Abstracts for the HPSG09 Conference

University of Goettingen

July 16-18 2009

Contents

Abstracts

Inflectional periphrasis in Persian

Olivier Bonami U. Paris-Sorbonne
Pollet Samvelian U. Paris 3

Grammars of Persian (e.g. Lazard et al., 2006) distinguish four conjugational periphrastic construction types. The passive construction is based on an inflected form of *šodan* 'become' preceded by a perfect participle (1). So-called 'perfect' forms are based on an inflected form of *budan* 'be' preceded by a perfect participle (2)—the auxiliary is a full word (2a), or a clitic (2b) depending on tense and mood. The future is formed with a special present tense form of *xâstan* 'want' followed by a bare stem (3). Finally, the progressive is based on an inflected form of *dâštan* 'have' followed by a finite form (4).[1]

(1) In tâblo foruxte mi-šav-ad.
this painting sold UNBD-become.S1-3SG
'This painting is sold.'

(2) a. Maryam in tâblo=râ foruxte bud.
Maryam this painting=DDO sold be.S2.3SG
'Maryam had sold this painting.'

b. Maryam in tâblo=râ foruxe=ast.
Maryam this painting=DDO sold=be.PRS.3SG
'Maryam has sold this painting.'

(3) Maryam in tâblo=râ xâh-ad foruxt.
Maryam this painting=DDO become.S1-3SG sell.S2
'Maryam will sell the painting'

(4) Maryam dâr-ad in tâblo=râ
Maryma have.PRS-3SG this painting=DDO
mi-foruš-ad.
UNBD-sell.S1-3SG
'Maryam is selling the painting.'

The differing properties of these four types of periphrasis stem from different origins as finite, infinitival or participial complements, and different degrees of grammaticalization, going from the quasi-analytic passive to the recently morphologized present perfect, through truly periphrastic forms that need to be integrated into inflectional paradigms despite being multi-word expressions. We assume that the different properties call for different analyses. We propose four contrasting analyses relying on the combination of an HPSG approach to feature geometry and syntactic combination, and an approach to paradigm organization and morphological exponence based on Paradigm Function Morphology (PFM; Stump, 2001). Interestingly, this combination of analytic tools allows us to treat the whole array of periphrastic constructions as lexical in origin—no phrasal construction or multi-word lexical entry of any kind is required.

[1] The glosses use the following abbreviations. BD: bounded aspect; DDO: definite direct object; EZ: Ezafe; NEG: negation; PRF: perfect; PRS: present; S1: first stem (a.k.a. the present stem); S2: second stem (a.k.a. the past stem); SBJV: subjunctive; UNBD: unbounded aspect.

1 Synthetic conjugation in HPSG/PFM

Before we address the analysis of periphrastic forms, we start with an account of synthetic conjugation. (5) lists the synthetic subparadigms of the lexeme *xaridan* 'buy', using the positive 2SG form as an illustration.

(5) a. Finite forms:
i. Simple present: *mi-**xar**-i*
ii. Simple bounded past: ***xarid**-i*
iii. Simple unbounded past: *mi-**xarid**-i*
iv. Simple subjunctive: *be-**xar**-i*
v. Imperative: *be-**xar***

b. Nonfinite forms:
i. Infinitive: ***xarid**-an*
ii. Present participle: ***xar**-ande*
iii. Perfect participle: ***xarid**-e*
iv. Gerund: ***xar**-ân*

Persian verbs exhibit a morphomic stem alternation (here *xar* vs. *xarid*). Neither stem is predictable from the other in general, and both stems are used in a combination of contexts which do not form a natural class. Affixal exponents realize unbounded aspect in the indicative (*mi-*), irrealis mood (*be-*), negation (*na-* or *ne-*, not illustrated here), type of nonfinite form (*-e* vs. *-ande* vs. *-an* vs. *-ân*), and subject agreement for finite forms. Within Paradigm Function Morphology, this rather simple position class system can be accounted for using the system of rule blocks outlined in (6). Remember that in PFM, realization rules are organized in successive blocks. When attempting to realize a given set of morphosyntactic feature, the most specific applicable rule within the block is chosen. (7) are sample rules from block V, written using the conventions of citepAckerman04: while (7a) asks that finite verbs with a 2.SG subject take the suffix *-i*, the more specific (7b) indicates that the suffix is dropped in the imperative.

(6)

III	II	I	IV	V
na- *ne-* *be-*	*mi-*	stem-selection	*-e* *-ande* *an*	*-am* *-i*/∅ *-ad*/∅ *-im* *-id* *-and*

(7) a. $X_V, \sigma : \{\text{PER } 2, \text{NB } sg\} \longrightarrow X$i

b. $X_V, \sigma : \{\text{PER } 2, \text{NB } sg, \text{MOOD } imper\} \longrightarrow X$

Since the integration of HPSG and PFM will be essential to our account of periphrastic conjugation, it is important that we specify how we intend to do it. The task is not trivial, because of PFM's reliance on comparisons of feature structure descriptions, which can not easily be formulated in existing

description languages for HPSG grammars. Rather than attempting a direct integration, we propose to use a PFM grammar to further constrain the class of signs satisfying an HPSG theory. Specifically, we rely on a slight reorganization of the feature geometry for head values as in (8), where MORSYN groups features that get realized in inflection and LID assigns a specific index to each lexeme (Spencer, 2005; Sag, 2007). We then define a version of PFM that is exactly like that of (Stump, 2001) except for the fact that (i) typed feature structures are used to model morphosyntactic feature bundles instead of category structures *à la* (Gazdar et al., 1985); and (ii) the rules output valence lists in addition to phonological representations (see below section 4 for motivation). The meta-constraint in (9) then links the two grammars.

(8) $$head \rightarrow \begin{bmatrix} head & \\ \text{LID} & lexemic\text{-}index \\ \text{MORSYN} & morsyn \end{bmatrix}$$

(9) A sign of type *word* meeting the description

$$\begin{bmatrix} \text{PHON} & \boxed{1} \\ \text{VAL} & \boxed{2} \\ \text{HEAD} & \begin{bmatrix} \text{LID} & \boxed{3} \\ \text{MORSYN} & \boxed{4} \end{bmatrix} \end{bmatrix}$$

is well-formed only if the PFM grammar licenses phonology 1 and valence 2 as a realization of the features 4 for the lexeme 3.

2 The passive

The passive in Persian is a typical complex predicate construction, whose properties are parallel to those of copula-predicative complement constructions. The auxiliary *šodan* is clearly the head: all inflectional information, e.g. negation (10), is realized on the auxiliary. The participle-auxiliary sequence is syntactically flexible: adverbs may intervene (11), the auxiliary may be scrambled over the participle (12). and long-distance fronting of the participle is possible (13).

(10) In tâblo foruxte ne-mi-šav-ad.
this painting sold NEG-UNBD-become.S1-3.SG
'This painting is not sold.'

(11) In tâblo foruxte hatman šod.
this painting sold certainly become.S2
'This painting was certainly sold.'

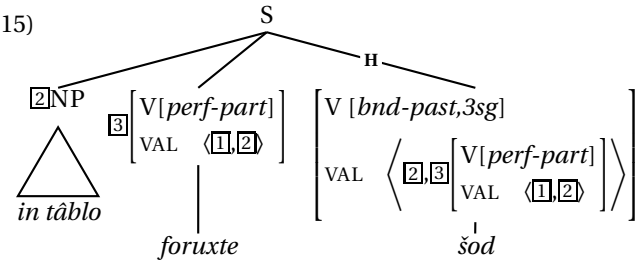
(12) In tâblo šod robude va foruxte.
this painting become.S2 stolen and sold
'It is this painting which was stolen and sold'

(13) Foruxte fekr mi-kon-am [agar in tâblo
sold thought UNBD-do.S1-1SG if this painting
__ be-šav-ad (...)]
SBJV-become.S1-3SG
'I think that if this painting is sold (...).'

To account for this we rely on an argument composition analysis in the spirit of (Hinrichs and Nakazawa, 1994) and subsequent work. Specifically we propose the lexical entry in (14) for the auxiliary lexeme *šodan*, giving rise to analyses such as that in (15). Note that under our analysis there is no passive participle, and subject demotion is effected directly in the auxiliary's entry. This is appropriate because (i) perfect participles are always active except in the periphrastic passive constructions (e.g. participial clauses with transitive head verbs take direct objects), and (ii) for semantic reasons there is no hope of using the same lexical entry for the auxiliary *šodan* and the full verb *šodan* (contrary to what happens in languages where the passive auxiliary coincides with the copula). Finally, note that under our analysis voice is not an inflectional category in Persian: the active-passive opposition is dealt with entirely within syntax.

(14)
$$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{LID} & \textit{šodan-aux} \end{bmatrix} \\ \text{CONT} & \boxed{1} \\ \text{VAL} & \boxed{L} \oplus \left\langle \begin{bmatrix} \text{FORM} & \textit{part} \\ \text{PERFECT} & + \\ \text{POL} & + \\ \text{CONT} & \boxed{1} \\ \text{VAL} & \langle \text{NP} \rangle \oplus \boxed{L} \\ \text{LEX} & + \end{bmatrix} \right\rangle \end{bmatrix}$$

(15)



3 Two sets of forms based on *budan*

There are five different subparadigms based on *budan*, illustrated here in (16). These contrast in two independent ways.

(16) a. Complex present: *xaride=i*
b. Complex bounded past: *xaride bud-i*
c. Complex unbounded past: *mixaride=i*
d. Complex subjunctive: *xaride bâš-i*
e. Compex perfect: *xaride bude=i*

3.1 Morphologized vs. truly periphrastic forms

In the complex present and the complex unbounded past, the perfect participle combines with the present clitic form of the auxiliary, which is homophonous with the exponent of subject agreement except for 3SG (there is also a nonclitic form of present *budan*, but it may not be used in this construction). In the complex bounded past and complex subjunctive, the perfect participle combines respectively with the bounded past and subjunctive forms of the auxiliary. Finally the complex perfect cumulates two forms of the auxiliary: the participle *bude* and the present form clitic (here *=i*). There is strong evidence that the forms historically based on the clitic auxiliary have undergone morphologization in contemporary Persian. First, the sequence cannot be interrupted

in any way; in particular, adverbs are excluded (17), as is participle fronting (18). Second, the distribution of the unbounded aspect marker *mi-* is otherwise unexplainable: it is the full construction, not the participle, that is unbounded. Finally, colloquial Persian allows a form of vowel reduction in the 3SG that is peculiar to these forms (19a): comparable contructions where the clitic auxiliary combines with an adjective do not give rise to the same pattern (19b).[2]

(17) *Rafte hatman=ast.
left certainly=be.S1.3SG
'(S)he has certainly left.'

(18) *Ne-mi-rafte sâlhâ Maryam be madrase=ast.
NEG-UNBD-gone years Maryam to school=be.S1.3SG
'For years, Maryam didn't go to school'

(19) a. raft'e=ast → raft'e:
left=be.S1.3SG
'(S)he left.'

b. bard'e=ast → bard'ast
slave=be.S1.3SG
'(S)he is a slave.'

Compare now the situation of forms that are based on a non-clitic auxiliary. The participle-auxiliary combination is more constrained than it is in the passive; in particular, neither adverbs (20) nor object clitics (21) can occur between the two verb forms, and negation must be realized on the participle (22). In addition, scrambling is excluded (23). However, the combination is not lexical, since the participle can be extracted (24).

(20) * Maryam dide hatman bud=aš
Maryam seen certainly be.S2=3SG

(21) a. Maryam dide bud=aš.
Maryam seen be.S2=3SG
'Maryam had seen him.'

b. * Maryam dide-aš bud.
Maryam seen=3SG be.S2

(22) Maryam Omid-râ na-dide bud.
Maryam Omid-DDO NEG-seen be.S2
'Maryam hadn't seen Omid.'

(23) * Maryam Omid-râ bud dide.
Maryam Omid-DDO be.S2 seen

(24) Foruxte fekr ne-mi-kon-am [_
sold thought NEG-UNBD-do.S1-1SG
bâš-ad in tâblo=râ].
be.SBJV-3SG this painting=DDO
'I don't think that s/he has sold this painting.'

[2] The only piece of evidence pointing in the other direction is the possibility for the auxiliary to have wide scope over a coordination of participles. However the existence of sublexical coordination in numerous languages calls into question whether this is a strong argument against a morphological analysis. We leave this issue for future research.

3.2 Semantic import

The use of a form based on *budan* may have two types of semantic import. The complex bounded past (25) and complex subjunctive (26) express respectively the past perfect and the subjunctive perfect. The complex unbounded past however does not express perfectivity at all. Rather, it has an evidential value (Windfuhr, 1982; Lazard, 1985; Jahani, 2000). Whereas the simple bounded past is used when the speaker has direct evidence for what she is asserting, the complex bounded past is used in contexts where the evidence is only indirect, as in (27).

(25) Qabl az inke Omid be-res-ad, Maryam
before from that Omid SBJV-arrive.S1-3SG Maryam
birun rafte bud.
out gone be.S2
'Maryam had left (before Omid arrived).'

(26) Fekr mi-kon-am Maryam mariz bude
thought UNBD-do.S1-1SG Maryam sick been
bašad.
be.SBJV-3SG
'I think Maryam has been sick.'

(27) (Banâ bar gofte-ye Omid) Maryam dar sâl-e 1950
According to-EZ Omid Maryam in year-EZ 1950
in xâne-râ mi-sâxte=ast.
this house-DDO UNBD-built=be.S1.3SG
'According to Omid, Maryam would have been building this house in 1950.'

The complex present is ambiguous between a perfect and an evidential value: it can be interpreted either as a present perfect (28a) or as a bounded past with indirect evidentiality (28b). Finally, the complex perfect expresses both perfectivity and indirect evidentiality: it is the indirect evidential equivalent of the complex bounded past (29). Note that this corresponds transparently to the fact the the complex perfect includes two realizations of the copula.

(28) a. Maryam tâze reside=ast.
Maryam new arrived=be.S1.3SG
'Maryam has just arrived.'

b. (Banâ bar gofte-ye Omid) Maryam in
According to-EZ Omid) Maryam this
xâne-râ dar sâl-e 1950 xaride=ast.
house-DDO in year-EZ 1950 bought=be.S1.3SG
'According to Omid, Maryam bought this house in 1950.'

(29) (Az qarâr), qabl az inke Omid
apparently before from that Omid
be-res-ad, Maryam birun rafte
SBJV-arrive.S1-3SG, Maryam out gone
bude=ast.
been=be.S1.3SG
'Apparently, Maryam had left before Omid arrived.'

As can be seen in (30), if the present perfect is ignored, morphosyntactic properties align nicely with morphologized vs. syntactic combinations: the morphologized forms are used

for indirect evidentiality, as stated by rules (31); while the truly periphrastic forms are used to express the perfect. The fact that the present perfect is unexpectedly synthetic calls for an paradigmatic analysis: this seems to be a standard case of syncretism, where the exponents used to realize a certain feature set (here indirect bounded past) are reused in some unrelated part of the paradigm. However, for such an analysis to be stated, one needs to treat perfect as an inflectional category, and thus to integrate the truly periphrastic forms in the inflectional paradigm (Ackerman and Stump, 2004).

(30)

		PAST		
	PRESENT	DIR. EV.	IND. EV.	SBJV
BD	***	bounded past	complex present	simple sbjv
UNBD	simple present	unbd past	cpl. unbd. past	
PRF	complex present	complex bnd. past	complex perfect	complex sbjv

(31) $X_V, \sigma : \{\text{EVID } \textit{indir}\} \longrightarrow X\mathsf{e}$ (block IV)

$X_V, \sigma : \{\text{EVID } \textit{indir}, \textit{3sg}\} \longrightarrow X\mathsf{ast}$ (block V)

4 Valence as exponence

As the last section stressed, what we need is a way to treat perfect forms as part of the inflectional paradigm, while allowing for the fact that they correspond to a combination of two words, one of which may be extracted. The solution we explore here can be stated informally as follows: a perfect form of a lexeme Y is a word whose phonology is borrowed from that of a form of the lexeme *budan*, but which subcategorizes for a perfect participle of this same lexeme Y. For instance, the 3SG positive complex bounded past of *xaridan* meets the description in (32). Notice the discrepancy between the lexemic index and the phonology. Because of the VC + specification, (32) gives rise to a verbal complex, which contrasts with our analysis if the passive (15) and accounts for the tighter solidarity of the verbal sequence. In addition, it will allow for the extraction of the participle within any HPSG approach to extraction.

(32)
$$\begin{bmatrix} \text{PHON} & \mathsf{bud} \\ \text{HEAD} & \begin{bmatrix} \text{LID} & \textit{xaridan} \\ \text{MORSYN} & \begin{bmatrix} \text{TNS} & \textit{past} \\ \text{PRF} & + \\ \text{AGR} & \textit{3sg} \\ \text{POL} & + \end{bmatrix} \end{bmatrix} \\ \text{VAL} & \left\langle \text{NP, NP, } \begin{bmatrix} \text{LEX} & + \\ \text{VC} & + \\ \text{LID} & \textit{xaridan} \\ \text{MORSYN} & \begin{bmatrix} \text{FORM} & \textit{part} \\ \text{PRF} & + \\ \text{POL} & + \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

The challenge now is to derive (32) in a principled way, while integrating it within an inflectional system where perfect forms may be realized either synthetically or periphrastically. The approach we propose is based on an extension of the power of realization rules in the spirit of (Spencer, 2005). In classical PFM, realization rules relate phonology-lexemic index pairs to phonology-lexemic index pairs. We propose that valence lists be added to the picture. The rule licensing (32) is given in (33). To realize a feature structure σ verifying [PRF +], one should refer the phonology to that of the corresponding bounded positive nonperfect form of *budan*, and add to the valence list a requirement for a form of Y realizing the same feature set except for the fact that it is a participle.[3]

(33)
$$\begin{bmatrix} \text{PHON} & X \\ \text{LID} & Y \\ \text{VAL} & Z \end{bmatrix}, \sigma : \begin{bmatrix} \text{PRF} & + \end{bmatrix} \longrightarrow$$
$$\begin{bmatrix} \text{PHON} & \mathsf{refer}\left(\begin{bmatrix} \text{PHON} & X \\ \text{LID} & \textit{budan} \\ \text{VAL} & Z \end{bmatrix}, \sigma \setminus \begin{bmatrix} \text{PRF} & - \\ \text{ASP} & \textit{bnd} \\ \text{POL} & + \end{bmatrix}, \text{I-V} \right) \\ \text{LID} & Y \\ \text{VAL} & Z \oplus \left\langle \begin{bmatrix} \text{LEX} & + \\ \text{VC} & + \\ \text{HEAD} & \begin{bmatrix} \text{LID} & Y \\ \text{MORSYN} & \sigma \setminus \begin{bmatrix} \text{FORM} & \textit{part} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

This approach to periphrastic perfect forms has two definite advantages. First, it is syntactically reductionist: periphrasis is reduced to valence; no phrasal constructions or lexical entries are needed; no competition between morphology and syntax (e.g. Poser, 1992; Bresnan, 2001) needs to be orcherstrated. Second, since (33) is an inflectional realization rule, it interacts with other such rules under the logic of rule specifity independently needed for affixal exponence. For instance, the fact that the present perfect is syncretic with the (synthetic) indirect bounded past can be accounted for by the rule in (34), which overrides the application of (33) because of specificity.

(34)
$$\begin{bmatrix} \text{PHON} & X \\ \text{LID} & Y \\ \text{VAL} & Z \end{bmatrix}, \sigma : \begin{bmatrix} \text{TNS} & \textit{prst} \\ \text{PRF} & + \end{bmatrix} \longrightarrow$$
$$\begin{bmatrix} \text{PHON} & \mathsf{refer}\left(\begin{bmatrix} \text{PHON} & \text{X} \\ \text{LID} & Y \\ \text{VAL} & Z \end{bmatrix}, \sigma \setminus \begin{bmatrix} \text{ASP} & \textit{bnd} \\ \text{PRF} & - \\ \text{EVID} & \textit{ind} \end{bmatrix}, \text{I-V} \right) \\ \text{LID} & Y \\ \text{VAL} & Z \end{bmatrix}$$

5 The future

The future shares many of the properties of the periphrastic perfect forms, two notable exceptions being the morphological makeup of the auxiliary and nonfinite form, and the order

[3] This is a portmanteau rule of referral covering blocks I to V, thus bypassing completely synthetic exponence. $\sigma \setminus \tau$ is the description that is identical to σ except where the features mentioned in τ differ from those in σ. As the reader can check, the meta-constraint in (9) makes sure that the valence requirement added by the realization rule will indeed constrain the valence of the word within the HPSG grammar.

between them (see the introduction). It can be accounted for using a rule that is quite similar to (33), and LP rules sensitive to verb form to account for the differing order. For lack of space we can not provide more details in this abstract.

6 The progressive

All unbounded forms may give rise to a progressive interpretation, but that interpretation can also be forced by using the periphrastic construction illustrated in (4). Unlike the ones we discussed so far, this construction results from the grammaticalization of a finite complement clause construction, and all relevant evidence points to the fact that an embedded clausal structure is still present.[4] The nonauxiliary verb is unmistakably a finite form; it occurs on the right of the auxiliary, as finite complement clauses occur on the right of their head. No complementizer can be used, but complementizers are optional for finite complements (35). Complements normally occur between the two verbs; they can scramble to the left of the auxiliary, but this is also possible with clausal complements (36). Finally, object clitic pronouns must be realized on the nonauxiliary verb, and cannot climb to the auxiliary (37).

(35) a. Maryam dâr-ad (*ke) ketâb
Maryam have.S1-3SG COMP book
mi-xân-ad
UNBD-read.S1-3SG
'Maryam is reading a book.'

b. Maryam mi-xâh-ad (ke) bâ Omid
Maryam UNBD-want.S1-3SG COMP with Omid
har ruz be sinemâ be-rav-ad
every day to theatre SBJV-go.S1-3SG
'Maryam wants to go to theatre with Omid everyday.'

(36) a. Maryam in ketâb=râ dâr-ad
Maryam this book=DDO have.S1-3SG
mi-xân-ad
UNBD-read.S1-3SG
'Maryam is reading this book.'

b. Maryam bâ Omid mi-xâh-ad (ke)
Maryam with Omid UNBD-want.S1-3SG COMP
har ruz be sinemâ be-rav-ad
every day to theatre SBJV-go.S1-3SG
'Maryam wants to go to theatre with Omid everyday.'

(37) a. Maryam dâr-ad mi-xân-ad=aš
Maryam have.S1-3.SG UNBD-read.S1-3SG-=3SG
'Maryam is reading it.'

b. * Maryam dâr-ad=aš mi-xân-ad
Maryam have.S1-3SG=3SG UNBD-read.S1-3SG

This data can be accounted for by assuming a slightly idiosyncratic lexemic entry for the auxiliary *dâštan*. This entry assumes that *prog* is a subtype of the ASPECT value *unbd* (unbounded). As a result of its lexeme-level specification, this auxiliary is defective for all subparadigms except the present, the unbounded past and the complex unbounded past, in accordance with the facts.

