole phrase/clause being available for questioning but eliminates the mechanism of feature percolation from the grammar altogether. In fact, the whole concept of 'pied-piping' becomes unnecessary because the operation of questioning applies to the maximum projection of the *wh*-word without looking 'inside' it. Despite its name, the Q-particle is not actually understood as being restricted to questions: Cable (2010b: 200ff.) suggests that a similar analysis may be applicable to other types of 'A-bar movement', in particular, the operation of focusing. This is also the basic claim in Horvath (2007), who argues that at least some types of focusing in Hungarian do not actually target the features of the focused phrase itself but are triggered by the focus-sensitive Exhaustivity Identification Operator *only* that c-commands the focus phrase and can be phonologically null.

It is not immediately clear to us how this type of analysis can account for the difference between (16a) and (16b) if the word within the scope of *only* does not have any bearing on grammaticality and the overall semantics because none of its features are targeted. In (16) we do have clear morphosyntactic evidence that both the sub-constituent of the relative clause and its head nominal are in focus: the former hosts the focus particle *-r′i*/*-l′i* 'only', while the latter appears to be specified as [FOC] because it cannot trigger object agreement on the main verb. The different position and scope of *-r′i*/*-l′i* create difference in interpretation, which effectively means that the word inside the island remains visible for the purpose of focusing. Since the maximal projection is also focused, we proposed that some kind of focus feature percolation may be responsible for the resulting structure. This operation has an important semantic effect: the focusing of a non-head sub-constituent and the percolation of the focus feature to the head results in the formation of a pairwise list in which the head denotes a set of entities defined in terms of the properties specified in the focus phrase, so both the head of the phrase and its sub-constituent are focused. The syntactic

implementation of this semantic analysis in its application to TN and TY is a subject of future work.

Abbreviations

ABL – ablative; ACC – accusative; AFOC – agent focus; AN – action nominaliser; COM – comitative; CVB – converb; DAT – dative; DS – different subject; GEN – genitive; FOC – focus; IMPF – imperfective; INTER – interogative; LOC – locative; NEUFOC – neutral focus; OBJ – object; OFOC – object focus; PST – past tense; PF – perfective; PL – plural; PTCP – participle; SFOC – subject focus; SG – singular; STAT – stative; SS – same subject

References

Abusch, D. 2010. Presupposition triggering from alternatives. *Journal of Semantics* 27. 37-80.

Beck, S. 2006. Intervention effects follow from focus interpretation. *Natural Language Semantics* 14.1–56.

Büring, D. 2006. Focus projection and default prominence. In: V. Molnár & S. Winkler (eds.), *The Architecture of Focus*, 321–346. Berlin: De Gruyter.

Cable, S. 2007. The grammar of Q: Q-particles and the nature of *wh*-fronting. Doctoral dissertation, MIT, Cambridge, MA.

Cable, S. 2010a. Against the existence of pied-piping: Evidence from Tlingit. *Linguistic Inquiry* 41. 563–594.

Cable, S. 2010b. *The grammar of Q: Q-particles, wh-movement, and pied-piping.* Oxford: OUP.

Coon, J. 2009. Interogative possessors and the problem with pied-piping in Chol. *Linguistic Inquiry* 40. 165-175.

Erteschik-Shir, N. 1973. *On the nature of island constraints*. PhD dissertation, MIT.

Erteschik-Shir, N. 2007. *Information structure. The syntax-discourse interface*. Oxford: OUP.

Ginzburg, J. & I. Sag. 2000. *Interrogative investigations: The form, meaning, and use of English interrogatives*. Stanford: CSLI.

Jin, D. 2013. Information Structure Constraints and Complex NP Islands in Chinese. Online Proceedings of the 20th International Conference on Head-Driven Phrase Structure Grammar, 110-120. http://web.stanford.edu/group/cslipublications/cslipublications/HPSG/2013/jin.pdf

Hamblin, C. L. 1973. Questions in Montague English. *Foundations of Language* 10. 41-53.

Hagstrom, P. 1998. *Decomposing questions*. Cambridge, MA: MIT dissertation.
Horvath, J. 2007. Separating ‘ Focus movement’ from focus. In: *Phrasal and clausal architecture*. In: S. Karimi, V. Samiian & W. Wilkins (eds.), 108-145. Amsterdam: John Benjamins.
Horn, L. 1996. Exclusive company: *only* and the dynamics of vertical inference. *Journal of Semantics* 13. 11–40.
König, E. 1991. *The meaning of focus particles.* London: Routledge.
Krifka, M. 1991. A compositional semantics for multiple focus constructions. In J. Jacobs (ed.), *Informationsstruktur und Grammatik*. Opladen: Westdeutscher Verlag.
Krifka, M. 1998. Additive particles under stress. *Proceedings from SALT VIII*, 111–128. Ithaca: CLC.
Krifka, M. & R. Musan. 2012. Information structure: Overview and linguistic issues. In: M. Krifka & R. Musan (eds.), *The expression of information structure*, 1-47. Berlin: De Gruyter.
Matić, D. 2014. Questions and syntactic islands in Tundra Yukaghir. In: R. van Gijn, J. Hammond, D. Matić, S. van Putten & A. V. Galucio (eds.). *Information structure and reference tracking and complex sentences*, 127-162. Amsterdam: Benjamins.
Nikolaeva, I. 2014. *A grammar of Tundra Nenets.* Berlin: Mouton de Gruyter.
Nishigauchi, T. 1990. *Quantification in the theory of grammar.* Dordrecht: Kluwer.
Rooth, M. 1992. A theory of focus interpretation. *Natural Language Semantics* 1. 75-116.
Selkirk, E. 1995. Sentence prosody: Intonation, stress, and phrasing. In J.A. Goldsmith (ed.), *The Handbook of Phonology*, 550-569. Oxford: Blackwell.
Van Valin Jr., R. D. 1994. Extraction restrictions, competing theories and the argument from the poverty of the stimulus. In: S. D. Lima, R. L. Corrigan & G. K. Iverson (eds.), *The reality of linguistic rules*, 243-259. Amsterdam: Benjamins.
Van Valin, Jr., R. D. 2005. *Exploring the syntax-semantics interface*. Cambridge: CUP.