(38)
$$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{LID} & \textit{dâštan-aux} \\ \text{MORSYN} & \boxed{1}\begin{bmatrix} \text{ASP} & \textit{prog} \end{bmatrix} \end{bmatrix} \\ \text{CONT} & \boxed{2} \\ \text{VAL} & \left\langle \boxed{3}, \begin{bmatrix} \text{MORSYN} & \boxed{1} \\ \text{MARKING} & \textit{none} \\ \text{CONT} & \boxed{2} \\ \text{VAL} & \langle \boxed{3} \rangle \end{bmatrix} \right\rangle \end{bmatrix}$$

References

Ackerman, F. and Stump, G. T. (2004). 'Paradigms and periphrastic expression'. In L. Sadler and A. Spencer (eds.), *Projecting Morphology*. Stanford: CSLI Publications, 111–157.

Ackerman, F. and Webelhuth, G. (1998). *A Theory of Predicates*. Stanford: CSLI Publications.

Bresnan, J. (2001). *Lexical-functional syntax*. Oxford: Basil Blackwell.

Gazdar, G., Klein, E., Pullum, G. K., and Sag, I. A. (1985). *Generalized Phrase Structure Grammar*. Harvard: Harvard University Press.

Hinrichs, E. and Nakazawa, T. (1994). 'Linearizing AUXs in German verbal complexes'. In J. Nerbonne, K. Netter, and C. Pollard (eds.), *German in HPSG*. Stanford: CSLI Publications, 11–37.

Jahani, C. (2000). 'Expressions of indirectivity in spoken Modern Persian'. In L. Johanson and B. Utas (eds.), *Evidentials: Turkic, Iranian and Neighbouring Languages*. Berlin: Mouton de Gruyter, 185–207.

Lazard, G. (1985). 'L'inférentiel ou passé distancié en persan'. *Studia Iranica*, 14:27–42.

Lazard, G., Richard, Y., Hechmati, R., and Samvelian, P. (2006). *Grammaire du persan contemporain*. Tehran: IFRI and Farhag-e Moaser.

Poser, W. J. (1992). 'Blocking of phrasal constructions by lexical items'. In I. A. Sag and A. Szabolcsi (eds.), *Lexical matters*. Stanford: CSLI Publications, 111–130.

Sag, I. A. (2007). 'Sign-based construction grammar: An informal synopsis'. Ms., Stanford University.

Spencer, A. (2005). 'Inflecting clitics in generalized paradigm function morphology'. *Lingue e Linguaggio*, 4:179–193.

Stump, G. T. (2001). *Inflectional Morphology. A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.

Windfuhr, G. (1982). 'The verbal category of inference in Persian'. In *Momentum Georg Morgenstierne II*. Leiden: E. G. Brill, 263–287.

[4] Persian raising and control constructions normally rely on a finite unsaturated complement clause. Infinitival complements are available only in a very formal register.

Construction-based adjunct extraction

Rui P. Chaves
University at Buffalo, SUNY

HPSG 2009

1 Introduction

Certain modifier phrases can be extracted, as noted by Pollard and Sag (1994). This is the case of adverbial phrases and prepositional phrases, as illustrated below.

(1) a. [On Monday], I think that [Kim went home very late _].

b. [Yesterday], it seems that [Kim arrived home very early _].

c. [How often] do you think that [Fred was late this week _]?

(2) a. *[Tall], I saw the _ man.

b. *[Who shot the sheriff], I met the man _ .

Pollard and Sag (1994, 387) propose a lexical rule that applies to verbs and adds adjuncts to SLASH. This account was later revised in Bouma et al. (2001). The latter assumes that post-verbal modifiers are in fact complements. And since these elements are complements, then adjunct extraction is handled in exactly the same way as the extraction of arguments.

The adverbs-as-complements analysis finds some cross-linguistic evidence in the literature, as noted in Bouma et al. (2001). However, this view is not without problems. First, there is no semantic sense in which these adverbs can be seen as verbal arguments.[1] Second, Levine (2003) notes that this analysis requires further assumptions in order to account for cases that would otherwise be standard VP adjunction, as in (3).

(3) Nobody can [[drink four beers and eat two hotdogs] [under fifteen seconds]].

If the PP is a complement, then something else must be assumed in order to capture this sentence. For example, one would have to assume that the PP is extraposed ATB or Right-Node Raised. But these hypotheses are at odds with the semantic interpretation that the PP obtains, which ranges over the total time interval denoted by the two conjuncts. In contrast, this reading is trivially obtained if the PP simply adjoins to the conjoined VP.

(4) Nobody can [[drink four beers]$_{e_1}$ and [eat two hotdogs]$_{e_2}$]$_{e_1+e_2}$ [under fifteen seconds].

There is good evidence that plurality-forming conjunction operates beyond the nominal part of speech, and that it can also form *event pluralities* (Bach 1986; Lasersohn 1995; Link 1998), as illustrated below. Sentence (5a), adapted from Oehrle (1987), describes the frequency of two joint event-types, not of independent frequencies of occurrence.

(5) a. Often, [[I go to the beach]$_{e_1}$ and [you go to the city]$_{e_2}$]$_{e_1+e_2}$.

[1] In the final version of this paper I will discuss case marking of adverbial nominals.

b. Sue [[[got dressed]$_{e_1}$ and [dried her hair]$_{e_2}$]$_{e_1+e_2}$, [in exactly twenty seconds]].

c. You can't simultaneously [[drink]$_{e_1}$ and [drive]$_{e_2}$]$_{e_1+e_2}$.

Levine (2003) raises further objections about the 'adverbs as complements' analysis, with the datum in (6), where the extracted phrase *In how many seconds flat* obtains a collective reading over the three events denoted by the entire embedded coordinate VP, not over each conjunct.

(6) In how many seconds flat do you think that [Robin found a chair, sat down and took off her logging boots]?

This utterance is a query about the total time occupied by the occurrence of three (possibly overlapping) events. Such a reading suggests that the topicalized constituent is not a complement of anything in the sentence. If it were, then the adjunct should be predicating over each of the conjuncts separately, not the higher VP coordination node.

Levine (2003) – but see also Sato and Tam (2008) – propose that slash paths are terminated by traces, and therefore a modifier like the one in (7) can trivially be adjoined to a conjoined VP. When it does, it adds itself as an unbounded dependency, which is percolated and linked to a filler like any other unbounded dependency.

(7)
$$\begin{bmatrix} \textsc{phon} \langle\rangle \\ \textsc{synsem} \begin{bmatrix} \textsc{loc} \boxed{1}\begin{bmatrix} \textsc{head} \ adv \end{bmatrix} \\ \textsc{slash} \left\{ \boxed{1} \right\} \end{bmatrix} \end{bmatrix}$$

There are two problems for this approach. First, the introduction of traces forces the grammar to become more complex in order to prevent overgeneration in coordination. In particular, the grammar will require extra constraints to capture the Conjunct Constraint, which falls out as a prediction of a traceless grammar (Bouma et al. 2001). See also Sag (2000) for a review of the standard arguments for the existence of traces, and their problems. A second issue is that nothing in Levine (2003) prevents the adverb trace in (7) from being adjoined to each of the VP conjuncts, instead of the coordinate mother VP. If it assumed – as standardly is assumed – that SLASH values are structure-shared between daughters and mother in symmetric coordination, then one would obtain an impossible interpretation where each conjunct event is the same. This is illustrated schematically in Figure 1. The problem that Levine (2003) falls prey is related to the one discussed in Levine and Hukari (2006, 159), where structure-sharing slashed event-modifying adjuncts originating in each conjunct yield a description that no sign can satisfy.

That is *not* to say that distributive interpretations of the adverb do not exist. These are a different matter entirely. For example, one can argue that an adverb like *yesterday* is distributive in the sense that when applying to a sum of events like $e_1 + e_2$ then it does not yield a collective predication (i.e. $\mathit{yesterday}(e_1 + e_2)$), but rather a distributive one (i.e. $\forall e \in e_1{+}e_2 \rightarrow \mathit{yesterday}(e)$). Which adverbs are distributive, which are collective, and which are ambiguous is a matter of lexical specification, similar to how verbs like *smile*, *meet*, and *hire* can interpret their pluralic NP arguments in different ways. In my view this is what is happening in the ambiguous example from Levine and Hukari (2000,186), shown below. Although the PP attaches to the higher VP coordination, it can apply distributively to each event in the sum $e_1 + e_2$, or collectively.

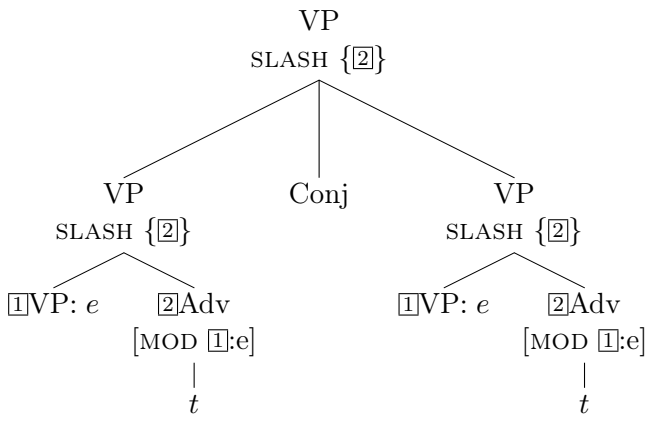


Figure 1: ATB adverbial extraction and impossible descriptions

(8) Robin [stands on his head and falls off his chair]$_{e_1+e_2}$ in order to attract people's attention.

Sag (2005) proposes that the extracted phrase *in how many seconds flat* is a complement, but one that is allowed to semantically outscope the verb structure that it modifies. The idea is that in a coordinate structure, it should follow that the PP extracted ATB is required to simultaneously outscope each of the verbs heading the conjuncts, thus obtaining wide scope over the entire coordination. The main problem of this approach, in my view, is that there is no evidence that the modifier *in X seconds* is semantically scopal. Consider the examples in (9). The intersective modifiers in (9a) do not give rise to scope ambiguities, while the scopal modifiers in (9b) trigger an ambiguity with respect to the wide or narrow scope interpretation of the indefinite subject:

(9) a. A spy photographed the documents yesterday / in twelve seconds.

b. A spy probably / usually photographed the documents.

Below I propose construction-based account of (6) which has nothing special to say about coordination, applies to both intersective and scopal modifiers, does not rely on traces, does not have to assume the adjuncts-as-complements analysis, and does not require extra assumptions about rightward extraction.

2 A traceless constructional formulation

In this section I provide what is in effect a rule-based traceless version on Levine (2003). Although this rule is for English, the approach can also be adapted to other languages, by making it either a lexical rule or a rule that applies to sentences rather than VPs, depending on the language. Furthermore, the usage of a feature ADJ as in Levine and Hukari (2006) or Sato and Tam (2008) can ensure that in gap threading languages the verb can have access modifiers higher up in the tree structure, as their SLASH values.

Following the grammar fragment in Ginzburg and Sag (2000) in general terms, I propose a unary-branching phrasal rule that allows adjuncts to modify a VP as members of SLASH:

(10) Adjunct Extraction Rule:

$$adj\text{-}extr\text{-}phr \Rightarrow \left[\begin{array}{l} \textsc{synsem} \left[\begin{array}{l} \textsc{loc}\ \boxed{1} \\ \textsc{slash}\ \boxed{2} \cup \left\{ \textsc{xp}\left[\textsc{mod}\ \boxed{3}\right] \right\} \end{array}\right] \\ \textsc{dtrs} \left\langle \boxed{3} \left[\textsc{synsem} \left[\begin{array}{l} \textsc{loc}\ \boxed{1}\left[\begin{array}{ll} \textsc{head} & \textit{verb} \\ \textsc{spr} & \langle [\ldots] \rangle \\ \textsc{comps} & \langle\rangle \end{array}\right] \\ \textsc{slash}\ \boxed{2} \end{array}\right]\right] \right\rangle \end{array}\right]$$

The Generalized Head Feature Principle will percolate the head's slash values in the tree structure. Moreover, since (10) is independent from coordination, it can also operate in non-coordinate cases. The account is illustrated in Figure 2. The VP daughter is modified by an adverbial phrase located in SLASH. Thus, *yesterday* modifies the event plurality that VP conjunction yields.

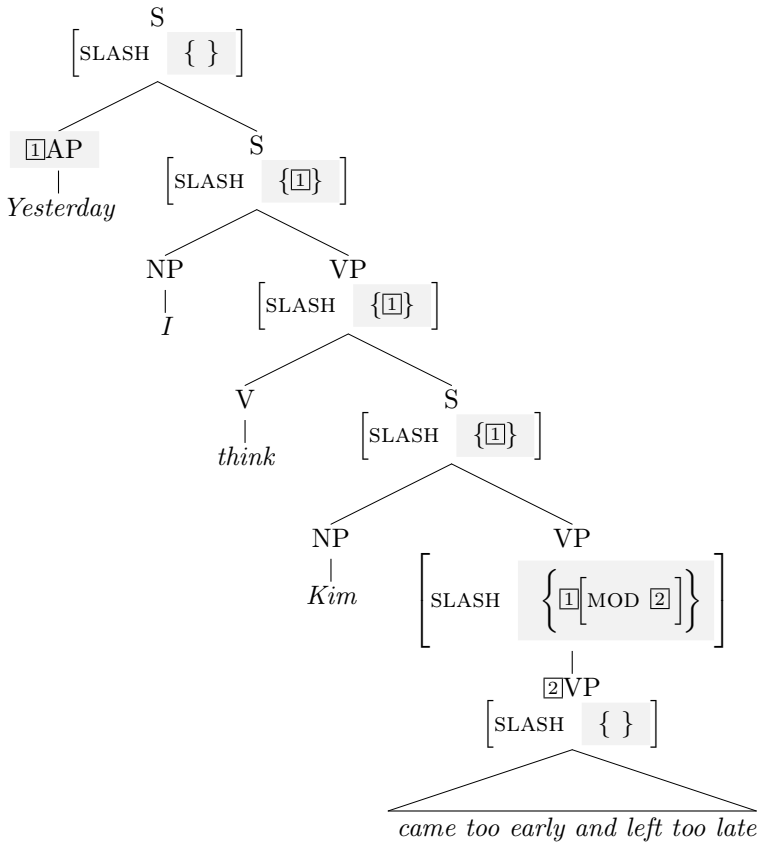


Figure 2: A modifier adjoining into SLASH

The next question to ask is what prevents ATB adjunct extraction as discussed in §1. In Copestake et al. (2006), existentially quantification of events is often omitted for simplification purposes, but if one makes such quantification explicit – as in $\exists e\ \mathit{left}(e, \mathit{kim})$ – then the adjunct extraction rule cannot apply to each VP conjunct because the resulting seman-

tic structure is ill-formed, according to general MRS well-formedness conditions. Consider the parse in (11).

(11) [Under how many seconds flat] did Kim [pack _ and escape _]?

Each adjunct is located in the SLASH value of each VP conjunct, and predicates the respective event. The modifiers must be one and the same at the coordination level, because the coordination rule imposes identity of SYN values (as for example, in Beavers and Sag (2004)). Thus, the adverbial phrase filler has to predicate the very same event across conjuncts, and has to be simultaneously located under the scope of each existential quantifier. This yields an ill-formed MRS structure because the underspecified representation cannot describe a tree structure, as depicted in Figure 3. Arrows denote underspecified semantic subordination constraints. Here, $\boxed{k}$ is the shared subject index of *Kim*, and $\boxed{e'} = \boxed{e1} = \boxed{e2}$. Both of these facts are consequences of the coordination rule.

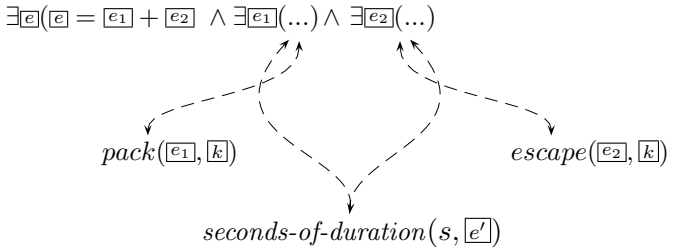


Figure 3: ATB adjunct extraction yielding an illegal MRS representation

3 Conclusion

This paper shows how a traceless construction-based account of modifier extraction in HPSG can be obtained for English, without some of the shortcomings of other accounts. There is no need to postulate traces, rightward extraction, or that non-scopal modifiers are scopal complements. The account boils down to a construction that can introduce adverbial unbounded dependencies in verbal structures.

4 Future work

There is another route for dealing with cumulative readings of extracted modifiers that has not been explored in the literature. I will argue that in this view, the phenomenon of cumulativity in adverbial extraction is but a special case of a more general mechanism that shows up elsewhere in the grammar, and that it has nothing to do with extraction or with modifiers. This traceless approach noes not fall prey to the problems reviewed in this paper, and dispenses extra grammar rules such as the ADJUNCT EXTRACTION RULE as well as extra features like ADJS.

References

Bach, E. (1986). The algebra of events. *Linguistics and Philosophy 9*, 5–16.

Beavers, J. and I. A. Sag (2004). Ellipsis and apparent non-constituent coordination. In S. Müller (Ed.), *Proceedings of the 11th International Conference on Head-Driven Phrase Structure Grammar, Katholieke Universiteit Leuven*, pp. 48–69. Stanford: CSLI Publications. `http://cslipublications.stanford.edu/HPSG/5/`.

Bouma, G., R. Malouf, and I. A. Sag (2001). Satisfying constraints on extraction and adjunction. *Natural Language and Linguistic Theory 1*(19), 1–65.

Copestake, A., D. Flickinger, I. A. Sag, and C. Pollard (2006). Minimal recursion semantics: An introduction. *Journal Research on Language & Computation 3*(4), 281–332.

Ginzburg, J. and I. A. Sag (2000). *Interrogative Investigations: the form, meaning and use of English interrogative constructions*. Stanford: CSLI Publications.

Lasersohn, P. (1995). *Plurality, Conjunction and Events*. Kluwer: Dordrecht.

Levine, R. and T. Hukari (2006). *The Unity Of Unbounded Dependency Constructions*. CSLI Publications, Leland Stanford Junior University.

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

Link, G. (1998). *Algebraic Semantics in Language and Philosophy*. CSLI Lecture Notes No. 74.

Oehrle, R. T. (1987). Boolean properties in the analysis of gapping. In G. J. Huck and A. E. Ojeda (Eds.), *Syntax and Semantics 20: Discontinuous Constituency*, pp. 201–240. Academic Press, Orlando, FL.

Pollard, C. and I. A. Sag (1994). *Head-driven Phrase Structure Grammar*. Chicago: University of Chicago Press and Stanford: CSLI.

Sag, I. A. (2000). Another argument against *wh*-trace. Jorge Hankamer Webfest. `http://ling.ucsc.edu/Jorge/sag.html`.

Sag, I. A. (2005). Adverb extraction and coordination: a reply to Levine. In S. Müller (Ed.), *The Proceedings of the 12th International Conference on Head-Driven Phrase Structure Grammar, Department of Informatics, University of Lisbon*, Stanford, pp. 322–342. CSLI Publications.

Sato, Y. and W. L. Tam (2008). Towards a unified account of adjuncts. In S. Müller (Ed.), *The Proceedings of the 15th International Conference on Head-Driven Phrase Structure Grammar*, Stanford, pp. 173–188. CSLI Publications.

Deriving superficial ergativity in Nias

Berthold Crysmann

Universität Bonn
crysmann@ifk.uni-bonn.de

Case assignment in the Indonesian language Nias (Brown, 1997, 2005; Donohue and Brown, 1999) is standardly assumed to belong to the ergative type. The main evidence for this typological categorisation comes from the fact that case marking of the S(ole) argument in intransitives patterns with that of the O(bject) argument in transitives. The A(gent) of transitives, however, features case marking which is clearly distinct from the S(ole) argument of intransitives, yielding a partitioning characteristic of other ergative languages.

(1) manavuli sui [n-ama-da Tohonavanaetu] ba
return again MUT-father-1.PL.IN.GEN Tohönavanaetu LOC
Maenamölö
Maenamölö

'Ama Tohonavanaetu came back again to Maenamölö.' (Donohue and Brown, 1999)

(2) I-a [m-bavi] [ama Gumi]
3.sg.RLS-eat MUT-pig father Gumi

'Father Gumi eats pig.' (Donohue and Brown, 1999)

Mutation in Nias is not a mere surface-phonological effect: as detailed in Donohue and Brown (1999), obliques can intervene between the verb and the mutated S or O argument, as shown below.

(3) I-be khö-nia g-ana'a.
3SG.RLS-give OBL-him MUT-gold

'He gave him (the) gold.' (Donohue and Brown, 1999)

Further evidence against a surface-phonological view is corroborated by the fact that Nias also features constructions with two mutated NPs, e.g. with experiencer verbs like *ta'u* 'fear', or in the imperfective of transitive verbs.

(4) A-ta'u m-ba'e n-ono matua
ST-fear MUT.monkey MUT.child

'The monkey is afraid of the boy.' (Brown, 2005)

Besides case marking morphology for nouns, the ergative pattern is also salient in the pronominal system: as detailed in Brown (1997), pronominal A arguments are realised as prefixes on the verb (in the realis), whereas both S and O arguments are realised as independent pronominals in their canonical syntactic position.

What is striking about Nias, however, is that absolutive case is morphologically marked (by mutation), wheras ergative is the unmarked case. Similarly, it is the (unmarked) ergative case which is used in citation contexts, for topicalised arguments, or in elliptical answers consisting of a single NP (Brown, 1997). According to Donohue and Brown (1999), this constitutes a clear exception to Dixon's (1994) claim that absolutive case should universally be expressed by unmarked case morphology.

Agreement in Nias is likewise peculiar: while ergative languages standardly show agreement with the absolutive argument (cf. Corbett, 2006), or else with the thematically highest argument (e.g., Udi; Harris, 1984), verbs in Nias agree with the ergative in the realis mode.

(5) a. I-tolo zi'ila ama-gu
3SG.RLS-help MUT.village.advisor father-1SG.POSS
'My father helped the village advisors.' (Brown, 2003; Corbett, 2006)

b. La-tolo n-ama-gu si-ila
3PL.RLS-help MUT-father-1SG.POSS village.advisor
'The village advisors helped my father.' (Brown, 2003; Corbett, 2006)

If no ergative is present, as with intransitives, agreement marking in the realis will be null (Brown, 2003, 2005; Corbett, 2006).

(6) Mofanö n-ama-gu
leave MUT-father-1SG.POSS
'My father left.' (Brown, 2003; Corbett, 2006)

Agreement in the realis thus appears to follow the divide between the A argument on one side and O and S arguments on the other (Donohue and Brown, 1999). In the irrealis, agreement is controlled by the highest role (S or A), thus following the pattern exemplified by Udi.

(7) Ya-ma-nana n-ono-nia ba va-a-lio.
3SG.I-DYN-hand MUT-child-3SG.POSS LOC MUT.NR-ST-quick
'Her child will be crawling soon.'

(8) Ya-mbu'a g-ömö-nia
3SG.I-repay MUT-debt-3SG.POSS
'He might repay his debt.'

Irrespective of the agreement pattern, case marking in these two modes is identical, featuring unmutated A arguments and mutated S and O arguments.

If we look at transitives in isolation, however, a different pattern emerges: the A argument appears to have all the characteristic properties of subjects in nominative languages, controlling agreement of the verb and featuring unmarked case morphology. This perspective of the "ergative" argument as the subject is further supported by the observation that it appears in a peripheral position: in a contoured VP, one would actually expect subjects to occur outside non-subject complements. As stated in Himmelmann (2005), VXS basic word order, together with evidence for a VP constituent is a common typological pattern in Western Austronesian languages, whereas VSO is not.

Adopting the idea that Nias ergatives more closely resemble nominative arguments in other languages, it remains to be explained, however, why the S argument of intransitives does not pattern with A arguments, as should be expected in a nominative-accusative system. Taking marked case morphology and lack of agreement as indicative, I suggest that intransitives are actually subject-less verbs, their S argument being mapped to direct object function. Subject-less verbs assigning objective case to their only argument are not unattested cross-linguistically, cf. German verbs like *frieren*. What is special about Nias, then, is that a linking pattern that can only be observed with lexical case assignment in German is grammaticalised in Nias, applying to structural case assignment of intransitives.

The hypothesis that S arguments, or more generally, mutated arguments in Nias are indeed non-subject complements is also supported by language-internal facts: as detailed in Brown (1997), possessor arguments of nouns, as well as complements of most prepositions also appear in mutated form. Conversely, unmutated case functions as the default case in a variety of constructions: besides its uses as the citation form or in elliptical answers, it features in topicalised structures, as well as in embedded contexts.

Taking as a starting point the division between argument structure and valence underlying the theory of ergativity advanced by Manning and Sag (1999); Manning (1994), I shall argue that Nias constitutes a third type, featuring canonical direct mapping of transitives, but non-canonical mapping of intransitives.

(9) *transitive-verb-lxm* →

$$\left[\begin{array}{l} \textsc{synsem}\,|\,\textsc{local}\,|\,\textsc{cat}\left[\textsc{val}\left[\begin{array}{ll}\textsc{subj} & \langle \boxed{1} \rangle \\ \textsc{comps} & \langle \boxed{2}\,|\,\boxed{3} \rangle\end{array}\right]\right] \\ \textsc{arg-st}\left\langle \boxed{1}\left[\textsc{case}\quad \textit{str}\right], \boxed{2}\left[\textsc{case}\quad \textit{str}\right]\,|\,\boxed{3}\right\rangle \end{array}\right]$$

(10) *intransitive-verb-lxm* →

$$\left[\begin{array}{l}\textsc{synsem}\,|\,\textsc{local}\,|\,\textsc{cat}\ \left[\begin{array}{ll}\textsc{val} & \left[\begin{array}{ll}\textsc{subj} & \langle\rangle\\ \textsc{comps} & \left\langle \boxed{1}\,|\,\boxed{2}\right\rangle\end{array}\right]\end{array}\right]\\ \textsc{arg-st}\left\langle \boxed{1}\left[\textsc{case}\quad \textit{str}\right]\,|\,\boxed{2}\right\rangle\end{array}\right]$$

Once we adopt the argument mapping sketched above, constraints on agreement and case assignment can be formulated quite straightforwardly, assigning basic unmutated case to subjects, and mutated case to members of COMPS:

(11)
$$\left[\textsc{synsem}\,|\,\textsc{loc}\left[\textsc{val}\quad\left[\textsc{subj}\left\langle\left[\textsc{case}\quad \textit{str}\right]\right\rangle\right]\right]\right]$$
$$\rightarrow\left[\textsc{synsem}\,|\,\textsc{loc}\,|\,\textsc{val}\,|\,\textsc{subj}\left\langle\left[\textsc{case}\quad \textit{basic}\right]\right\rangle\right]$$

(12)
$$\left[\textsc{synsem}\,|\,\textsc{loc}\,|\,\textsc{cat}\left[\textsc{val}\quad\left[\textsc{comps}\left\langle\ldots\ \boxed{1}\left[\textsc{case}\quad \textit{str}\right]\ldots\right\rangle\right]\right]\right]$$
$$\rightarrow\left[\textsc{synsem}\,|\,\textsc{loc}\,|\,\textsc{cat}\,|\,\textsc{val}\,|\,\textsc{comps}\left\langle\ldots\ \boxed{1}\left[\textsc{case}\quad \textit{mutated}\right]\ldots\right\rangle\right]$$

Similarly, agreement in the realis mode will target the subject valency, selecting default zero agreement for (subject-less) intransitives.

(13)
$$\left[\textsc{synsem}\,|\,\textsc{loc}\left[\textsc{cat}\left[\begin{array}{l}\textsc{head}\left[\textsc{vform}\quad \textit{realis}\right]\\ \textsc{val}\,|\,\textsc{subj}\left\langle\left[\textsc{loc}\,|\,\textsc{cont}\,|\,\textsc{index}\ \boxed{i}\right]\right\rangle\end{array}\right]\right]\right]$$
$$\rightarrow\left[\textsc{synsem}\,|\,\textsc{loc}\,|\,\textsc{cat}\left[\textsc{head}\,|\,\textsc{agr}\ \boxed{i}\right]\right]$$

(14)
$$\left[\textsc{synsem}\,|\,\textsc{loc}\left[\textsc{cat}\left[\begin{array}{l}\textsc{head}\left[\textsc{vform}\quad \textit{realis}\right]\\ \textsc{val}\,|\,\textsc{subj}\ \langle\rangle\end{array}\right]\right]\right]$$
$$\rightarrow\left[\textsc{synsem}\,|\,\textsc{loc}\,|\,\textsc{cat}\left[\textsc{head}\,|\,\textsc{agr}\quad \textit{none}\right]\right]$$

In the irrealis, by contrast, agreement will be with the highest member on ARG-ST:

(15)
$$\left[\textsc{synsem}\,|\,\textsc{loc}\left[\textsc{cat}\left[\begin{array}{l}\textsc{head}\left[\textsc{vform}\quad \textit{irrealis}\right]\\ \textsc{arg-st}\left\langle\left[\textsc{loc}\,|\,\textsc{cont}\,|\,\textsc{index}\ \boxed{i}\right],\ \ldots\right\rangle\end{array}\right]\right]\right]$$
$$\rightarrow\left[\textsc{synsem}\,|\,\textsc{loc}\,|\,\textsc{cat}\left[\textsc{head}\,|\,\textsc{agr}\quad \boxed{i}\right]\right]$$

The non-canonical mapping for intransitives can also be applied to account for case assignment and agreement with experiencer verbs like 'fear', which do not take any unmutated argument, and therefore, in the realis, do not display agreement morphology.

The imperfective of transitives (marked with *maN-*) constitutes another instance of double mutated case marking:

(16) Man-uri zawi ya
IPF-keep.alive MUT.cattle MUT.3SG
'He keeps cattle.' (Brown, 2005)

As expected, agreement is null in the realis mode, and the A argument is realised by a free pronoun, not by a bound pronominal. The imperfective can easily be integrated into our current analysis by means of the following lexical rule:

(17)
$$\left[\text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \left[\text{VAL} \begin{bmatrix} \text{SUBJ} & \left\langle \boxed{1} \right\rangle \\ \text{COMPS} & \boxed{2} \end{bmatrix}\right]\right]$$
$$\mapsto \left[\text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \begin{bmatrix} \text{HEAD} & \left[\text{VFORM} \quad \textit{imperf}\right] \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \langle\rangle \\ \text{COMPS} & \boxed{2} \oplus \left\langle \boxed{1} \right\rangle \end{bmatrix} \end{bmatrix}\right]$$

So far, we have not addressed default case assignment, as witnessed, e.g., with topicalised, pre-verbal constituents which receive unmutated case. A conceivable approach would be to assume that case assignment for these constituents applies to members of SLASH:

(18)
$$\left[\text{SYNSEM} \mid \text{NLOC} \mid \text{INH} \left[\text{SLASH} \left\{ \boxed{1} \left[\text{CAT} \mid \text{HEAD} \mid \text{CASE} \quad \textit{str}\right], \ldots \right\}\right]\right]$$
$$\rightarrow \left[\text{SYNSEM} \mid \text{NLOC} \mid \text{INH} \left[\text{SLASH} \left\{ \boxed{1} \left[\text{CAT} \mid \text{HEAD} \mid \text{CASE} \quad \textit{basic}\right], \ldots \right\}\right]\right]$$

To conclude, once we assume that mapping of intransitives is non-canonical in Nias, both agreement and case assignment will represent typologically unmarked options, referring to the central notion of subject. Lack of agreement in intransitives will thus be reduced to the absence of a subject. The alternative view (Handschuh, 2008; Wichmann, 2005) which treats Nias as the mirror image of marked nominative languages, however, not only assigns Nias to an extremely rare linguistic type (marked absolutive), but also needs to postulate that agreement in this type should differ drastically from the patterns standardly observed in the languages of the world. Extending

the linking theory of Manning and Sag (1999); Manning (1994) which derives ergative-absolutive patterns in transitives by an inverted mapping of thematic arguments to grammatical function, I have argued that superficial ergativity in Nias emerges as a result of non-canonical mapping in intransitives.

References

Brown, L. (1997), "Nominal mutation in Nias," in C. Odé and W. Stokhof, eds., *Proceedings of the 7th Intl' Conference on Austronesian Linguistics*, pp. 395–413, Amsterdam, Editions Rodopi B.V.

Brown, L. (2003), "Nias: An exception to universals of argument marking," paper presented at the Association of Linguistic Typology 5, Sardinia, 15–18 Sep.

Brown, L. (2005), "Nias," in AlexanderAdelaar and N. Himmelmann, eds., *The Austronesian Languages of Asia and Madagaskar*, pp. 562–589, Routledge, London.

Corbett, G. (2006), *Agreement*, Cambridge UP, Cambridge.

Dixon, R. M. W. (1994), *Ergativity*, Cambridge UP, Cambridge.

Donohue, M. and L. Brown (1999), "Ergativity: Some additions from Indonesia," *Australian Journal of Linguistics* **19**(1), 57–76.

Handschuh, C. (2008), "Split marked-s case systems," in M. Richards and A. L. Malchukov, eds., *Scales*, no. 86 in Linguistische Arbeitsberichte, pp. 73–82, Universität Leipzig, Leipzig.

Harris, A. (1984), "Case marking, verb agreement, and inversion in Udi," in D. M. Perlmutter and C. G. Rosen, eds., *Studies in Relational Grammar 2*, pp. 243–258, University of Chicago Press, Chicago.

Himmelmann, N. (2005), "The Austronesian languages of Asia and Madagaskar: Typological characteristics," in AlexanderAdelaar and N. Himmelmann, eds., *The Austronesian Languages of Asia and Madagaskar*, pp. 110–181, Routledge, London.

Manning, C. and I. Sag (1999), "Dissociations between argument structure and grammatical relations," in A. Kathol, J.-P. Koenig and G. Webelhuth, eds., *Lexical and Constructional Aspects of Linguistic Explanation*, Studies in Constraint-based Lexicalism, CSLI Publications, Stanford.

Manning, C. D. (1994), "Ergativity: Argument structure and grammatical relations," Ph.D. thesis, Stanford.

Wichmann, S. (2005), "Tlapanec cases," in R. Beam de Azcona and M. Paster, eds., *Conference on Otomanguean and Oaxacan Languages, March 19-21, 2004, University of California at Berkeley*, pp. 133–145, Berkeley, University of California at Berkeley.

Inflectional morphology in Turkish VP-coordination

Antske Fokkens[♣] **Laurie Poulson**[♢] **Emily Bender**[♢]
[♣] Department of Computational Linguistics, Saarland University
[♢] Department of Linguistics, University of Washington

1 Introduction

This paper presents an analysis of the interaction between verbal morphology and VP coordination in Turkish. The interaction is of interest because of three general properties of the language: (i) VP conjuncts must agree in their tense and aspect features, (ii) tense/aspect (TAM) morphemes may appear only on the final conjunct, either obligatorily (in the *-ip* construction) or optionally (in *ve* coordination), and (iii) verbal affixes expressing necessity and possibility also take scope over the whole coordinated structure while attaching morphologically only to the final conjunct. Generalizations (i) and (ii) are shown to be straightforwardly analyzed in terms of local constraints on identity of TAM features. Generalization (iii) is more surprising, and would seem to point to the necessity for a notion of phrasal affixes. In fact, however, it is amenable to a constructional analysis along the lines of (Tseng, 2003).

The analyses proposed in this paper are implemented in a grammar fragment derived from the LinGO Grammar Matrix (Bender et al., 2002), through its web-based customization system.[1] Our grammar fragment focuses on capturing syntactic and semantic constraints on verbal morphology in coordination and uses semantic representations in the format of Minimal Recursion Semantics (Copestake et al., 2005, MRS). Like all Matrix-derived grammars, it is compatible with the LKB (Copestake, 2002).

The implementation confirms that the analyses manage to account for all examples presented below. It also shows the cross-linguistic applicability and the practical utility of the Grammar Matrix, because we were able to implement and test the analyses presented in this abstract quickly.

2 Verbal Morphology in Turkish

2.1 Properties of Turkish Verbs

This section describes some basic properties of verbal morphology in Turkish. We provide an overview of morphemes that may be added to the stem and present conditions of completeness and well-formedness of the verbs. The description is based on, among others, Kornfilt (1997) and Lewis (2000) and we base our interpretation of the data partially on the descriptions provided by Sezer (2001) and Kabak (2007).

Table 1: Morpheme slots of inflectional morphemes

1	2	3	4	5
-DI def. witn. past *-sE* subj. cond. *-mIş* inferential pres. perf *-Iyor* continuous *-yEcEG* future *-Ir/-Er* aorist *-mEli* neccesitive *-mEkte* continuous	*-(i)DI* def. witn. past *-(i)sE* ind. cond. *-mIş* inferential	*-(i)sE* ind. cond *-(i)mIş* inferential	AGR	*-DIr*

The distinction between derivational and inflectional morphemes is not clear cut in Turkish. Traditionally, morphemes that can be followed by the infinitive marker *-mEk* are considered derivational. Derivational morphemes are *-DIr/t* (causative), *-Il* (passive), *-(y)A* (abilitative), *-mA* (negation) and *-(y)Abil* (potentiality). In addition to the derivational morphemes, there are five slots that may host an inflectional morpheme. The inflectional morphemes are presented in Table 1. In the representation, capital letters represent phonemes whose realization depends on vowel harmony. A finite verb must bear an inflectional marker from slot 1 and an agreement marker. At least one inflectional marker must be phonologically overt (Kabak, 2007).[2]

Turkish has two paradigms of agreement markers: the *k*-paradigm for definite past and conditional (*-DI* and *-sE*, respectively) and the *z*-paradigm for all other TAM[3] morphemes. Which paradigm is used depends on the last TAM morpheme attached to the verb.

[1] http://www.delph-in.net/matrix/customize/matrix.cgi; accessed on 2/25/09

[2] Some linguists assume that secondary tense markers are hosted by an auxiliary suffix *-i/(y)* (Lees (1962) and Sezer (2001), among others), though this suffix has also been analyzed as a phonological element (Erguvanlı-Taylan, 1999). Our analysis is compatible with both views.

[3] Throughout the paper, we use the term TAM morphemes to refer to all inflectional morphemes in slots 1-3.

2.2 Verbal Morphology with Lexical Rules

The analysis we propose makes use of the morphotactic infrastructure added to the Matrix customization system by O'Hara (2008), which provides implementations for some wide spread phenomena in morphology. The grammar created this way only requires minor changes for the basic morphology to work.

The morphotactic support allows the definition of multiple morphological "slots" for each stem type. It provides implementations for optional and obligatory morphemes that may add syntactic and semantic features to the derived form. It also allows lexical rules to require earlier, or force later, slots as well as to forbid other slots from appearing. These properties are enforced by binary features on the verb that are related to specific morphological slots. For instance, if optional morpheme2 requires morpheme1 in order to be licensed, bare verbs will carry a feature [MORPHEME2 −]. The lexical rule associated with morpheme1 turns this value into +, which allows the (otherwise prohibited) morpheme2-rule to apply.

The morphotactic infrastructure in the customization system does not provide an analysis for the two agreement paradigms. In this case, we have an obligatory slot of morphemes that interact in a different ways with other morphemes. In order to account for the different agreement paradigms, we created two subtypes of the agreement-lexical-rule, and distinguished them with the binary feature AGR-PARADIGM. The three inflection-rules have two subtypes as well: one supertype of the so-called "true" tenses *-DI* and *-sE*, and one supertype for the other morphemes appearing in the same slot. Rules inheriting from the former supertype turn AGR-PARADIGM to +, whereas rules inheriting from the latter assign it the value −. The value of AGR-PARADIGM controls which agreement rule applies.

The analysis described above ensures that the right morphology is present on independent finite verb forms. In the next section, we present two structures that more or less correspond to VP-coordination in English. In these structures, the morphological requirements on a non-final conjunct differ from those on independent verbs. We propose an analysis along the lines of the basic morphological rules presented above.

3 Coordinated VPs

3.1 Suspended affixation and the "-ip" structure

Turkish has several structures that correspond largely to VP-coordination in English. Namely, simple juxtaposition, the coordination word *ve*, the coordination clitic *de*, and the suffix *-ip*.[4] These structures interact in different ways with the morphology of their conjuncts. We focus on the structures with the suffix *-ip* and the word *ve*, presented in examples (1) and (2).

(1) Çocuk-lar film izle-**yip** pizza
child-PL movie watch-COORD pizza
yi-yor-dı-lar.
eat-CONT-PAST-3PL
"The children were watching a movie and eating pizza."

(2) Çocuk-lar film izli-yor **ve** pizza
child-PL movie watch-CONT and pizza
yi-yor-dı-lar.
eat-CONT-PAST-3PL
"The children were watching a movie and eating pizza."

Both coordination structures share the property that all conjuncts must have the same tense, aspect and mood, even though they may not be overtly marked on non-final conjuncts. The difference lies in the morphological requirements of the first conjunct: the verb marked with *-ip* does not bear any other markers, whereas the progressive marker *-yor* is (obligatorily) repeated in the *ve* structure. In fact, adding or inserting any inflectional marker on a verb bearing *-ip* renders the sentence ungrammatical. In (2), reflecting the phenomenon of "suspended affixation", two of the three suffixes are only marked on the final verb. Additional inflection markers may be present on the preceding conjunct, as long as they are also found on the following conjunct.

Kabak (2007) poses as the primary condition for suspended affixation that the verb must end in a terminal morpheme.[5] Agreement morphemes are terminal, as well as TAM morphemes from the first inflectional slot, except for *-DI* and *-sE*. This difference between the "true tense" morphemes *-DI* and *-sE* on the one hand, and aspect morphemes such as *-Iyor* on the other hand is illustrated in the example below:

(3) Film izle-di-∅ ve pizza ye-di-m
movie watch-PAST-3SG and pizza eat-PAST-1SG
"(S)he watched a movie, and I ate pizza."

(4) Film izle-yor ve pizza yi-yor-um.
movie watch-CONT and pizza eat-CONT-1SG
"I am watching a movie and eating pizza."

[4] Some linguists consider verbs marked by *-ip* "converbs" (Tikkanen, 2001), though in descriptive literature (Lewis, 2000; Kornfilt, 1997) it is generally treated as a coordination marker. Empirical studies have, to our knowledge, not been able to settle the matter up to date. A converb would require a similar treatment when scopal affixes are at play, hence our decision to only present the coordination analysis in this abstract.

[5] A detailed discussion on suspended affixation in *ve* coordination can be found in Kabak (2007). For reasons of space, we limit our presentation to facts necessary to understand the analysis, which is compatible with all observations made by Kabak (2007).

In (3), suspended affixation is not possible. It can only be interpreted as two coordinated sentences. In example (4), on the other hand, both verbs are understood to have the same subject.

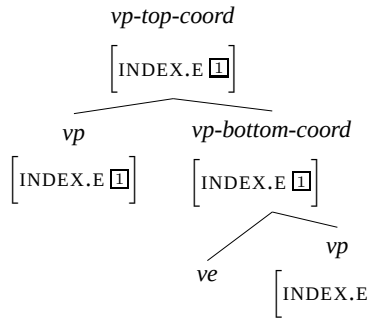
3.2 Analyzing coordinated VPs

The previous section explained that in the *-ip* structure, the verb bearing *-ip* receives the information expressed by inflectional morphemes from the conjunct that follows it. Coordinated VPs within the *ve* structure are also interpreted as if both had the inflection of the final conjunct, even if this may not be overtly present on the preceding verb. On our analysis, the *-ip* suffix takes the same slot as primary inflection morphemes. Evidence for this assumption comes from the fact that all derivational morphemes can precede *-ip*, but *-ip* cannot co-occur with any of the inflectional morphemes.

We assume that the required identity of tense morphemes is a semantic constraint (i.e. coordinated VPs must express events that take place in the same time, with the same mood, aspect, etc.), and implement this constraint via a sharing of semantic features (see below). The requirement that morphemes on a conjunct must be a subset of those on the following conjunct(s) is treated by morpho-syntactic constraints.

Just as for the verbal morphology, the coordination analysis here builds upon the implementation of coordination in the Matrix customization (Drellishak and Bender, 2005). A coordinated structure consists of a bottom-coord-phrase combining the coordination marker with the right element of the coordination, and a top-coord-phrase that adds the left conjunct as specifier, as in (5). Alternatively, when the coordination mark is inflection on the nonfinal conjuncts, the bottom-coord-phrase is a unary rule, and the top-coord-phrase joins appropriately inflected left conjuncts to bottom-coord-phrases, as in (10). In the Matrix definition of basic coordinated verb phrases, the TAM features of the coordinated phrase are identical to those of the right conjunct. Semantically ill-formed structures (i.e. structures in which left and right conjunct have a different TAM interpretation) can easily be excluded by sharing the TAM features of the left conjunct as well. Unification now fails when left and right conjunct provide conflicting semantics.

(5)



This enforces the right semantics, but optional morphemes can still be placed on either conjunct. Additional constraints are necessary to make sure all morphemes present on the first conjunct, are also found on the following one(s). Moreover, we need to make sure that the non-final conjunct ends in a terminal morpheme. For this purpose, we introduce a special lexical rule for nonfinal conjuncts. It takes a verbal form ending in a terminal morpheme as its daughter and creates a word that must be the left daughter of a coordinated structure. The rule is presented in reduced form below. The specification on DTR refers to the lexical rule that adds terminal morphemes to the stem.

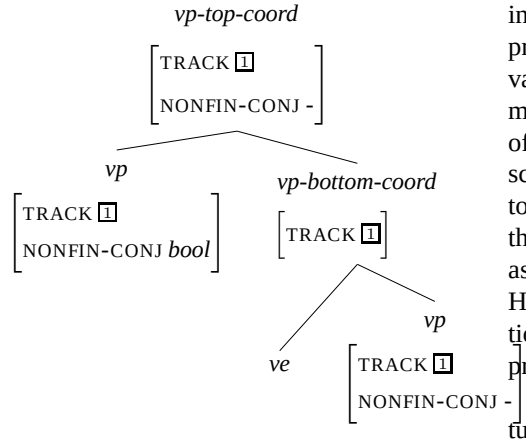
(6)
$$\begin{bmatrix} \text{TRACK} \begin{bmatrix} \text{2ND-INFL } \textit{luk} \\ \text{AGR } \textit{luk} \end{bmatrix} \\ \text{NONFIN-CONJ } + \\ \text{DTR } \textit{term-morph-infl-lex-rule} \end{bmatrix}$$

Here, we use the boolean feature NONFINAL-CONJ to make sure that appropriately inflected constituents appear as the left-hand conjunct in coordinated structures (with *ve*), and not elsewhere. When the nonfinal conjunct rule applies, NONFIN-CONJ is set to +. Subject-head phrases cannot have a head daughter whose NONFINAL-CONJ value is +. Coordinated phrases also require their right hand daughters to be [NONFINAL-CONJ −], but do not constrain this value on the left conjunct. The entire coordinated phrase has value [NONFINAL-CONJ -]. This way, verbs ending in a terminal morpheme that are not fully inflected can only be part of a well-formed structure if they appear as a left conjunct.

What remains is to ensure that the overtly expressed morphology on the left-hand conjunct is a subset of that on the right. To do this, we employ the substructure under the feature TRACK, posited by (O'Hara, 2008) to model dependencies between morphemes. Because we need three situations ("allowed", "prohibited", "applied") for some of the slots, we make use of the type *luk* (subtypes *na* and *bool(ean)*) for their value. When a lexical rule applies, its associated feature receives the value *na*.

The nonfinal conjunct lexical rule changes all values back to *luk*, as shown above. In the coordinated phrase, the TRACK features of both conjunct daughters must unify. If suspended affixation has applied, the *luk* values on the left daughter will unify, regardless of the values found on the right conjunct. Otherwise, the structure is only permitted if the values on both conjuncts are identical. Note that the NONFINAL-CONJ feature cannot be part of the TRACK features, since its value changes in the coordination structure.

(7)



This analysis has one draw-back: in the standard morphotactic analysis provided by the Matrix customization system, TRACK is part of words and lexemes only. In order to use it in phrases, we needed to upgrade it to signs. It seems, however, a property of the data that morphological properties have an impact on well-formedness at a phrasal level. In that sense, the expansion of the features to phrases reflects exactly what makes Turkish coordinated phrases so interesting.

4 Necessity and Ability

4.1 The scope of -(y)Abil and -mEli

Consider the following examples:

(8) Çocuk-lar film izle-yip pizza yi-meli-ler.
child-PL movie watch-COORD pizza eat-NEC-3.PL
"The children must watch a movie and eat pizza."

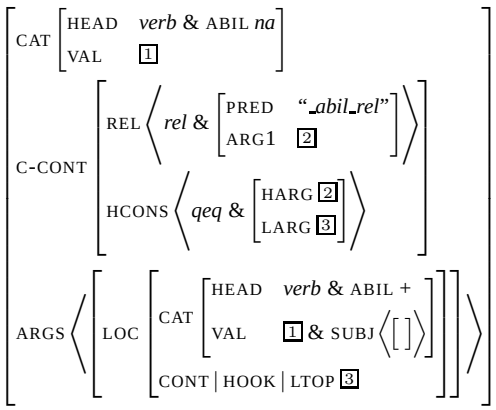
(9) Çocuk-lar film izle-yip pizza
child-PL movie watch-COORD pizza
yi-yebil-ir-ler.
eat-ABIL-AOR-3PL
"The children can watch a movie and eat pizza."

On our analysis, matching of TAM information is handled because these features of verbs are recorded as "variable properties" on their event variables. Following general practice in MRS (Copestake et al., 2005), the event variable of the elementary predication introduced by a verb is also "published" through the verb's INDEX value. Furthermore, this INDEX value is shared with larger constituents that are projections of that verb, and thus the coordination construction has access to the information it needs to ensure matching across conjuncts. However, the derivational affixes *-mEli* and *-(y)Abil* apparently contribute information that is usually handled in terms of (scopal) elementary predications: neccessity and ability. Thus it is more surprising to see this information shared across conjuncts.[6]

4.2 A Constructional Analysis

If we treat *-(y)Abil* and *-mEli* as predicate introducing morphemes, we cannot obtain the correct interpretation of coordinated VPs by simply sharing the value of both events. Nor can we just allow the semantics of these morphemes to attach "low"; Instead of (merely) the second verb, the suffixes must have scope over the entire coordinated VP.[7] This seems to suggest that these affixes attach to phrases rather than words, but "phrasal affixes" would violate the assumption of lexical integrity, generally held in HPSG. Instead, we propose a constructional solution, in the spirit of the analysis that Tseng (2003) proposes for apparent phrasal affixes in French.

Both *-(y)Abil* and *-mEli* contribute a HEAD feature, which is referenced by a special construction that takes a VP daughter and adds the potential or necessity semantics, respectively. The AVM below represents a simplified representation of the unary "ability"-phrase.



The ability predicate is added to the phrase through the C-CONT value.[8] The daughter VP is the first argument of the introduced predicate

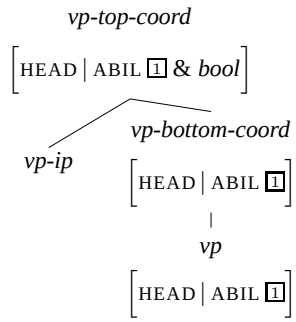
[6] Other derivational morphemes seem not have this property. According to Lewis (2000), the negation morpheme *-mA* also has wide-scope in the *-ip* structure, but none of the speakers we consulted got this reading.

[7] Furthermore, rather than an event variable, the relevant argument position takes a *handle* as its value, which is equal modulo quantifiers to the label of the coordinated VP (cf Copestake et al. (2005)).

[8] Per standard MRS solutions, the CONT value of a phrase is constructed from the CONT values of the daughter(s) and the C-CONT value of the rule licensing the phrase.

and falls under the predicate's scope constraints in HCONS. The construction assigns the value *na* to the HEAD features associated with the *-(y)Abil* inflection. When a phrase has an ability or necessity feature with value *na*, it cannot be the right daughter of an *-ip* coordination. On the other hand, head-subj phrases require that their head daughter's potential or necessity feature is *na* or -. These constraints ensure that the associated constructions can only apply after the coordinated VP has been formed, but must apply before the subject is added to the phrase if *-(y)Abil* or *-mEli* was present on the verb. The tree below represents the ABIL feature in an *ip*-coordination.

(10)



5 Conclusion

This paper presents an analysis for Turkish verbal morphology. The first part of the paper discusses an analysis for basic properties of Turkish inflectional morphemes, which can be implemented with help of the Matrix customization system with only minor changes.

The second part of the paper addresses the morphology of coordinated VPs. We show that both on a syntactic level, as well as on a semantic level, morphological processes partially operate on a phrasal level.

For the syntactic properties at hand (verification whether certain processes took place on each conjunct), passing features that keep track of morphological processes up to the phrase manage to accurately account for the data. But the scopal properties of suffixes *-(y)Abil* and *-mEli* seem to suggest that these affixes attach to phrases rather than words. This property would violate HPSG assumptions on lexical integrity. We show, however, that the data can be analyzed with the help of a construction.

The broader implications of this project are twofold: First, it can be seen as a case study in the cross-linguistic applicability and practical utility of the Grammar Matrix, which has allowed us to quite quickly produce a grammar fragment with which to test these ideas. Second, we have shown that the approach to apparent phrasal affixes of Tseng (2003) is not idiosyncratic to French, but also quite applicable to the unrelated language Turkish.

References

Bender, Emily, Flickinger, Dan and Oepen, Stephan. 2002. The Grammar Matrix: An Open-Source Starter-Kit fir the Rapid Development of Cross-linguistically Consistent Broad-Coverage Precision Grammars. In *Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th Conference on Computational Linguistics*, pages 8 – 14, Taipei, Taiwan.

Copestake, Ann. 2002. *Implementing Typed Feature Structure Grammars*. Stanford, CA: CSLI Publications.

Copestake, Ann, Flickinger, Dan, Pollard, Carl and Sag, Ivan. 2005. Minimal Recursion Semantics: An Introduction. *Research on Language and Computation* 3(4), 281–332.

Drellishak, Scott and Bender, Emily. 2005. A Coordination Module for a Crosslinguistic Grammar Resource. In Stefan Müller (ed.), *Proceedings of the HPSG05 Conference*, Stanford: CSLI Publication.

Erguvanlı-Taylan, Eser. 1999. Review: Turkish by Jaklin Kornfilt. *Anthropological Linguistics* 41(2), 253 – 258.

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

Kornfilt, Jaklin. 1997. *Turkish*. New York: Routledge.

Lees, Robert B. 1962. A compact analysis for the Turkish personal morphemes. In Nicholas Poppe (ed.), *American Studies in Altaic Linguistics*, volume 13 of *Indiana University Uralic and Altaic Series*, The Hague, the Netherlands: Mouton.

Lewis, Geoffrey. 2000. *Turkish Grammar*. New York: Oxford University Press, second edition.

O'Hara, Kelly. 2008. *A Morphosyntactic Infrastructure for a Grammar Customization System*. Masters Thesis, University of Washington.

Sezer, Engin. 2001. Finite inflection in Turkish. In Eser Erguvanlı Taylan (ed.), *The Verb in Turkish*, Chapter 1, pages 1 – 46, Philadelphia: John Benjamins.

Tikkanen, Bertil. 2001. Converbs. In Martin Haspelmath and Wulf Oesterreicher (eds.), *Language Typology and Language Universals: An International Handbook*, volume 2, Chapter 83, Berlin, Germany: W. de Gruyter.

Tseng, Jesse. 2003. Phrasal Affixes and French Morphosyntax. In G. Jäger, P. Monachesi, G. Penn and S. Wintner (eds.), *Proceedings of Formal Grammar 2003*, pages 177–188, Stanford: CSLI Publications.

Negative Concord in Romanian as Polyadic Quantification

Gianina Iordăchioaia
Institut für Linguistik: Anglistik
Universität Stuttgart
gianina@ifla.uni-stuttgart.de

Frank Richter
IMS, Universität Stuttgart &
SfS, Universität Tübingen
fr@sfs.uni-tuebingen.de

1 Introduction

We present an analysis of the syntax and semantics of the core of Romanian Negative Concord (NC) constructions as polyadic quantification in Lexical Resource Semantics (LRS). Following a proposal by de Swart and Sag (2002) for French, we express the truth conditions associated with Romanian NC constructions by means of negative polyadic quantifiers. Going beyond de Swart and Sag's largely informal treatment of the logical representations for polyadic quantification in HPSG, we extend the logical representation language of Lexical Resource Semantics (LRS, Richter and Sailer (2004)) and modify the interface principles of LRS to accommodate polyadic quantification. Apart from the immediate benefit of a theory of Romanian NC, we obtain an interesting result for constraint-based approaches to model-theoretic semantics like LRS: Resumptive polyadic quantifiers, which are at the heart of this approach to NC, are a notorious problem for frameworks which use the lambda calculus in combination with a functional theory of types to define a compositional semantics for natural languages. LRS overcomes these fundamental logical limitations and is powerful enough to specify by standard HPSG devices a precise systematic relationship between a surface-oriented syntax and semantic representations with polyadic quantifiers.

2 Data

Sentential negation in Romanian is usually expressed by a verbal prefix, *nu* (Barbu (2004)). In the absence of other negative elements, it contributes semantic negation (1a). If in addition an n-word is present such as *niciun* in *niciun student (no student)* in (1b), only a NC reading is available, a double negation interpretation (DN) is not. The negation marker (NM) *nu* is obligatory. In constructions with two n-words, both a NC reading and a DN reading are available (1c).

(1) a. Un student **nu** a venit.
a student NM has come
'Some student didn't come.'

b. **Niciun** student *(**nu**) a venit.
no student NM has come
i. 'No student came.' (NC)
ii. # 'No student didn't come.' (DN)

c. **Niciun** student nu a citit **nicio** carte.
no student NM has read no book
i. 'No student read any book.' (NC)
ii. 'No student read no book. (Every student read some book.)' (DN)

The observations in (1b) and (1c) suggest that (a) n-words are exponents of semantic negation (as can be confirmed by other tests), and (b) the negative marker *nu* is semantically non-negative in the presence of n-words. This is confirmed by the test in (2) and (3): Negative functions are anti-additive: f is anti-additive iff for each X and Y, $f(X \vee Y) = f(X) \wedge f(Y)$. In the absence of n-constituents, *nu* receives an anti-additive interpretation (2):

(2) a. Studenţii **nu** au citit romane *sau* poezii.
students-the NM have read novels or poems
'The students haven't read novels or poems.'

b. = Studenţii **nu** au citit romane *şi* studenţii **nu** au citit poezii.
students-the NM have read novels and students-the NM have read poems
= 'The students haven't read novels and the students haven't read poems.'

If the disjunction that *nu* takes as argument contains n-constituents, anti-additivity disappears, and the two n-constituents are interpreted independently under the scope of negation (3):

(3) a. Studenţii **nu** au citit **niciun** roman *sau* **nicio** poezie.
students-the NM have read no novel or no poem
'The students read no novel or no poem.'

b. ≠ Studenţii **nu** au citit **niciun** roman *şi* studenţii **nu** au citit
students-the NM have read no novel and students-the NM have read
nicio poezie.
no poem
≠ 'The students read no novel and the students read no poem.'

These data are consistent with an analysis that assumes that determiner n-words and negative NP constituents are quantifiers of Lindström type $\langle 1,1\rangle$ and $\langle 1\rangle$ (Lindström (1966)), respectively, and they may combine to form a polyadic quantifier (of type $\langle 1^n, n\rangle$ and $\langle n\rangle$) by resumption (Keenan and Westerståhl (1997)). The negative marker *nu* is analyzed as a negative quantifier of type $\langle 0\rangle$ that is absorbed under resumption with other negative polyadic quantifiers. The relevant technical details will be briefly outlined in our LRS implementation of polyadic quantification and resumption below.

3 LRS with Polyadic Quantifiers

For our analysis we will need a higher-order logical language with negative polyadic quantifiers. Here we briefly outline its crucial properties and indicate how to integrate it with LRS.

We assume a simple type theory with types e and t. Functional types are formed in the usual way. The syntax of the logical language provides function application, lambda abstraction, equality and negative polyadic quantifiers. By standard results this is enough to express the usual logical connectives and monadic quantifiers. In reference to the simple type theory, we call our family of languages Ty1. Var and $Const$ are a countably infinite supply of variables and constants of each type:

Definition 1 Ty1 Terms*: Ty1 is the smallest set such that:*
$Var \subset Ty1$, $Const \subset Ty1$,
for each $\tau, \tau' \in Type$, *for each* $\alpha_{\tau\tau'}, \beta_\tau \in Ty1$: $(\alpha_{\tau\tau'}\beta_\tau)_{\tau'} \in Ty1$,
for each $\tau, \tau' \in Type$, *for each* $v_{i,\tau} \in Var$, *for each* $\alpha_{\tau'} \in Ty1$: $(\lambda v_{i,\tau}.\alpha_{\tau'})_{(\tau\tau')} \in Ty1$,
for each $\tau \in Type$, *and for each* $\alpha_\tau, \beta_\tau \in Ty1$: $(\alpha_\tau = \beta_\tau)_t \in Ty1$,
for each $\tau \in Type$, *for each* $n \in \mathbb{N}^0$, *for each* $i_1, i_2, ..., i_n \in \mathbb{N}^+$, *for each* $v_{i_1,\tau}, v_{i_2,\tau}, ..., v_{i_n,\tau} \in Var$,
for each $\alpha_{t1}, \alpha_{t2}, ..., \alpha_{tn}, \beta_t \in Ty1$: $(NO(v_{i_1,\tau}, ..., v_{i_n,\tau})(\alpha_{t1}, ...\alpha_{tn})(\beta_t))_t \in Ty1$.

The standard constructs receive their usual interpretation. Here we only state the interpretation of the polyadic quantifiers:

Definition 2 The Semantics of Ty1 Terms *(clause for negative polyadic quantifiers only)*
For each term $\alpha_\tau \in Ty1$*, for each model* M *and for each variable assignment* $a \in Ass$*, for each* $\tau \in Type$*, for each* $n \in \mathbb{N}^0$*, for each* $i_1, i_2, ..., i_n \in \mathbb{N}^+$*, for each* $v_{i_1,\tau}, v_{i_2,\tau}, ..., v_{i_n,\tau} \in Var$*, for each* $\alpha_{t1}, \alpha_{t2}, ..., \alpha_{tn}, \beta_t \in Ty1$*:*

$[\![NO(v_{i_1,\tau}, ..., v_{i_n,\tau})(\alpha_{t1}, ..., \alpha_{tn})(\beta_t)]\!]^{M,a} = 1$ *iff*
for every $d_{i_1}, d_{i_2}, ..., d_{i_n} \in D_{E,\tau}$,

$[\![\alpha_{t1}]\!]^{M,a[v_{i_1,\tau}/d_{i_1}]} = 0$ *or* $[\![\alpha_{t2}]\!]^{M,a[v_{i_2,\tau}/d_{i_2}]} = 0$ *or* ...

or $[\![\alpha_{tn}]\!]^{M,a[v_{i_n,\tau}/d_{i_n}]} = 0$ *or* $[\![\beta_t]\!]^{M,a[(v_{i_1},...,v_{i_n})/(d_{i_1},...,d_{i_n})]} = 0$.

(4) shows the truth conditions that we obtain for the translation of the Romanian counterparts of *John didn't come* (4a) and *No teacher didn't give no book to no student*, where all NPs are n-constituents and form a ternary negative quantifier by resumption (4b):

(4) a. For $n = 0$, $[\![NO()()(come'(j))]\!]^{M,a} = 1$ iff $[\![come'(j)]\!]^{M,a} = 0$

b. For $n = 3$, $v_{i_1} = x, v_{i_2} = y, v_{i_3} = z$, $\alpha_{t1} = teacher'(x)$, $\alpha_{t2} = book'(y)$, $\alpha_{t3} = student'(z)$ and $\beta_t = give'(x, y, z)$,
$[\![NO(x, y, z)(teacher'(x), book'(y), student'(z))(give'(x, y, z))]\!]^{M,a} = 1$ iff
for every $d_1, d_2, d_3 \in D_{E,e}$,
$[\![teacher'(x)]\!]^{M,a[x/d_1]} = 0$ or $[\![book'(y)]\!]^{M,a[y/d_2]} = 0$ or
$[\![student'(z)]\!]^{M,a[z/d_3]} = 0$ or $[\![give'(x, y, z)]\!]^{M,a[(x,y,z)/(d_1,d_2,d_3)]} = 0$

Minor adjustments suffice to integrate these logical representations in LRS. In the signature, the appropriateness of *gen-quantifier* of Richter and Kallmeyer (2007) is generalized to lists of variables (instead of single variables), and the restrictor of quantifiers now contains a list of expressions:

```
me  TYPE  type
    gen-quantifier  VAR  list
                    RESTR  list
                    SCOPE  me
```

The theory of well-formed logical expressions restricts polyadic generalized quantifiers to the form given in DEFINITION 1. The relational restrictions in (5) guarantee that 1 is a list of variables, they all have the same type 3, the expressions in the restrictor list 2 are of type t, and there are exactly as many restrictor expressions as variables on the two lists:

(5) $$\textit{gen-quantifier} \rightarrow \begin{bmatrix} \text{TYPE } \textit{truth} \\ \text{VAR } \boxed{1} \\ \text{RESTR } \boxed{2} \\ \text{SCOPE}\,|\,\text{TYPE } \textit{truth} \end{bmatrix}$$
$$\wedge \texttt{variable-list}(\boxed{1}) \ \wedge \texttt{same-type-list}(\boxed{3},\boxed{1})$$
$$\wedge \texttt{truth-list}(\boxed{2}) \ \wedge \texttt{same-length}(\boxed{1}, \boxed{2})$$

The LRS PROJECTION PRINCIPLE (EXCONT and INCONT percolation, inheritance of PARTS lists) remains unchanged. The clause of the SEMANTICS PRINCIPLE governing the combination of quantificational determiners with nominal heads is adjusted to polyadic quantifiers:

(6) THE SEMANTICS PRINCIPLE, clause 1
1. if the non-head is a quantifier, then its INCONT value is of the form $Q(v, \phi, \psi)$, the INCONT value of the head is a component of a member[1] of the list ϕ, and the INCONT value

[1] The symbol "$\lhd_\in$" is the infix notation of the new relation `subterm-of-member`, a generalized subterm relation. Note that v and ϕ are shorthand for a list of variables and a list of expressions in $Q(v, \phi, \psi)$. ψ is a single expression.

of the non-head daughter is identical to the EXCONT value of the head daughter:

$$\left[\text{DTRS}|\ \text{SPR-DTR}|\text{SS}|\ \text{LOC}\begin{bmatrix}\text{CAT}|\ \text{HEAD} & \textit{det}\\ \text{CONT}\ |\text{MAIN} & \textit{gen-quantifier}\end{bmatrix}\right] \rightarrow$$
$$\left(\left[\text{DTRS}\begin{bmatrix}\text{H-DTR}\ |\text{LF} & \begin{bmatrix}\text{EXCONT} & \boxed{1}\\ \text{INCONT} & \boxed{2}\end{bmatrix}\\ \text{SPR-DTR}|\text{LF} & \begin{bmatrix}\text{INCONT} & \boxed{1}\begin{bmatrix}\textit{gen-quantifier}\\ \text{RESTR}\ \boxed{3}\end{bmatrix}\end{bmatrix}\end{bmatrix}\right] \wedge \boxed{2} \lhd_{\in} \boxed{3}\right)$$

Resumption will be implemented in LRS as identity of quantifiers contributed by lexical elements. Thus no special technical apparatus for the resumption operation has to be introduced in preparation of our analysis of negative concord in Romanian in the next section.

4 The Analysis of Romanian NC

The analysis of simple negated sentences without n-constituents follows immediately from the lexical analysis of verbs with the negative marker prefix *nu*, which we derive by lexical rule (not shown here, but see Przepiórkowski and Kupść (1997) for a comparable analysis of the Polish negative marker, and Ionescu (1999) for similar assumptions about Romanian). For the verb in (1a) we get:

(7) *nu a venit* ('NM has come', derived by Lexical Rule)

$$\begin{bmatrix}\textit{word}\\ \text{PHON}\ \langle \texttt{nu, a, venit}\rangle\\ \text{SS}\mid\text{LOC}\begin{bmatrix}\text{CAT}\ \begin{bmatrix}\text{HEAD}\ |\text{NEG}\ +\\ \text{VAL}\ |\text{SUBJ}\ \langle \text{NP}\boxed{1a}\rangle\end{bmatrix}\\ \text{CONT}\ \begin{bmatrix}\text{INDEX}\ |\text{VAR}\ \textit{no-var}\\ \text{MAIN}\ \boxed{3a}\ \textit{come}'\end{bmatrix}\end{bmatrix}\\ \text{LF}\ \begin{bmatrix}\text{EXC}\ \boxed{0}\\ \text{INC}\ \boxed{3}\ \textit{come}'(\boxed{1a})\\ \text{PARTS}\ \langle \boxed{3}, \boxed{3a}, \boxed{7}\ \textit{no}(u,\gamma,\delta)\rangle\end{bmatrix}\end{bmatrix} \ \&\ \boxed{3}\lhd\boxed{0}\ \ \&\ \boxed{3}\lhd\delta\ \ \&\ \boxed{7}\lhd\boxed{0}$$

With standard LRS mechanisms in combination with a language-specific constraint that excludes the existential quantifier originating from *un student* from occurring in the immediate scope of negation, we obtain $\textit{some}(x, \textit{student}'(x), \textit{no}((),(),\textit{come}'(x)))$ as the truth condition for (1a).

For the analysis of NC (1b), we adapt the NEG CRITERION of Richter and Sailer (2004) to Romanian and the polyadic quantifier approach:

(8) THE NEG CRITERION for Romanian
For every finite verb, if there is a type $\langle 0\rangle$ *no* quantifier in the external content of the verb that has scope over the verb's MAIN value, then any other negative quantifier in the verb's external content that also has scope over the verb's MAIN value must be on the verb's PARTS list.

$$\forall\boxed{0}\ \forall\boxed{1}\ \forall\boxed{2}\ \forall\boxed{3}$$
$$\left(\begin{bmatrix}\textit{word}\\ \text{SS}\ |\text{LOC}\begin{bmatrix}\text{CAT}\ |\text{HEAD}\ \begin{bmatrix}\textit{verb}\\ \text{VFORM}\ \textit{fin}\end{bmatrix}\\ \text{CONT}\ |\text{MAIN}\ \boxed{3}\end{bmatrix}\\ \text{LF}\ |\text{EXC}\ \boxed{0}\end{bmatrix} \wedge \boxed{1}\ \textit{no}((),(),\delta)\lhd\boxed{0} \wedge \boxed{2}\ \textit{no}(v,\alpha,\beta)\lhd\boxed{0} \wedge \boxed{3}\lhd\delta \wedge \boxed{3}\lhd\beta \right.$$
$$\left. \rightarrow \exists\boxed{4}\left(\left[\text{LF}\ |\text{PARTS}\ \ \boxed{4}\right] \wedge \boxed{2}\in\boxed{4}\right)\right)$$

The obligatoriness of the negative marker in negative concord constructions is an immediate consequence of the NEGATIVE CONCORD CONSTRAINT of Romanian (9). If a sentential negation (2) outscopes a verb (within the verb's EXCONT), the verb must be negatively marked, which in turn can only be the case if it is licensed as output of the negation lexical rule.

(9) THE NC CONSTRAINT (NCC)

$$\forall \boxed{0}\ \forall \boxed{1}\ \forall \boxed{2}$$
$$\left(\begin{bmatrix} word \\ \text{SS}\,|\text{LOC} \begin{bmatrix} \text{CAT}\,|\text{HEAD} & \begin{bmatrix} verb \\ \text{VFORM} & fin \end{bmatrix} \\ \text{CONT}\,|\text{MAIN} & \boxed{1} \end{bmatrix} \\ \text{LF}\,|\text{EXCONT} \quad \boxed{0} \end{bmatrix} \wedge \boxed{2}\ no(v, \alpha, \beta) \lhd \boxed{0} \wedge \boxed{1} \lhd \beta \right) \rightarrow \left[\text{SS}\,|\text{LOC}\,|\text{CAT}\,|\text{HEAD}\ [\text{NEG}\ +]\right]$$

Assuming that n-words introduce on their PARTS list a negative quantifier of unspecified type $\langle 1^n, n\rangle$ with exactly one new variable, these two principles suffice to guarantee the correct analysis of (1b) and (1c), shown in (10) and (11), respectively.

(10) $no(x, student'(x), come'(x))$

(11) a. $no(x, student'(x), no(y, book'(y), read'(x, y)))$ (DN)

b. $no((x, y), (student'(x), book'(y)), read'(x, y))$ (NC)

In (1b), the verb and the n-word each contribute a negative polyadic quantifier. The verb does not contribute a variable for the quantifier, whereas the negative determiner does. If the two negative quantifiers were not identical, they would be subject to the NEG CRITERION, because the quantifier contributed by the verb would have an empty variable list, i.e. it would be of type $\langle 0\rangle$. But then the quantifier contributed by the n-word would have to be on the PARTS list of the verb. This cannot be the case, since the verb only contributes one negative quantifier. Therefore the quantifiers contributed by the n-word and by the verb must be identical, with one variable on the VAR list, resulting in (10). The NEG CRITERION has nothing to say about this case, because there is no type $\langle 0\rangle$ quantifier in the formula. Since an n-word always contributes a negative quantifier, the NCC guarantees that the verb must have the negation prefix and contribute a negative quantifier in the presence of an n-word.

In sentences with more than one n-word such as (1c), the negative quantifier contributed by the verb must undergo resumption with at least one of the two quantifiers contributed by the n-words for the reasons just described. If one n-word does not undergo resumption with the NM and the other n-word, we obtain the DN reading as in (11a). However, there is also the possibility that all the negative quantifier contributions in the sentence are identified. The number of variables contributed by the individual n-words determines the type of the resumptive quantifier. For (1c) with two n-words, each contributing one variable, the second available alternative is resumption of all three negative quantifiers and leads to a quantifier of type $\langle 1^2, 2\rangle$ for the NC reading, shown in (11b).

In the talk, we will also show how our analysis treats cases in which NC crosses clause boundaries of embedded subjunctive clauses. In these constructions, a negated matrix verb may license n-words in an embedded subjunctive clause. The matrix negation then enters into a negative polyadic quantifier with the embedded n-words. Our analysis of the syntax-semantics interface will provide an account of the conditions when this is possible.

5 Conclusion

The present analysis of NC in Romanian applies the approach that was pioneered by an analysis of French in de Swart and Sag (2002). Our theory considerably extends de Swart and Sag's proposal by explicitly integrating a higher-order logic with polyadic quantification in HPSG. We expect that

the formulation of the polyadic quantifier approach to NC in LRS will make it possible to unify this line of research with the typological approach to NC in Polish, French and German presented in Richter and Sailer (2006). Last but not least, adding polyadic quantification to LRS opens the door to exploring a whole range of new semantic phenomena in HPSG such as cumulative and *same/different* (unreducible) polyadic quantifiers (Keenan (1992), Keenan and Westerståhl (1997)). Since our constraint-based syntax-semantics interface supports the integration of polyadic quantifiers, HPSG theories can take full advantage of them. This brings within reach an explicit specification of the syntax and semantics of constructions that require unreducible polyadic quantifiers for an adequate rendering of their truth conditions and have, for that reason, turned out to be problematic in other grammar frameworks.

References

Barbu, Ana-Maria (2004). The negation NU: lexical or affixal item. In E. Ionescu (Ed.), *Understanding Romanian Negation. Syntactic and Semantic Approaches in a Declarative Perspective*, pp. 68–82. Bucharest U. Press.

Ionescu, Emil (1999). A Quantification-based Approach to Negative Concord in Romanian. In G. Kruijff and R. Oehrle (Eds.), *Proceedings of Formal Grammar 1999*, Utrecht, pp. 25–35.

Keenan, Edward L. (1992). Beyond the Frege Boundary. *Linguistics and Philosophy 15*, 199–221.

Keenan, Edward L. and Westerståhl, Dag (1997). Generalized Quantifiers in Linguistics and Logic. In J. van Benthem and A. ter Meulen (Eds.), *Handbook of Language and Logic*, pp. 837–893. Amsterdam: Elsevier Science.

Lindström, Per (1966). First Order Predicate Logic with Generalized Quantifiers. *Theoria 32*, 186–195.

Przepiórkowski, Adam and Kupść, Anna (1997). Negative Concord in Polish. Technical report, Institute of Computer Science, Polish Academy of Sciences.

Richter, Frank and Kallmeyer, Laura (2007). Feature logic-based semantic composition: A comparison between LRS and LTAG. In A. Søgaard and P. Haugereid (Eds.), *Postproceedings of The 1st International Workshop on Typed Feature Structure Grammars*, Volume 7 of *CST Working Papers, København, Denmark*, pp. 31–83.

Richter, Frank and Sailer, Manfred (2004). Basic Concepts of Lexical Resource Semantics. In *ESSLLI 2003 – Course Material I*, Volume 5 of *Collegium Logicum*. Kurt Gödel Society Wien.

Richter, Frank and Sailer, Manfred (2006). Modeling Typological Markedness in Semantics: The Case of Negative Concord. In S. Müller (Ed.), *Proceedings of the 13th International Conference on Head-Driven Phrase Structure Grammar*, pp. 305–325. CSLI Publications.

de Swart, Henriëtte and Sag, Ivan A. (2002). Negation and Negative Concord in Romance. *Linguistics and Philosophy 25*, 373–417.

On Predication

Stefan Müller
Deutsche Grammatik
Institut für Deutsche und Niederländische Philologie
Fachbereich Philosophie und Geisteswissenschaften
FU Berlin

Stefan.Mueller@fu-berlin.de

1 Introduction

Van Eynde (2008) suggested a revision of the treatment of predication that can be found in (Pollard and Sag, 1994, p. 360) and (Ginzburg and Sag, 2000, p. 409). This paper points out several shortcomings of van Eynde's analysis and suggests returning to a raising analysis. We will present the problems that van Eynde identified with Pollard and Sag's account and the solution he suggested in Section 2, and explain which of the problems are not real problems. Section 3 shows where the shortcomings of van Eynde's proposal are and Section 4 provides our own analysis. We draw some conclusions in Section 5.

2 Problems and alleged Problems and Proposed Solutions

Pollard and Sag (1994, p. 360) sketch the lexical rule in (1) that takes nouns as used in normal referential NPs like *a teacher* in (2a) and maps them onto another lexical item that can be used predicatively like in (2b).

(1) $\text{N}[-\text{PRD}, \text{SUBJ}\ \langle\rangle]:[\text{RESTRICTION}\ \{\boxed{2}\}]_{\boxed{1}}$
$\mapsto$
$\text{N}[+\text{PRD}, \text{SUBJ}\ \left\langle\ \text{XP}_{\boxed{1}}\ \right\rangle]:\boxed{2}$

(2) a. A teacher laughs.
b. John is a teacher.

Ginzburg and Sag (2000, p. 409) give the following variant of the lexical rule in (1):

(3) Singular Predicative Noun Lexical Rule:

$$\begin{bmatrix} \text{SS|LOC|CAT|HEAD}\ n \\ \text{ARG-ST}\ \langle\ \boxed{1}\ \rangle \oplus \boxed{A} \\ lx \end{bmatrix} \Longrightarrow_{LR}$$

$$\begin{bmatrix} \text{SS|LOC|CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{AGR|NUM}\ sg \\ \text{PRED}\ + \end{bmatrix} \\ \text{SPR} & \langle\ \boxed{1}\ \rangle \\ \text{SUBJ} & \langle\ \boxed{2}\ \rangle \end{bmatrix} \\ \text{ARG-ST}\ \langle\ \boxed{2}, \boxed{1}\ \rangle \oplus \boxed{A} \\ word \end{bmatrix}$$

The lexical rule in (1) adds a subject to the valence features of the noun and by doing so makes it parallel to predicative adjectives. The copula and verbs like *seem* and *consider* are treated as raising verbs that raise the element in SUBJ and make it their own subject or – in the case of *consider* – object. This analysis is also assumed by other researchers working on different languages (see for instance (Müller, 2002)).

Pollard and Sag suggest that the element in the set of restrictions of the noun in the input of the rule is represented as the main semantic contribution of the resulting noun. So the contribution of *teacher* in (2b) is *teacher'*(1), while it is $\boxed{1}\{$*teacher'*$(\boxed{1})\}$ for (2a).[1] As Pollard and Sag point out, this does not extend to proper nouns like those in (4) for semantic reasons:

(4) Cicero is Tully.

Van Eynde mentions such examples and adds one with a pronoun:

(5) That must be her.

But Pollard and Sag (1987, p. 66) point to the solution already: There is a distinction to be made between the *be* of predication and the *be* of identity.

Van Eynde discusses examples like (6) and argues that nominalizations pose problems for the raising approach since the subject of *choose* is *she* and the object is *this hotel* and in predicative structures like (6) one would be forced to assume that *her* is a specifier and *this hotel* is the subject.

(6) This hotel was her first choice.

We think that this is a very weak argument since we have other processes that change the realization of grammatical functions. A prominent example is the passive of transitive verbs, in which an object is realized as the subject.

The next phenomenon that van Eynde discusses is gerunds:

(7) a. The greatest pleasure on earth is eating oysters and drinking champagne.
b. His main worry now is to get rid of his detractors.

[1] The curly brackets around 2 in the input are missing in Pollard and Sag's version of the lexical rule.

It is clear that the subject of *eating* is not *the greatest pleasure on earth*. Neither is *his main worry* the subject of the infinitive. Rather these are equational structures in which the NP is equated with another element the subject of which remains implicit. This is unproblematic if one assumes an equational copula. The translation of (7a) into German shows that implicit subjects are nothing unusual:

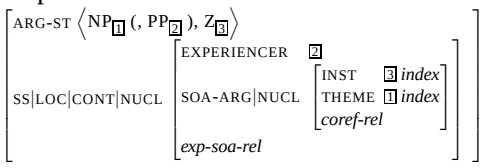
(8) Das größte Vergnügen auf der Welt ist das Essen von Austern und das Trinken von Champagner.

In his analysis of sentences like (7) van Eynde suggests that *eating* is nominal and *to* is a preposition with nominal semantics. Of course this analysis can be employed in an analysis with an identity copula as well.

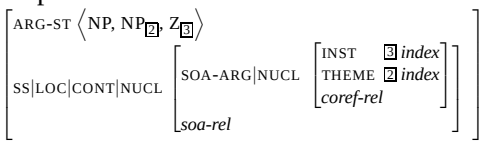
Van Eynde suggests the following alternative to the raising analysis: Lexical items for *seems* as in (9a) are constrained by (10) and items like *consider* in (9b) are constrained by (11).

(9) a. John seems a nice guy.
b. Bob considers his brother a genius.

(10) a1-pred-lex ⇒



(11) a2-pred-lex ⇒



By assuming these lexical entries van Eynde can analyze the sentences in (9) with normal nouns without having to assume a separate predicative lexical item for the predicative usage of the noun.

Van Eynde assumes that all predicate selectors contribute such semantic information and explicitely includes the copula *be* here. He argues that the dative iudicantis depends on the copula which he takes as evidence for its relational status:

(12) Es ist mir zu kalt.
it is me.DAT too cold
'It is too cold for me.'

Traditionally it is said that this dative depends on the *zu* (How this is captured in HPSG is a different question. The analysis is not trivial since dative and *zu* can be discontinuous). Note however, that van Eynde would be forced to assume empty copulas in prenominal position if he were to apply his argument to the following data:

(13) a. bis auf das mir zu kalte Ziel
until on the me.DAT too cold goal
Spitzbergen
Spitsbergen
'except for the goal Spitsbergen, which is too cold for me'

b. Kann man im Adapterbetrieb
can one in.the battery.mode
die mir zu warme
the me.Dat to warm
Book-Unterseite etwas kühler
bottom.of.the.Book a.bit cooler
bekommen, wenn man [...]
get if one
'Is it possible to get the bottom of the Book, which is to warm for me, a bit cooler if one ...'

Here we have *mir zu warme* and *mir zu kalte*, with *zu* present but in a prenominal context in which copulas are never present.

3 Problems of the Identity Approach

The analysis works for the given examples, but the argumentation against the raising analysis is not convincing. In addition, the identity analysis faces several problems.

The first problem is that pronouns and proper names cannot be used as predicates in such constructions:

(14) a. * He seems him.
b. * He seems John Malcovich.

Here the copula has to be used:

(15) a. He seems to be him.
b. He seems to be John Malcovich.

In addition there is a very general problem of the analysis: It does not extend to subjectless predicates. Müller (2002, p. 72–73) discusses the following examples:[2]

(16) a. weil schulfrei ist.
because school.free is
'because there is no school.'

b. Für dich ist immer offen.
for you is always open
'It is always open for you.'

c. In der Mensa ist es laut.
in the commons is it.EXPL loud
'It is loud in the commons.'

The adjective *laut* also has a non-expletive version, and (16c) is actually ambiguous between the expletive and the non-expletive reading. With the expletive predicate, (16c) means that

[2] (16b) is quoted from (Haider, 1986, p. 18).

the people, machines, or whatever, in the commons are loud, whereas in the non-expletive reading the *es* could refer to a child. So the examples in (16) either do not involve an NP argument or an expletive one. In any case there is nothing present that could be "coreferential" with the adjectival predicate.

So concluding this discussion of van Eynde's paper one must say that there is no compelling argument against the raising analysis and there are two areas in which the identity analysis has empirical problems.

4 The Analysis

As van Eynde pointed out in earlier work, there is one aspect in which his analysis is clearly superior to the one suggested in (Pollard and Sag, 1994, p. 194–195) and (Ginzburg and Sag, 2000): The latter analyses fail for examples that contain modifiers in the predicative phrase.

(17) He is a good candidate.

The classical analysis of adjuncts assumes that nominal modifiers attach to an $\overline{\text{N}}$ and identify their referential index with the referential index of the noun. But if the semantic contribution of *candidate* is a predicate rather than an index, modification cannot apply as usual.[3]

This problem can be solved by assuming a unary projection instead of the lexical rule in (3). The unary projection applies to a full NP and licenses the predicative NP with an appropriate SUBJ value. Note that in this analysis there is still ambiguity between NPs that can function as complements and NP that can function as predicates – something that van Eynde critizized – but the ambiguity is reduced considerably since it is only present at the NP level and not for all nominal projections. So there is no predicative version of *good candidate*.

It is interesting to note that we find a similar phenomenon in temporal NPs: As Flickinger (2008, p. 91–92) points out, it is not just simple NPs that can act as modifiers of verbs. The time nouns can be embedded inside of a more complex NP, as (18) shows.

(18) a. Kim disappears those days.
b. Kim disappears some of those days.

Therefore a treatment in which the time noun has a MOD value that allows it to modify a verb is not appropriate. Further evidence for an analysis as unary projection is provided by parallel German examples:

(19) a. Er arbeitete den größten Teil der Nacht.
he worked the.ACC largest part of.the.GEN night
'He worked almost all night.'

b. Er arbeitete die halbe Nacht.
he worked the.ACC half.ACC night
'He worked half of the night.'

In (19a) the time expression *der Nacht* is genitive but the whole NP is accusative. This accusative is called a semantic case. It is connected to the function of the NP and not assigned by the verb. It is clear from data like (19a) that an analysis like the one suggested by Müller (2007, p. 226) that assigns both function (i.e. MOD value) and case lexically cannot explain the data in (19a). Hence we have evidence that a unary projection that accounts for the property of predication is nothing unusual.

We assume the following entry for the copula in German:

(20) *sein* (copula):

$$\left[\text{SUBCAT } \boxed{1} \oplus \boxed{2} \oplus \left\langle \begin{bmatrix} \text{PRD} + \\ \text{SUBJ } \boxed{1} \\ \text{SUBCAT } \boxed{2} \end{bmatrix} \right\rangle\right]$$

This lexical entry is similar to the one suggested by Müller (2002, p. 103) in that both the elements of SUBJ and of SUBCAT are raised to the SUBCAT list of the copula. The elements at the SUBCAT list of the embedded predicate are raised in addition to the elements in SUBJ since German forms a verbal complex and predicative constructions like copula constructions and resultative constructions take part in complex formation. The formation of verbal complexes is analyzed via argument attraction (Hinrichs and Nakazawa, 1989a,b, 1994; Kiss, 1995) and Müller (2002) extended this analysis to predicative constructions.

Note that nothing is said about the actual members of the lists. It is therefor possible to handle the cases in (21) as well as the subjectless examples that were given in (16).

(21) a. weil er auf seinen Sohn stolz ist
because he.NOM on his son proud is
'because he is proud of his son'

b. weil er klug ist
because he.NOM smart is
'because he is smart'

[3] This may not be an issue if an MRS semantics (Copestake et al., 2005) is assumed. However, one would have to be willing to claim that the type of the index of *candidate* is not changed by the predication lexical rule.

c. Ihm wurde schlecht.
him.DAT got bad
'He got sick.'

In the analysis of (21a), 1 contains the subject (*er*) and 2 the PP (*auf seinen Sohn*). In the analysis of (21b), 1 contains the subject (*er*) and 2 is the empty list. In the analysis of (21c), 1 is the empty list and 2 contains the dative object *ihm*. In the analysis of (16a), both 1 and 2 are the empty list.

It is further interesting to note that the treatment of raising in (20) differs from the characterization of raising as it is given in Ginzburg and Sag (2000, p. 22). Ginzburg and Sag assume the following constraint:

(22) [ARG-ST ⟨ [LOC 1], [SUBJ ⟨ [LOC 1]⟩] ⟩]

This version of raising differs from earlier proposals in that only LOCAL values are shared instead of whole *synsem* objects. The reason for this treatment is that one would get problems with the lexical SLASH amalgamation that was suggested by Bouma et al. (2001): if the whole *synsem* object was shared there would be SLASH amalgamation in the subject and in the phrase from which the subject is raised, an unwelcome result (Ginzburg and Sag, 2000, p. 21, fn. 8). So if one would assume an amalgamation account of nonlocal dependencies for German, one would be forced to either use disjunctions that refer to the arity of the respective lists or use a relational constraint that walks through lists and produces a copy of the list that contains elements that share the LOCAL values with the elements of the list from which they are raised. Given that the amalgamation analysis is not uncontroversial (Levine and Hukari, 2006), we suggest to drop it and return to an analysis that introduces nonlocal dependencies in syntax (through a trace or a unary branching projection).

There is another important aspect regarding the lexical item in (20): The predicate is selected via SUBCAT rather then VCOMP or XCOMP as was suggested by Chung (1993), Rentier (1994), Müller (1997), and Kathol (1998). Müller (2002, p. 103) gave the following lexical item for the copula:

(23) *sein* (copula, according to Müller (2002)):

$$\begin{bmatrix} \text{SUBCAT} & \boxed{1} \oplus \boxed{2} \\ \text{XCOMP} & \left\langle \begin{array}{l} \text{ADJ}[\text{MOD } \textit{none}, \text{PRD } +, \text{SUBJ } \boxed{1}, \text{SUBCAT } \boxed{2}, \\ \quad \text{XCOMP } \langle\rangle, \text{LEX } +] \end{array} \right\rangle \end{bmatrix}$$

The problem with this lexical item is that it specifically selects a predicative adjective. Müller selected all verbs that take part in complex formation via XCOMP, but those that were realized as full phrases – that is in so-called incoherent constructions – were selected via SUBCAT. The motivation for this is that full phrases can be permuted like other arguments in German, while lexical elements in general have to be adjacent to the verbal complex. With a uniform selection of verbal complements via SUBCAT it is possible to treat optionally coherent verbs like *versuchen* with one lexical item (Kiss, 1995, p. 178). The control verb does not specify whether it forms a verbal complex with the embedded verb or not. It does not mention the LEX value auf the embedded verbal element. Because of this we can analyze both examples with a predicate complex as in (24a) and examples like (24b) with so-called intraposition:

(24) a. Karl hat das Buch nicht [zu lesen
Karl has the book not to read
versucht].
tried
'Karl did not try to read the book.'

b. Karl hat [das Buch zu lesen] nicht
Karl has the book to read not
versucht.
tried
'Karl did not try to read the book.'

In comparison verbs like *scheinen* ('to seem') or modals, that obligatorily construct coherently select a verbal complement that is LEX+. Consequently they do not allow for intraposition of a VP complement, but require complex formation.

Müller (2002, p. 112) criticized Kiss's analysis of optional coherence because it also licences unwanted structures like (25) and hence results in spurious ambiguities.

(25) weil Karl das Buch [[dem Mann zu
because Karl the book the man to
geben] verspricht].
give promises
'because Karl promises to give the book to the man.'

In (25) *versprechen* is combined with a partly saturated verbal projection *dem Mann zu geben* and the non-saturated argument *das Buch* is raised and combined with *dem Mann zu geben verspricht* in a later step. However, this structure is excluded if arguments are required to be saturated and elements of the predicate complex are required to be LEX +. With the new treatment predicate selection via SUBCAT, it is not required that predicative PPs or NPs are part of the predicate complex as was suggested by Müller (2002) for PPs in resultative constructions. Instead they can be analyzed via head-argument structures.

Returning to the copula, it allows the embedding of fully saturated phrases like predicative NPs and PPs but also allows for the formation of a predicate complex consisting of adjective and copula. Since coherence is optional we can

explain so-called focus movement of adjectives as in (26), something that was noted by Müller (2002, p. 69) but not treated in his analysis.

(26) a. Sie wuchsen in einem
they grew in a
gesellschaftlichen Klima auf,
social climate PART(up)
das *freier* in Deutschland nie *war*.[4]
that freer in Germany never was

'They grew up in a social climate that was freer than ever in Germany.'

b. Dabei könnte die Begründung des
that.at could the reason for.the
Urteils *absurder* nicht *sein*: [...][5]
verdict more.absurd not be

'Yet the reason for the verdict could not be more absurd.'

5 Conclusion

We have shown that the arguments provided by van Eynde for a copula-based analysis are not convincing. In addition there are problems with pronouns in predication structures and the analysis does not extend to subjectless constructions.

We suggested returning to a raising analysis of predication that raises the complete value of SUBJ of the embedded predicate rather than identifying LOCAL values of raised subjects. The predication lexical rule was recoded as a unary branching immediate dominance schema, which allows the inclusion of modifiers in the NP. In addition it was suggested to dispense with the XCOMP feature and to return to a SUBCAT-based analysis. This makes it possible to treat the various predication structures as optionally coherent constructions.

The analysis is part of an implemented fragment of German.

References

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

Chung, C. 1993. Korean Auxiliary Verb Constructions Without VP Nodes. In S. Kuno, I.-H. Lee, J. Whitman, J. Maling, Y.-S. Kang and Y. joo Kim (eds.), *Proceedings of the 1993 Workshop on Korean Linguistics*, pp 274–286, Cambridge, Massachusetts: Harvard University, Department of Linguistics.

Copestake, A., Flickinger, D. P., Pollard, C. J. and Sag, I. A. 2005. Minimal Recursion Semantics: an Introduction. *Research on Language and Computation* 4(3), 281–332.

Flickinger, D. P. 2008. Transparent Heads. In Müller (2008).

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

Haider, H. 1986. Fehlende Argumente: vom Passiv zu kohärenten Infinitiven. *Linguistische Berichte* 101, 3–33.

Hinrichs, E. W. and Nakazawa, T. 1989a. Flipped out: AUX in German. In *Aspects of German VP Structure*, Eberhard-Karls-Universität Tübingen.

Hinrichs, E. W. and Nakazawa, T. 1989b. Subcategorization and VP Structure in German. In *Aspects of German VP Structure*, Eberhard-Karls-Universität Tübingen.

Hinrichs, E. W. and Nakazawa, T. 1994. Linearizing AUXs in German Verbal Complexes. In J. Nerbonne, K. Netter and C. J. Pollard (eds.), *German in Head-Driven Phrase Structure Grammar*, pp 11–38, Stanford: CSLI Publications.

Kathol, A. 1998. Constituency and Linearization of Verbal Complexes. In E. W. Hinrichs, A. Kathol and T. Nakazawa (eds.), *Complex Predicates in Nonderivational Syntax*, pp 221–270, San Diego: Academic Press.

Kiss, T. 1995. *Infinite Komplementation. Neue Studien zum deutschen Verbum infinitum*. Tübingen: Max Niemeyer Verlag.

Levine, R. D. and Hukari, T. E. 2006. *The Unity of Unbounded Dependency Constructions*. Stanford University: CSLI Publications.

Müller, S. 1997. Yet another Paper about Partial Verb Phrase Fronting in German. Research Report RR-97-07, DFKI, Saarbrücken, a shorter version appeared in *Proceedings of COLING 96*, pages 800–805.

Müller, S. 2002. *Complex Predicates: Verbal Complexes, Resultative Constructions, and Particle Verbs in German*. Stanford: CSLI Publications.

Müller, S. 2007. *Head-Driven Phrase Structure Grammar: Eine Einführung*. Tübingen: Stauffenburg Verlag, first edition.

Müller, S. (ed.). 2008. *Proceedings of the 15th International Conference on Head-Driven Phrase Structure Grammar, NICT, Keihanna, Japan*, Stanford, CSLI Publications.

Pollard, C. J. and Sag, I. A. 1987. *Information-Based Syntax and Semantics*. Stanford: CSLI Publications.

Pollard, C. J. and Sag, I. A. 1994. *Head-Driven Phrase Structure Grammar*. Chicago, Lon-

[4] taz, 01.07.1995, p. 10.
[5] taz, 17.02.1999, p. 12.

don: University of Chicago Press.
Rentier, G. 1994. Dutch Cross Serial Dependencies in HPSG. In COLING Staff (ed.), *Proceedings of COLING 94*, pp 818–822, Kyoto, Japan: Association for Computational Linguistics.
van Eynde, F. 2008. Predicate Complements. In Müller (2008), pp 253–273.

Preposed Sentential Negation in Danish

Submission for HPSG09

1 Introduction

In Danish embedded clauses sentential negation occupies a position to the left of the VP, i.e. after the subject immediately preceding the finite verb (1a). Even inherently negated indefinite complements are attracted to this position as in (1b) ([Sells, 2000]). However, in certain non-veridical contexts sentential negation may also appear between the complementizer and the subject (1c). I will refer to this as *preposed negation*. Note that the negation element *ikke*/'not' is morphologically and phonologically identical in the two uses.

(1) a. *fordi (*ikke) han (ikke) har nogen venner*
because (*not) he not has any friends

b. *fordi han (ingen venner) har (*ingen venner)*
because he (no friends) has (*no friends)

c. *hvis (ikke) han (ikke) har nogen venner*
if (not) he (not) has any friends

In the paper I propose that preposed negation is an optional complement of the non-veridical complementizer just as sentential negation in English is analyzed as a complement of the auxiliary in [Kim and Sag, 2002, Warner, 2000], while ordinary sentential negation in Danish is an adjunct. I further show that preposed negation is associated with VERUM-focus ([Höhle, 1992, Romero, 2006]), emphasizing the negation of an *all_focus* proposition in the antecedent of the conditional. Thus, in non-veridical contexts sentential negation may scope either over a proposition or a VERUM-predicate. The HPSG analysis is spelled out in the last section providing sample semantic representations using Minimal Recursion Semantics ([Copestake et al., 2005, Warner, 2000]) and providing lexical entries for the sentential negation *ikke*/'not' and the complementizer *hvis*/'if'.

2 Preposed negation

Preposed negation is found in non-veridical contexts, notably in conditional clauses but also in interrogatives (embedded (2a) and non-embedded (2b)) and optative clauses (2c).

(2) a. *jeg spekulerer på om ikke det er for sent*
I wonder PREP if not it is too late

b. *mon ikke det er for sent*
I.guess.COMPL not it is too late

c. *bare ikke han kommer*
if.only.COMPL not he comes

Preposing is, however, not restricted to negation. Sentential adverbs relating to the polarity of the clause, may be preposed, even more at a time (3c) forming a complex ADVP where the additional adverbs modify the negation.

(3) a. *hvis alligevel du vil deltage*
if anyway you want to participate

b. *hvis godt du vil deltage*
if indeed you want to participate

c. *hvis [altså alligevel ikke] du vil deltage*
if that.is anyway not you want to participate
'if you don't want to participate anyway, that is'

Not only negation but also inherently negated indefinite objects may be preposed subject to the same restriction as post-subject negated indefinites, namely that postnominal PPs must remain in situ ([Christensen, 2005]). Negated indefinites will not be further considered here.

(4) *hvis ingen arvinger (*til formuen) der er (til formuen)*
if no heirs (*to the.fortune) there are (to the.fortune)

A non-veridical context is not a sufficient condition for preposing although previous accounts actually explain this construction through the close bond between conditional semantics and negation ([Jensen, 2001]). The complementizer also has to be monosyllabic with the possible exception of *dersom*/'in case'. Otherwise bisyllabic conditional complementizers like *såfremt*/'provided that' and *ifald*/'in case' do not allow preposing.

(5) * *såfremt* / *ifald* *ikke du* *vil* *deltage*
provided that / in case not you want to participate

Also verbinitial conditional clauses do not allow preposing arguing against a purely semantic explanation of the phenomenon.

(6) * *har* *ikke du* *lyst,* *skal du ikke gøre det*
have not you the.desire, don't do it
'if you don't feel like it, don't do it"

The following discussion is limited to sentential negation in the context of *hvis*/'if'.

3 The Syntax of Preposed Negation

Preposed negation takes scope over the whole clause. It cancels out ordinary sentential negation (7) and just like ordinary sentential negation it licenses the negative polarity item (NPI) *heller*/'either' in a conjoined negated clause ([Borsley and Jones, 2005]) (8). Sentential negation with the (presuppositional) adverb *heller*/'either' requires a preceding sentence with sentential negation.

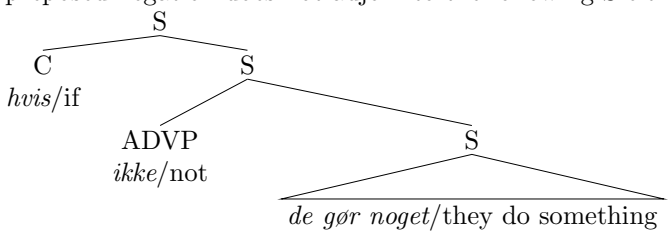
(7) *hvis ikke de* *ikke gør noget* → *hvis de gør noget*
if not they not do something → if they do something

(8) *hvis ikke det regner og* *det heller ikke sner*
if not it rains and it either not snows
'if it doesn't rain and it doesn't snow either'

Preposed negation is not part of a lexical structure, i.e. a kind of composite, inherently negated complementizer. The negation part is unambiguously a syntactic phrase as already shown in (4). If the negative quantifier phrase were to be analyzed differently from a preposed negation element, we would be left to explain why negative quantifier phrases always occur in exactly the same position as sentential negation (preposed or not). Also the negation part may contain the NPI-degree word *slet*/'at all' which cannot occur alone.

(9) *Jaa,* *hvis slet* *ikke du* *har* *nogen fornemmelse af* *hvordan ...*
Well, if at.all not you have any idea PREP how to ...

The preposed negation does not adjoin to the following S either, i.e. it does not exhibit the following structure.



First of all the restriction to monosyllabic complementizers is hard to explain on the basis of this structure. But more importantly sentential negation cannot attach to S but only to VP regardless of whether the complement clause is fronted or not. Cf.

(10) * *han havde fortalt at* *ikke han havde stjålet bilen*
he had told that not he had stolen the.car

Negation attaching to an S is only possible in the so-called *meta-negation* ([Christensen, 2005]) where negation attaches to a paranthetical complementizer clause expressing denial of an otherwise expected (conversational) implicature. In this use the sentence cannot be embedded and negation is to the left of the complementizer (in preposed negation it is to the right of the complementizer).

(11) *ikke at jeg er utaknemmelig, men ...*
not that I am ungrateful, but ...

Secondly the structure above falsely predicts that a conjoined clause may also occur with a preposed negation. Preposed negation, however, must be adjacent to the complementizer.

(12) ??/* *hvis ikke Peter kommer og ikke Louise er syg*
if not Peter comes and not Louise is ill

Finally the complementizer and the negation may be stranded in elliptical structures. This is unexpected if the negation adjoins to an S, since there is no S to adjoin to ([Kim and Sag, 2002]).

(13) *Hvis ikke, er det ikke ulovligt at have dem stående*
If not, is it not illegal to have them around

Thus preposed negation is syntactically tied to the complementizer as in the analysis of ([Johanessen, 2000]) who assumes that the negation cliticizes to C in Norwegian. Licensing by the complementizer also allows an account of the restriction to monosyllabic complementizers as a case of lexical selection either through COMPS or MOD. However, there is evidence that preposed negation does not modify the complementizer. Preposed negation is always interpreted within the scope of the complementizer: IF(¬(p)) while in a modificational structure the negation would take scope over the complementizer.

The complementizer is the syntactic and semantic head of the construction. It is still possible that the lexical complementizer clitizises to the phrasal preposed negation forming a kind of complex head, but this does not explain how a negation element in the preposed position is licensed.

Thus I propose that the preposed negation is a complement of the complementizer just as sentential negation in English is analyzed as a complement of the auxliary in [Kim and Sag, 2002, Warner, 2000]. This is further corroborated by the fact that preposing of the negation is lexically determined. Only certain monosyllabic complementizers with a non-veridical semantics license preposing. And the failure of finite verbs to license preposing (example (6)) can be explained as a lexical fact about finite verbs.

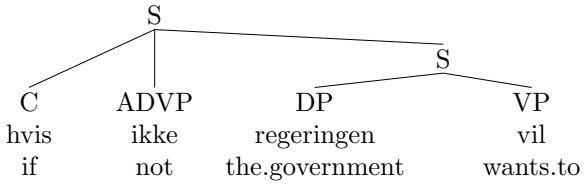
An analysis of ordinary sentential negation as a complement of the finite verb is, however, not viable. In Danish, complements appear to to right of their head, but sentential negation appears to the left of the head and may even be separated from its head by adjuncts. In addition, it would make the ungrammaticality of (6) totally mysterious.

(14) a. *fordi han ikke ser filmen*
because he not sees the.movie

b. *fordi de ikke som raske mennesker hurtigt er i stand til at slukke ilden*
because they not such as healthy people quickly are capable of extinguishing the fire

For these reasons I assume that ordinary sentential negation is an adjunct and preposed negation is an optional complement of the triggering complementizers. The example in (15) is assigned the structure below.

(15) *hvis ikke regeringen vil*
if not the.government wants.to



4 The Semantics of Preposed Negation

Sentences with preposed and ordinary negation are truth-conditionally equivalent, except in the case of scope-sensitive quantified subjects. However, there is a crucial difference between the two kinds of negation: Corpus investigations[1] reveal that preposed negation does not occur with strong NPIs (in the sense of ([Sailer, 2006, van der Wouden, 1997]), and such examples are judged to be marginal at best. Weak NPIs do occur with preposed negation, but they also occur in conditional clauses and may thus be licensed by the conditional semantics.

[1]This investigation was carried out on *KorpusDK* from *Det Danske Sprog og Litteraturselskab*: http://ordnet.dk/korpusdk.

(16) a. *hvis du ikke løfter en finger for dem, hjælper de heller ikke dig*
if you not lift a finger for them, help they either not you

b. ?? *hvis ikke du løfter en finger for dem, hjælper de heller ikke dig*
if not you lift a finger for them, help they either not you

(17) a. *hvis du ikke giver en skid for hans undskyldninger, hvorfor accepterer du dem så?*
if you not give a damn about his excuses, why do you accept them then?

b. ?? *hvis ikke du giver en skid for hans undskyldninger, hvorfor ...*
if not you give a damn about his excuses, why ...]

In a similar vein preposed negation allows for (strong) positive polarity items (PPI).

(18) a. ??/* *hvis du bare ikke kan det pis, skal du lade være*
if you just not can that crab, don't do it
'if you are not a top-professional, then don't do it'

b. *hvis ikke du bare kan det pis, skal du lade være*
if not you just can that crab, don't do it

Finally, preposed negation is preferred with non-topic-worthy subjects suggesting that the clause does not exhibit a basic topic-focus structure but rather *all_focus*.

(19) a. *hvis ikke nogen gør noget*
if not anyone does something

b. ??/* *hvis nogen ikke gør noget*
if anyone not does anything

This is further corrobated by the fact that indefinite subjects occurring with preposed negation favour a weak (existential) reading, while indefinite subjects with ordinary negation favour a strong (presuppositional/generic) reading. In the theory of indefinites in ([Diesing, 1992]) (VP-internal) focussed subjects favour the weak reading.

(20) a. *Han ville uden tvivl have slået sig ihjel, hvis ikke [en rotte] i det samme var kommet løbende hen over gulvet*
He would beyond doubt have been killed, if not [a rat] in that moment had come running across the.floor

b. *Hvis [en atlet] ikke vil eller glemmer at fortælle Anti-Doping Danmark, ...*
If [an athlete] not wants.to or forgets to tell Anti-Doping Denmark ...

I follow ([Romero, 2006]) in assuming that preposed negation introduces and scopes over a VERUM-predicate in the sense of ([Höhle, 1992]). Ordinary negation in turn scopes over a proposition. Thus the examples in (16a) and (16b) may be paraphrased: *if you do not lift a finger* vs. *if it is not true that you lift a finger*. These paraphrases make clear that preposed negation scopes over a VERUM predicate, which in turn scopes over a positive proposition explaining the observed exclusion of NPIs and licensing of PPIs. However, contrary to the analysis in ([Romero, 2006]) negation of a VERUM-predicate by preposed negation does not generally trigger a presupposition as to whether the speaker believes the proposition to be true or not, as otherwise suggested for Swedish by ([Jensen, 2001]). I believe this to be a matter of complex interaction between context, negation, VERUM and information structure.

5 Analysis

Using the framework of MRS (partly) from ([Warner, 2000]) I suggest the following semantic representations for *hvis Peter ikke vinder*/'if Peter not wins' (1) and *hvis ikke Peter vinder*/'if not Peter wins' (2) respectively.

(1)
$$\left[\begin{array}{l}\textsc{content}\ \left[\begin{array}{l}\textsc{key}\ \boxed{1}\\ \textsc{rels}\ \left\langle \boxed{1}\begin{bmatrix}\textit{cond\_rel}\\ \textsc{hdl}\ \text{h1}\\ \textsc{arg}\ \text{h2}\end{bmatrix}, \begin{bmatrix}\textit{neg\_rel}\\ \textsc{hdl}\ \text{h3}\\ \textsc{arg}\ \text{h4}\end{bmatrix}, \begin{bmatrix}\textit{win\_rel}\\ \textsc{hdl}\ \text{h5}\\ \textsc{arg1}\ \text{h6}\end{bmatrix}, \begin{bmatrix}\textit{name\_rel}\\ \textsc{hdl}\ \text{h6}\\ \textsc{name}\ \textsc{Peter}\\ \textsc{index}\ \text{x}\end{bmatrix}\right\rangle\\ \textsc{conds}\ \left\{\text{h2} \geq \text{h3},\ \text{h4} \geq \text{h5}\right\}\end{array}\right]\end{array}\right]$$

(2)
$$\left[\begin{array}{l}\text{CONTENT}\ \left[\begin{array}{l}\text{KEY}\ \boxed{1}\\ \text{RELS}\ \left\langle \boxed{1}\left[\begin{array}{l}\textit{cond\_rel}\\ \text{HDL}\ \text{h1}\\ \text{ARG}\ \text{h2}\end{array}\right], \left[\begin{array}{l}\textit{neg\_rel}\\ \text{HDL}\ \text{h3}\\ \text{ARG}\ \text{h4}\end{array}\right], \left[\begin{array}{l}\textit{verum\_rel}\\ \text{HDL}\ \text{h5}\\ \text{ARG}\ \text{h6}\end{array}\right], \left[\begin{array}{l}\textit{win\_rel}\\ \text{HDL}\ \text{h7}\\ \text{ARG1}\ \text{h8}\end{array}\right], \left[\begin{array}{l}\textit{name\_rel}\\ \text{HDL}\ \text{h8}\\ \text{NAME}\ \text{Peter}\\ \text{INDEX}\ \text{x}\end{array}\right]\right\rangle\\ \text{CONDS}\ \left\{\text{h2} \geq \text{h3},\ \text{h2} \geq \text{h7},\ \text{h4} \geq \text{h5},\ \text{h6} \geq \text{h7}\right\}\end{array}\right]\end{array}\right]$$

Lexical entry for *ikke*/'not' as a complement:
$$\left[\begin{array}{l}\text{HEAD}\ \textit{adv}\\ \text{CONTENT}\ \left[\begin{array}{l}\text{KEY}\ \boxed{1}\\ \text{RELS}\ \left\langle \boxed{1}\left[\begin{array}{l}\textit{neg\_rel}\\ \text{HDL}\ \boxed{2}\\ \text{ARG}\ \boxed{3}\end{array}\right], \left[\begin{array}{l}\textit{verum\_rel}\\ \text{HDL}\ \boxed{4}\\ \text{ARG}\ \boxed{5}\end{array}\right]\right\rangle\end{array}\right]\\ \text{CONDS}\ \left\{\boxed{3} \geq \boxed{4}\right\}\end{array}\right]$$

Lexical entry for *hvis*/'if':
$$\left[\begin{array}{l}\text{HEAD}\ \left[\begin{array}{l}\textit{compl}\\ \text{MOD}\ \left\langle \text{S}\right\rangle\end{array}\right]\\ \text{COMPS}\ \left\langle \left(\text{ADVP:}\left[\text{CONTENT}\ \left[\begin{array}{l}\text{KEY}\ \mid\ \text{HDL}\ \boxed{4}\\ \text{RELS}\ \boxed{\cdots}\oplus\left\langle\left[\begin{array}{l}\textit{verum}\\ \text{ARG}\ \boxed{5}\end{array}\right]\right\rangle\end{array}\right]\right]\right), \text{S:}\left[\boxed{6}\right]\right\rangle\\ \text{CONTENT}\ \left[\begin{array}{l}\text{KEY}\ \boxed{1}\\ \text{RELS}\ \left\langle \boxed{1}\left[\begin{array}{l}\textit{cond\_rel}\\ \text{ARG}\ \boxed{3}\end{array}\right]\right\rangle\\ \text{CONDS}\ \left\{\boxed{3} \geq \boxed{4},\ \boxed{3} \geq \boxed{6},\ \boxed{5} \geq \boxed{6}\right\}\end{array}\right]\end{array}\right]$$

The complementizer *hvis*/'if' selects an S which it outscopes (3 outscopes 6 in the CONDS). If the optional ADVP is present, the complementizer also outscopes the ADVP, i.e. the negation (3 outscopes 4 in the CONDS). Finally the VERUM outscopes the S (3 outscopes 4 in the CONDS). For ease of exposition the semantics of the matrix-clause to which the conditional clause adjons (the value of the MOD-feature) has not been accounted for.

References

[Borsley and Jones, 2005] Borsley, R. D. and Jones, B. M. (2005). *Welsh Negation and Grammatical Theory*. University of Wales Press, Cardiff.

[Christensen, 2005] Christensen, K. R. (2005). *Interfaces: Negation - Syntax - Brain*. PhD thesis, University of Aarhus.

[Copestake et al., 2005] Copestake, A., Flickinger, D., Pollard, C., and Sag, I. A. (2005). Minimal Recursion Semantics. *Research on Language and Computation*, 3:281–332.

[Diesing, 1992] Diesing, M. (1992). *Indefinites*, volume 20 of *Linguistic Inquiries Monograph*. MIT Press.

[Höhle, 1992] Höhle, T. N. (1992). Über Verum-Fokus im Deutschen. In Jacobs, J., editor, *Informationsstruktur und Grammatik*, volume 4 of *Linguistische Berichte*, pages 112–142. Westdeutscher Verlag.

[Jensen, 2001] Jensen, B. (2001). On sentential negation in the Mainland Scandinavian languages. In *Oxford University Working Papers in Linguistics, Philology and Phonetics*, pages 115–137.

[Johanessen, 2000] Johanessen, J. B. (2000). Negasjonen *ikke*: kategori og syntaktisk posisjon. Tekstlaboratoriet, Universitetet i Oslo.

[Kim and Sag, 2002] Kim, J.-B. and Sag, I. A. (2002). Negation without Head-Movement. *Natural Language and Linguistic Theory*, 20:339–412.

[Romero, 2006] Romero, M. (2006). Biased *yes/no* Questions: The Role of VERUM. *Sprache und Datenverarbeitung*, 30:9–24.

[Sailer, 2006] Sailer, M. (2006). *Don't believe* in Underspecified Semantics. In Bonami, O. and Hofherr, P. C., editors, *Empirical Issues in Syntax and Semantics*, volume 6, pages 375–403.

[Sells, 2000] Sells, P. (2000). Negation in Swedish: What it's not at. In Butt, M. and King, T. H., editors, *Proceedings of the LFG00 Conference*. University of Berkeley.

[van der Wouden, 1997] van der Wouden, T. (1997). *Negative Contexts - Collocation, polarity and multiple negation*. Routledge Studies in Germanic Linguistics. Routledge: London, New York.

[Warner, 2000] Warner, A. (2000). English Auxiliaries without Lexical Rules. In Borsley, R., editor, *Syntax and Semantics: The Nature and Function of Syntactic Categories*, volume 32, pages 167–220. New York: Academic press.

Not as Hard a Problem to Solve as You Might Have Thought

Paul Kay (UC Berkeley) and Ivan A. Sag (Stanford University)

Two understudied phenomena of English are intimately intertwined but insofar as they are studied at all are not usually related. The discontinuous modifier phenomenon (DM) illustrated in (1) and the adjectival pre-determination phenomenon (APD) illustrated in (2) are independent. That is, each of these phenomena may occur independently of the other:

(1) a. [so] willing to help out [that they called early]

b. [too] far behind on points [to quit]

c. [more] ready for what was coming [than I was]

d. [as] prepared for the worst [as anyone]

e. (the) [same] courage in the face of adversity [as yours]

(2) a. [this] delicious A LASAGNA

b. [that] friendly A POLICEMAN

c. [How] hard A PROBLEM (was it)?

The oddity, the "non-core" property, of DM examples like those in (1) is that they appear to call for a discontinuous constituent analysis. The oddity of APD examples like those in (2) is that they present an adjective modifying a determined noun phrase (or DP), rather than a nominal (a common noun or "$\overline{\text{N}}$") – specifically an NP determined by what appears to be the singular indefinite article *a*.

Although, as we have seen in (1) and (2), DM and APD may appear independently, they frequently occur intertwined as in (3):

(3) a. [too] heavy A TRUNK [(for me) to lift]

b. [so] lovely A MELODY [that some people cried]

c. [more] sincere AN APOLOGY [than her critics acknowledged]

d. [as] good A SINGER [as many professionals]

Unsurprisingly, the initial lexical licenser determines the three-way distributional distinction displayed in (1), (2) and (3).

Licensers of DM but not APD include those comparative governors listed in (4):[1]

(4) *same...as, similar...to, equal...to/with, identical... to/with,* ADJ-*er...than, rather...than, prefer...to than, superior...to, inferior...to, other...than, else... than, differ...from, different...from/to/than, dissimilar...to/from*

Licensers of APD but not DM include:

(5) *this, that, how*

And licensers of both DM and APD are listed, exhaustively we believe, in (6):

(6) *so, too, more, less, as, such, how*

It is notable that comparative licensers are split between those that do not [(4)] and those that do [(6)] license APD. There are licensers of APD but not DM, DM but not APD, and both DM and APD.

More than one DM can occur in a clause, as exemplified in (7).

(7) a. [so] much more satisfied than the last time [(that) he couldn't stop smiling]

b. [too] many fewer supporters than her opponent [(for her) to rely on appeals to her base]

c. [enough] bigger an audience than last time [to require standing room only]

In examples such as (7) the multiple DMs form nested dependencies. The corresponding crossed dependencies in (8) are impossible:

(8) a.**[so] much more satisfied [(that) he couldn't stop smiling] than the last time

b.**[too] many fewer supporters [(for her) to rely on appeals to her base] than her opponent

c.**[enough] bigger an audience [to require standing room only]than last time

Other DMs may, however, participate with arguments or modifiers in either nested [(9b,d)] or crossed [(9a,c)] dependencies:

(9) a. Kim was more [willing] than Pat [to wash the dishes].

b. Kim was more [willing] [to wash the dishes] than Pat.

c. I sent out more [books] yesterday than ever before [that I really liked].

d. I sent out more [books] yesterday [that I really liked] than ever before.

[1] See Huddleston and Pullum 2002, p. 1104.

In general,

(10) All DM licensers except *so, too,* and *enough* can participate in crossed dependencies with arguments and other modifiers.

Constructions in SBCG are defined as constraints on construct types; construct types are classes of constructs inter-related in a multiple inheritance hierarchy; and constructs are the feature structures that collectively constitute the model of a language. More specifically, constructions are constraints on construct types that contain non-null mother (MTR) and daughters (DTRS) values. (The DTRS feature is list-valued.)

The major constructions required to account for the above and related data are the Head-Functor Construction, the APD Construction (aka the"big mess" construction), and the Head-Extraposition construction. In addition, lexical entries for the licensing lexical items listed in (4), (5) and (6) are required.

The Head-Functor Construction, is directly adapted from Van Eynde (J. Linguistics 42 (2006), 139186) with the extension that the marking feature is posited allow a range of values corresponding to individual markers or classes thereof. The construction is shown in (11):

(11) **Head-Functor Construction:**

$$\textit{hd-func-cxt} \Rightarrow \left[\begin{array}{ll} \text{MTR} & \left[\begin{array}{ll} \text{SYN} & \left[\begin{array}{ll} \text{CAT} & \left[\text{SEL}\ \ \text{X}\right] \\ \text{VAL} & \text{L}_1 \\ \text{MKG} & \text{Y} \\ \text{EXTRA} & \text{L}_2 \end{array}\right] \end{array}\right] \\ \text{DTRS} & \left\langle \left[\begin{array}{ll} \text{SYN} & \left[\begin{array}{ll} \text{CAT} \left[\text{SEL}\ \ \text{H}\right] \\ \text{MKG}\ \ \text{Y} \\ \text{EXTRA}\ \ \text{L}_2 \end{array}\right] \end{array}\right], \text{H:} \left[\begin{array}{ll} \text{SYN} & \left[\begin{array}{ll} \text{CAT}\ [\text{SEL}\ \text{X}\ ] \\ \text{VAL}\ \text{L}_1 \end{array}\right] \end{array}\right] \right\rangle \end{array}\right]$$

Our analysis of the APD (big mess) phenomenon differs from those of Van Eynde (2007) and Kim (2009) in locating the idiosyncrasy in the AP (modifier-adjective) constituent (e.g., *how big*), rather than the NP (AP-nominal constituent) constituent (e.g., *how big a mess*). The SBCG APD construction is given in Figure 1.

Unlike the Head-Functor Construction, the APD Construction does not describe a *headed* construct-type; there is no head daughter. Accordingly, and critically, the second daughter and the mother do not agree in their SEL value, the daughter selecting an unmarked nominal but the mother selecting a noun phrase marked with the indefinite determiner *a*. The *deg'* MRKG value shared by mother and first daughter licenses presence of just those lexical items that license the construction; they seem to constitute a subclass of (semantic) degree modifiers that resists more restrictive semantic definition. On this analysis the construction combines a *deg'* modifier with an adjective to produce an AP that selects a singular indefinite NP.

apd-cxt $\Rightarrow$

$$\begin{bmatrix} \text{MTR} & \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{CAT} & \textit{adj} \\ & \\ \text{SEL} & \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{CAT} & \textit{noun} \\ \text{MKG} & \textit{a} \end{bmatrix} \\ \text{SEM} & \text{Y} \end{bmatrix} \\ & \\ \text{MKG} & \text{deg'} \\ \text{EXTRA} & \text{L}_1 \end{bmatrix} \end{bmatrix} \\ \text{DTRS} & \left\langle \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{SEL} & \text{X} \end{bmatrix} \\ \text{MKG} & \text{DEG'} \\ \text{EXTRA} & \text{L}_1 \end{bmatrix} \end{bmatrix}, \text{X:} \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \textit{adj} \\ \\ \text{SEL} & \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{CAT} & \textit{noun} \\ \text{MKG} & \textit{unmkd} \end{bmatrix} \\ \text{SEM} & \text{Y} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

Figure 1: Adjectival-Pre-Determiner Construction

Both the Head-Functor and the APD construction have a list-valued EXTRA(position) feature that identifies the mother's value with that of the first daughter (as do several other constructions). Thus, members of the EXTRA list are passed up from first daughter to mother until they are realized by the Head-Extraposition Construction, displayed in (12):

(12) **Head-Extraposition Construction:**

$$\textit{hd-extra-cxt} \Rightarrow \begin{bmatrix} \text{MTR} & \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{VAL} & \text{L}_1{:}\langle(\text{Y})\rangle \\ \text{EXTRA} & \text{L}_2 \end{bmatrix} \end{bmatrix} \\ \text{DTRS} & \left\langle \mathbf{H} \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{VAL} & \text{L}_1 \\ \text{EXTRA} & \langle\, \text{X} \,\rangle \oplus \text{L}_2 \end{bmatrix} \end{bmatrix}, \text{X} \right\rangle \end{bmatrix}$$

This construction realizes the initial element of the first (head) daughter's EXTRA list as the second daughter and eliminates this element from the mother's EXTRA list. (See earlier discussions by Pollard and Sag (1994), Keller (1995).)

The correct ordering possibilities of arguments and various extraposed phrases, as exemplified in (7) and (8), are assured by a construction that pumps a complement of a lexical predicator to an element of its EXTRA LIST (See (13)) along with appropriate lexical entries for the lexical licensers of extraposed phrases and the Head-Extraposition Construction. Two example of relevant lexical entries are shown in (14) and (15). Note that *more* shuffles the extra complement it introduces into a random spot on the EXTRA list of its selectee, whereas

so appends its extra complement to the (right) end of its selectees EXTRA list. This arrangements allows the extra complement of *more* to commute with arguments or other extraposed elements, thus allowing both nested and crossed dependencies, while the complement of *so*, being the final element of the EXTRA list, must be realized at the lowest position and hence can participate in only nested dependencies. (See (8), (9), (10) above.)

(13) **Lexical Extraposition Construction:**

$$\textit{lex-extra-cxt} \Rightarrow \begin{bmatrix} \text{MTR} & \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{VAL} & \text{L}_1 \\ \text{EXTRA} & \text{L}_2 \oplus \langle\, \text{X} \,\rangle \end{bmatrix} \end{bmatrix} \\ \text{DTRS} & \left\langle \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{VAL} & \text{L}_1 \oplus \langle\, \text{X} \,\rangle \\ \text{EXTRA} & \text{L}_2 \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

(14) $$\begin{bmatrix} \text{FORM} & \langle\, \text{SO} \,\rangle \\ \text{SYN} & \begin{bmatrix} \text{CAT} & \left[\text{SEL} \;\; [\text{SYN}\; [\text{EXTRA}\; \text{L}_1]]\right] \\ \text{EXTRA} & \text{L}_1 \oplus \langle \text{S}[\textit{that}] \rangle \end{bmatrix} \end{bmatrix}$$

(15) $$\begin{bmatrix} \text{FORM} & \langle\, \text{MORE} \,\rangle \\ \text{SYN} & \begin{bmatrix} \text{CAT} & \left[\text{SEL} \;\; [\text{SYN}\; [\text{EXTRA}\; \text{L}_1]]\right] \\ \text{EXTRA} & \text{L}_1 \bigcirc \langle \text{XP}[\textit{than}] \rangle \end{bmatrix} \end{bmatrix}$$

The preceding discussion is abbreviated. There are further details about these distributions that are covered in the paper.

REFERENCES

Keller, Frank. 1995. Extraposition of S and PP. Proceedings of EACL.

Huddleston, Rodney, and Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: CUP.

Kim, Jongbok. 2009. The Big Mess construction revisited. Paper presented at BLS 35.

Pollard, Carl, and Ivan A. Sag. 1994. Head-Driven Phrase Structure Grammar. Chicago: U. of Chicago Press and Stanford: CSLI Publications.

Sadler, Louisa, and Douglas J. Arnold. 1994. Prenominal adjectives and the phrasal/lexical distinction. In *Journal of Linguistics* 30: 187-226.

Van Eynde, Frank. 2007. The Big Mess Construction. In Stefan Mller, ed., *The Proceedings of the 14th International Conference on Head-Driven Phrase Structure Grammar*. 415433. Stanford: CSLI Publications. URL: http://cslipublications.stanford.edu/HPSG/8/.

Accounting for underlying forms in HPSG phonology

Filip Skwarski
University of Warsaw

1. Why are underlying forms desirable in HPSG phonology?

The paper aims to present an approach to HPSG phonology which would account for underlying forms of particular phonemes. Phonology in HPSG is largely an undeveloped field, and probably the most significant attempt to expand the value of PHON (in most syntax- and morphology-related papers typically presented as a string equivalent to the word's pronunciation or orthographic spelling) was undertaken by Steven Bird (Bird 1994, Bird 1995).

The framework he proposed expanded PHON into an actual phonological description and allowed formulating constraints determining the well-formedness of surface representations. His approach was based around the principles of COMPOSITIONALITY and a requirement that a framework be MONOSTRATAL (Bird 1995, 1.4.5). The latter meant, in simplified terms, that any phonological representation has only one level, corresponding to forms actually appearing in the surface representations, and no abstract representation is stored.

Although such an approach would seem to be desirable in a computational framework, the phonological phenomena in various languages cannot be adequately described without a further reference to an underlying representation of a phoneme (Shoun 2005, 4.4.). Abby Shoun points out the issue in her study, citing eg. problems arising when representing consonants in Bengali, but does not develop an actual implementation of reference to underlying forms in HPSG phonology - which is the aim of this paper.

Evidence for usefulness of underlying representations can be seen in consonant alternations and voicing processes in languages where those phenomena are complicated, even though Bird seemed to disregard events such as final devoicing as purely phonetic processes which need not be described with binary features (Bird 1995, 3.3.2).

In Polish (my native language), for example, the phoneme /g/ exhibits the following alternations:

(1)	księga	a tome (nom.sg.)	[kɕɛŋga]	[g]
	ksiąg	of tomes (gen.pl.)	[kɕɔŋk]	[k]
	księdze	to a tome (dat.sg.)	[kɕɛndzɛ]	[dz]
	książka	a book (nom.sg.)	[kɕɔ̃w̃ʃka]	[ʃ]
	książek	of books (gen.pl)	[kɕɔ̃w̃ʒɛk]	[ʒ]

Although these alternations result from historical palatalisation and voice assimilation processes, all of them are fully productive in modern Polish, in specific morphology-related cases, like noun declension patterns.

Likewise, in Polish - unlike eg. German - the process traditionally called "final obstruent devoicing" is intertwined with a process of "voice assimilation". Voiced obstruents are devoiced word-finally and before voiceless obstruents, while voiceless obstruents become voiced before voiced obstruents, including across word boundaries (Rubach 1982, 4.2, 4.3). As a result, /d/ and /t/ can both surface as [t] and [d] accordingly, phonetically identical with the "default" form of their opposite-voiced counterpart. Before sonorants (except, in most cases, across word boundaries), obstruents retain their "underlying" voice values, and so, in a traditional monostratal framework, we would have no way of arriving at this basic form if we describe sonorants as either alternations of their surface representations or underspecifications (as suggested by Bird 1995, 1.5).

(2a)	kod	code	[kɔt]	[t]
	kody	codes	[kɔdɨ]	[d]
	kod dostępu	access code	[kɔd dɔstɛmpu]	[d]
	kod miasta	city code	[kɔt mʲasta]	[t]
	kod pocztowy	postal code	[kɔt pɔtʃtɔvɨ]	[t]

(2b)	kot	a cat	[kɔt]	[t]
	koty	cats	[kɔtɨ]	[t]
	kot perski	a Persian cat	[kɔt pɛrski]	[t]
	kot mały	a small cat	[kɔt mawɨ]	[t]
	kot domowy	a housecat	[kɔd dɔmɔvɨ]	[d]

The above data demonstrates that obstruents in Polish can behave in three ways depending on context: assimilate their voice to that of the following segment (before other obstruents, including across word boundaries), retain their "underlying" voice feature (before sonorants, except word-finally), or become voiceless regardless of their "underlying" voice feature (word-finally before a pause or before sonorants) This example will be used as a basis for representing the possibilities of accounting for underlying forms in HPSG phonology, in a further section.

It should be noted that despite the departure resulting from the introduction of underlying descriptions, other crucial ideas of the framework suggested by Bird are not violated. Accounting for the underlying representation does not require "rule ordering", but nonetheless can be used to solve problems traditionally dealt with via rule ordering, eg. opacity in Turkish (as shall be seen in a further paragraph). Similarly, Bird's idea of deletion as an alternation with zero (Bird 1995, 3.2) is still used in the approach I present.

2. Representing the representations

This section is concerned with estabilishing the structural side of the framework which would involve underlying features. A well-functional framework should achieve the following aims:

(a) Allow formulated rules to operate at various levels of the structure (stem, word, syllable, utterance, etc.)
(b) Accurately provide just one surface form for any phoneme in the complete utterance.
(c) Append lower-level representations into higher-level representations (words into phrases, syllables into feet, etc.)
(d) Allow for interactions between the underlying representation and the surface representation in cases where the underlying representation is directly relevant to the surfacing form.

Principle (a) is dictated by the observation that certain phonological phenomena operate within boundaries, such as word boundaries or phrase boundaries, and formulated constraints have to be formulated in a way accounting for this (Bird 1994, 2.2).

Principles (b) and (c) are related: because of observation mentioned in (a), various constraints operating solely on one level of the structure (word, phrase, etc.) would predict different criteria of well-formedness. For example, a constraint demanding that the word-final segment be voiceless would apply to the PHON structure of a *word* object, but not to the PHON structure of a *phrase* object. Similarly, constraints operating across word boundaries would not say anything about the PHON structure of a *word* object.

As a result, for a situation like the exemplary interaction between Polish final devoicing and voice assimilation processes (2a & 2b), we are left with a choice of either predicting different phonological structures for different levels of syntactic and morphological representation, or postulating that all surface representations at all levels

have to be the same. Höhle (Höhle 1999) appears to use (presumably for simplification) the first case scenario, and in his representations, the components of an utterance or a morphologically complex word may contain phonemes (defined for the purposes of sorting) where eg. the voicing feature may differ from the voicing feature of the corresponding phoneme sort in the complete utterance / word. Applying this to our Polish devoicing example would yield a situation in which the phrase "kod dostępu" would have a PHON listing [kɔd dɔstɛmpu], but its first daughter element would have a structure ending in a voiceless obstruent: [kɔt].

Such a solution is possible to account for predictions made at different levels, but causes problems with principle (c), that is, it requires a separate system for appending daughter elements together (since we cannot simply append [kɔt] and [dɔstɛmpu] to get [kod dɔstɛmpu]. Again, introducing underlying representation seems to be an advantage here, as it does not require clearly defined and sorted phonemes (which is superfluous, Shoun 2005), but allows forms to combine precisely because higher level structures are appended based on the underlying structure of their elements, while the surface structure may be separately predicted. Furthermore, because surface structure may be generated based on both the underlying structure and the context, it is possible to make individual surface forms equivalent on all levels of syntactic / morphological representation, thus better adhering to the principle of COMPOSITIONALITY (Bird 1995, 1.4.5).
To summarise - a system I propose is a system where the underlying and the surface forms are stored separately, where the higher level lists are appended separately from underlying and surface form lists of daughter elements, and where the underlying and surface forms can interact through formulated rules. The more complicated interactions (principle (d)) will be seen further on in the section on Turkish epenthesis/deletion opacity.

Below is an exemplary PHON structure provided according to my proposed framework for the English word "cat":

word					
PHON|SEG_LIST	*segs*				
	UR_LIST	[k, æ, t]			
	SR_LIST	[k, æ, t]			
	FIRST	*simple*			
		UR	[k]		
		SR	[k]		
	REST	*segs*			
		UR_LIST	[æ, t]		
		SR_LIST	[æ, t]		
		FIRST	*simple*		
			UR	[æ]	
			SR	[æ]	
		REST	*simple*		
			UR	[t]	
			SR	[t]	

Note that unlike in typical approaches to PHON expansions, the objects utilised in phonological descriptions are not just lists, but objects of type I called here "*segs*" (for "segments"), which are expanded to contain representations containing both the underlying (UR) and surface (SR) forms, as well as lists used to coordinate and append these elements together. The IPA symbols in brackets are obviously simplifactions of complete phonetic structures of the segments, with which I will not deal in detail (see: Bird 1995, "4. A theory of segmental structure", and Höhle 1999). The FIRST and REST features of *segs* are a simplification and a reference to lists, in reality, HPSG ontology would demand them to be named distinctly.

One of the framework's issues now is the relationship between SR and UR. The two basic possibilities of accounting for a possible number of surface representations (here based on the simple example in (2), but note also case (1)) - are alternatives (3) and underspecification (4):

(3)

simple	
UR	[d]
SR	[t] v [d]

simple	
UR	[t]
SR	[t] v [d]

(4)

simple		
UR	*segment*	
	LG|VOICED	+
	SL|PLACE|CORONAL	+
SR|SL|PLACE|CORONAL	+	

simple		
UR	*segment*	
	LG|VOICED	–
	SL|PLACE|CORONAL	+
SR|SL|PLACE|CORONAL	+	

However, because we are equipped with underlying representation now, it is possible to leave the surface representation of any lexical item fully unspecified, and unambiguously derive it based on the UR and the phonological context. While this requires a number of rules to relate particular underlying features to surface ones, it saves the need of stating the SR for every single lexical item. This approach is perhaps closer to the original phonology of derivations and transformations, but should not be confused with any sort of rule ordering and destructive processes. It is, rather, based around constraints and context-sensitive correspondence.

2.1. Word Final Obstruent Devoicing Meets Obstruent Voice Assimilation

The analysis in this section is based around the data and processes in (2), with the goal of adequately describing Polish obstruent voicing processes through HPSG constraints. As mentioned before, there are three elements of the process:

1. Obstruents before other obstruents, including across word boundaries, assimilate their voice to that of the following obstruent, regressively (obstruent clusters have to agree in voicing).
2. Obstruents before sonorants, but not across word boundaries, retain their underlying, distinctive voice.
3. Word-finally, voiced obstruents become voiceless before sonorants or a pause (all word-final obstruents must be voiceless before sonorants or a pause).

The above is true for mainstream Polish, but in south-western variants, the voicing context may be different (Höhle 1999). This will not be dealt with here, although the provided example may easily be altered to account for different voicing phenomena.

The first of the aforementioned rules can be translated into HPSG phonology rather easily, by forming a constraint requiring the surface voice of any two sonorants to be the same - or, in negative terms, prohibiting any cluster with differently voiced obstruents (we furthermore note that the constraint operates on the level of the entire utterance, introducing a subtype of *segs*: *utterance_segs*, required through a constraint demanding that the first *segs* of an *utterance* be of the *utterance_segs* type, and that the REST feature of *utterance_segs* may only be either an actual segment articulation, or another *utterance_segs*).

(5)

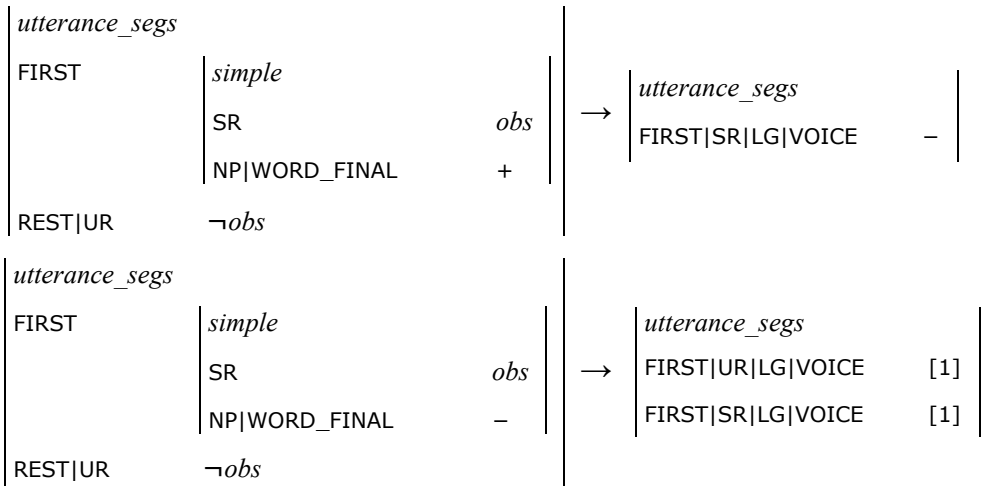
utterance_segs		→	*utterance_segs*	
FIRST|SR	*obs*		FIRST|SR|LG|VOICED	[1]
REST|SR	*obs*		REST|SR|LG|VOICED	[1]

(6)

utterance_segs		→	*utterance_segs*	
FIRST|SR	*obs*		FIRST|SR|LG|VOICED	[1]
REST|FIRST|SR	*obs*		REST|FIRST|SR|LG|VOICED	[1]

The third and the second rule, however, are more complicated, since they crucially revolve around the notion of the word boundary. Because they are interweaved (the word-final context has an exception of preceeding obstruents), however, we cannot just resort to *word_segs*. Instead, however, the structure of *simple* (our object corresponding to a single segment) can be expanded to include a binary feature WORD_FINAL, itself a feature of NP (for "non-phonetic"). We can introduce it through *word_segs*, and translate the rules in (2) and (3) by invoking it:

(7)



The above two constraints effectively require that any obstruent followed by a non-obstruent be voiceless when word-final, and have its surface voicing equivalent to underlying voicing if not word-final. However, they only take into account a situation when the REST element is an actual segment, and not another *segs* beginning with a non-obstruent. Another, analogical pair of constraints is required, their conditions being to this one as (6) was to (5).

2.2. The Issue of Opacity in Turkish: Simultaneous Vowel Epenthesis and Consonant Deletion

Here, we return to the previous topics of representing epenthesis and deletion, as well as the more complex interactions between the underlying and the surface forms in formalted rules. Turkish (Sanders 2003) is an exemplary language where the processes of consonant-triggered Epenthesis and (in some dialects) vowel-triggered k-Deletion are present:

(8) /bɑʃ/ + /m/ -> [bɑʃɯm]

(9) /ɑjɑk/ + /ɯ/ -> [ɑjɑɯ]

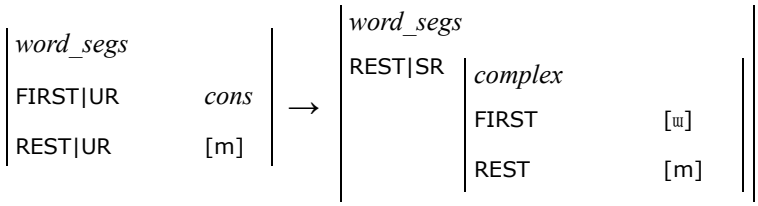
However, in a case traditionally solved through rule ordering, epenthetic vowel may trigger k-Deletion, resulting in both the presence of the epenthetic vowel, and the deletion of /k/:

(10) /ɑjɑk/ + /m/ -> [ɑjɑɯm]

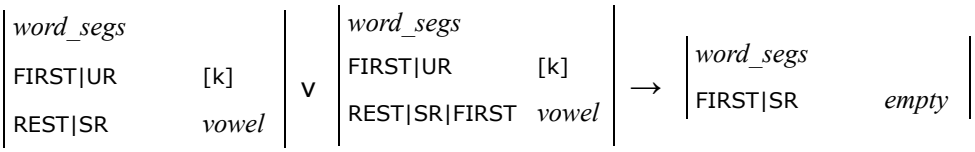
In a rigidly monostratal framework, introducing the constraints prohibiting both consonant clusters and [k] before vowels could, possibly, lead to a situation where neither [ɑjɑkɨm] nor [ɑjɑkm] are considered well-formed while the form [ɑjɑɨm] is, but, first of all, we would have no way to arrive at that form, and secondly, any cluster of two vowels would be acceptable - while in Turkish, that is not the case: except in borrowings, vowel clusters emerge due to the deletion of /k/. To account for this fact, the framework would have to postulate the presence, but not articulation, of /k/, as a ghost segment in the cases where it is deleted, but still present for the purpose of epenthesis.

Ghost segments are an undesired element in phonology, particularly in HPSG, working against the idea of epenthesis as an alternation with zero, and violating the principle perhaps even moreso than introducing underlying features. It is possible, however, to solve the situation in a framework which accounts for underlying representation, by formualting constraints in such a way that they may refer to the underlying context, while affecting the surface structure.

Thus, a constraint translated from the rule of Epenthesis demands any consonant preceeded by an underlying consonant to surface as a complex segment consisting of a vowel and a consonant. It also illustrates the way in which epenthesis is handled in my framework: here, in a simplified situation, /m/ may surface either as a *simple* object, where the surface form is a single phone (as in the cases seen before), or a *complex* object, with its own structure.



The rule of k-Deletion is translated as a constraint demanding that /k/, followed by a surface vowel this time, may not surface, ie. it has to be an *empty* type object, whose phonemic structure contains no surface representation, and so does not appear when appended in higher-level objects.



(note that the context of above constraints is simplified, ignoring situations where the REST object is another *segs*)

Both rules apply precisely at the same time and need not be ordered, but since they refer to different levels of representation, they allow two processes to overlap. On a final note, the *empty* and *complex* types of representation require further rules to establish the technical side of the expanded framework and properly assemble UR and SR lists (in simplified terms, similar to how lists assemble from *segs*), which for space issues are not included in this abstract.

References

Bird, S. 1995. *Computational Phonology: A Constraint-Based Approach*. Cambridge University Press.

Bird, S. and Klein, E. 1994. *Phonological Analysis in Typed Feature Systems.* Computational Linguistics, vol. 20.

Höhle, T. E. 1999. *An Architecture for Phonology.* Slavic in Head Driven Phrase Structure Grammar. Stanford.

Rubach, J. 1982. *Analysis of Phonological Structures*. Państwowe Wydawnictwo Naukowe, Warszawa.

Sanders, R. N. 2003. *Opacity and Sound Change in the Polish Lexicon.* Ph.D. Dissertation: University of California, Santa Cruz.

Shoun, A. 2005. *Interpreting and Generating Formal Grammars: A Study of HPSG Phonology.* Honors essay: Dept. of Linguistics, University of North Carolina.

Phonological change and grammaticalization in HPSG: the case of French final consonants

Jesse Tseng
Cognition Langues Langage Ergonomie UMR 5263
CNRS & University of Toulouse, France
Jesse.Tseng@univ-tlse2.fr

This paper explores the use of HPSG for modeling historical phonological change, focusing on the evolution of word-final consonants in French. The stages of development are presented as a series of overlapping grammars that show how the loss of final consonants, originally a phonological development in Middle French, gave rise to the grammaticalized external sandhi phenomenon known as consonant liaison in modern French.

A striking aspect of French orthography is the prevalence of silent letters, particularly at the ends of words. Final written consonants can be very roughly divided as follows (leaving aside unassimilated loanwords):

(1) a. usually silent: s, x, t, d, g, p, b
b. usually pronounced: c, f, l, r
c. nasals: n, m (generally silent but nasalize preceding vowel)

Silent letters are an indication of the conservative nature of French spelling, but silent final consonants are retained also because there are synchronic grammatical processes that appear to reactivate the corresponding sounds.

Inflectional morphology and the phenomenon known as "consonant liaison" provide many examples of $\emptyset \sim$ C alternations:

(2) a. grand [gʁɑ̃] 'big.*msg*' vs. grande [gʁɑ̃d] 'big.*fsg*'
b. tout nu [tuny] 'totally naked' vs. tout autre [tutotʁ] 'totally different'

In (2a), the written final "d" is silent in the masculine form of the adjective, but it corresponds to a pronounced [d] in the feminine. In (2b), the adverb *tout* is pronounced [tu] before consonant-initial *nu*, but the final "t" is pronounced as a liaison consonant ([tut]) before vowel-initial *autre*.

The synchronic analysis of such phenomena has been the subject of active debate, much of which is centered on the question of abstractness in phonological representation. The data in (2) can be dealt with by assuming underlying phonemic forms containing a final consonant, which must be deleted to derive the appropriate "truncated" forms. The alternations are not always as straightforward as in these examples, however, and more concrete approaches assuming representations closer to the surface forms prove to be more adequate. The analyses proposed in the HPSG literature naturally tend to be surface-oriented (although capturing empirical generalizations always involves some degree of abstractness); see for example Bonami *et al.* (2004) for consonant liaison and Bonami and Boyé (2003) for prenominal adjectives.

Analyses of French phonology in the generative tradition are often deeply inspired by knowledge of diachronic development. A good example of this kind of approach is Casagrande (1984), who proposes highly abstract underlying forms that are basically Latin words, and then a series of ordered rules formalizing the historical sound changes that eventually produced the modern French forms. For a psychologically plausible synchronic analysis, this is obviously taking things too far, but it is nonetheless useful to ask to what extent and how diachronic evidence should be taken into account. And of course the description and analysis of the historical development can be an objective in its own right.

The hypothesis adopted in this paper is that while the grammar of a language can change radically within the space of a few generations, this global change is the sum of smaller, individual changes

that can be modeled in terms of successive, overlapping alternative grammars corresponding to periods of variation (i.e. coexisting, competing analyses) eventually leading to reanalysis (cf. the approach to grammaticalization of Harris and Campbell 1995).

1 Middle French

The general evolution of final consonants in French is relatively well understood (Bourciez and Bourciez, 1967; Zink, 1986). Certain classes of consonants were lost before the end of the Old French period (up to the 13th century); these consonants—primarily nominal case endings and 3rd person verb endings—left no trace in the orthography and have no real relevance in later stages. A more significant development during this period is the loss of most final vowels (other than *-a*), which exposed a large number of new final consonants going into Middle French ("MidFr", 14th–16th centuries).

As mentioned above, French spelling is highly conservative, and the presence of a final consonant in a modern written form is a good indication that the word had a pronounced final consonant in the 13th century (with isolated exceptions, like *et* 'and'). The precise identity of the grapheme may not have remained stable (for example, "z" is often replaced by "s", "s" by "x", and "t" by "d") but final consonant letters are rarely eliminated altogether, even if the corresponding sounds subsequently disappeared from the pronunciation. This means, unfortunately, that the abundant texts of the Middle French period give us very little direct evidence for the sound changes under consideration here. Indirect evidence can be gleaned from rhymes in poetry, but this highly specific context does not reveal changes in the pronunciation of words in connected speech.[1]

Starting in the 16th century, however, the first published grammars of French provide explicit descriptions of actual usage as well as normative recommendations (which become increasingly influential in the following centuries). There are also contemporary proposals for radically reforming French orthography and replacing it with a more transparently phonetic system (e.g. Jacques Peletier in 1550 and Giles Vaudelin in 1713).

From these various sources, it can be established that in MidFr a weakening process affected all final consonants, to varying degrees, subject to the following conditions (Thurot, 1883):

(3) a. Final consonants were lost before a following consonant-initial word.
b. Final consonants were preserved before a following vowel-initial word.
c. Final consonants were preserved before a pause.

The first two contextual conditions are of course the origin of the phenomenon of liaison in modern French. However, the alternation is widely assumed to have started out as a purely phonological process of final consonant deletion in pre-C environments. Thus, while the examples provided by the early grammarians already exhibit many aspects of modern French pronunciation, we also observe some striking differences.

(4) a. mortel peril = morteperil (Palsgrave) vs. ModFr [mɔʁtɛlpeʁil] 'mortal danger'
b. quel monstre = qué' monstre (Duez) vs. ModFr [kɛlmɔ̃stʁ] 'what a monster'

(5) Vous me dites tousiours que vostre pays est plus grand de beaucoup et plus abondant que le nostre
= Vou me dite touiours que votre pays est plu gran de beaucoup et plus abondan que le notre
'You always tell me that your country is much larger and more abundant than ours' (H. Estienne)
vs. ModFr *toujours* [tuʒuʁ], *pays* [pei], *beaucoup* [boku]

This change has been explained as the result of a weakening of the boundaries between words, such that a -C#C- boundary came to be treated like a medial consonant cluster (and thus subject to reduction), while in a -C#V- sequence, the consonant could be resyllabified as the onset of the following syllable. Unfortunately there is insufficient evidence from the relevant period to reveal the precise progression

[1] The presence or absence of a final consonant does not affect the number of syllables, so the scansion of lines of poetry is not affected.

of this evolution. For example, we might expect some final consonants to have been preserved for a time before words starting with "r" or "l" (since consonant-liquid sequences such as "tr-" and "pl-" are possible syllable onsets). Also, while it is unsurprising for the final "s" in -Vs#V- contexts to be pronounced as [z] (like a word-internal, intervocalic "s"), we might expect it to remain unvoiced at first in -Cs#V- contexts, since this has always been the pronunciation of medial "s" in most words like *verser* and *penser*. These hypothetical early steps cannot be confirmed by available data.

Therefore, instead of speculating about and attempting to model the original mechanisms of the change at the level of phrasal phonology, we focus on a later stage at which the alternation is somewhat grammaticalized, so that the alternating forms and the contextual conditions listed in (3) are encoded directly in the grammar. A number of formal approaches are possible in HPSG. We adopt a variant of the PHON-CONTEXT model of Asudeh and Klein (2002), which allows the phonological realization of each daughter within a syntactic construction to vary as a function of the grammatical properties of the immediately following daughter.

(6) P-CONTEXT model (simplified and modified)
construction →

$$\left[\text{DTRS}\ \left\langle \left[\text{PHON} \mid \text{P-CTXT}\ \boxed{1}\right], \boxed{1}\left[\text{PHON} \mid \text{P-CTXT}\ \boxed{2}\right], \ldots, \boxed{n}\left[\text{PHON} \mid \text{P-CTXT}\ \textit{p-ctxt}\right]\right\rangle\right]$$

To model the realization of final consonants, the words in question are assigned lexical entries with complex PHON specifications:

(7)
$$\left[\text{PHON}\ \left[\begin{array}{l}\text{SEGS}\ \boxed{1}\\ \text{P-CTXT}\ \textit{nil}\end{array}\right] \vee \left[\begin{array}{ll}\text{SEGS} & \textsf{s-z}(\boxed{1})\\ \text{P-CTXT} & \left[\text{SEGS}\ \left\langle \textit{vow}, \ldots \right\rangle\right]\end{array}\right] \vee \left[\begin{array}{ll}\text{SEGS} & \textsf{trunc}(\boxed{1})\\ \text{P-CTXT} & \left[\text{SEGS}\ \left\langle \textit{cons}, \ldots \right\rangle\right]\end{array}\right]\right]$$

In this analysis, the pre-pausal form (null P-CTXT) is taken as the basic form (corresponding to the historically original form), the pre-V form is identical except that final [s] undergoes voicing,[2] and the pre-C form is derived by truncation of the final consonant. The adverb *plus*, for instance, has the three realizations [plys], [plyz], and [ply], respectively, while *tout* has the full form [tut] and the truncated form [tu].

While the lexical entry in (7) is appropriate for many words in Middle French, it is an idealization. The transition from the earlier Old French stage (where roughly speaking each word had a single, context-independent pronunciation) involved a long period during which competing forms co-existed, with diverse factors determining their variation. By the time of the first grammars (16th century), some words (or entire classes of words) had not yet reached the stage represented by (7), while others had already undergone further evolution. This partly explains the distribution sketched in (1) above.

2 Modern French

In the transition to modern French ("ModFr", 17th century onwards), two major changes must be accounted for. First, the basic or citation form of alternating words is no longer the form with the final consonant, but the "truncated" form. And second, the liaison alternation becomes highly grammaticalized (losing its purely phonological character and becoming subject to lexical and syntactic constraints).

It is evident from contemporary descriptions that by the end of the MidFr period, condition (3c) was no longer systematically respected. The explanation may be phonological: the weakening of final consonants that first led to truncation before #C- expanded (for whatever reason) to the environment ___]$_{\text{PhonP}}$. It may be analogical: truncated forms were more frequent than non-truncated forms in phrase-internal contexts. And perhaps most significantly, the prevalence of truncated forms diminished the status of the final consonant, which was no longer part of the core phonological content of the word. It became unnecessary for determining lexical identity, so much so that for a large number of words (nouns in

[2] Final [f] probably also became [v], but other voiceless obstruents were not affected.

particular) the final consonant disappeared entirely in all contexts (8a). In other cases, conservative normative influences succeeded in reversing the phonological evolution, and the final consonant was restored in all positions (8b).

(8) a. *clef* [klɛ] 'key', *outil* [uti] 'tool', *coup* [ku] 'blow'
b. *chef* [ʃɛf] 'leader', *net* [nɛt] 'net', *dormir* [dɔʁmir] 'sleep'

In all such cases, the result is a maximal simplification of the MidFr lexical entry in (7), with the neutralization of all contextual distinctions.

For the words that continue to exhibit a liaison alternation in ModFr, the evolution of (7) can be described as follows (leaving aside the voicing alternations encoded by the function s-z):

(9) a.
$$\left[\text{PHON}\ \left[\begin{array}{l}\text{SEGS}\ \ \boxed{1}\vee\boxed{2}\\ \text{P-CTXT}\ \ \textit{nil}\end{array}\right]\vee\left[\begin{array}{l}\text{SEGS}\ \ \boxed{1}\\ \text{P-CTXT}\ \ \left[\text{SEGS}\ \left\langle \textit{vow},\ldots\right\rangle\right]\end{array}\right]\vee\left[\begin{array}{l}\text{SEGS}\ \ \boxed{2}\ \textsf{trunc}(\boxed{1})\\ \text{P-CTXT}\ \ \left[\text{SEGS}\ \left\langle \textit{cons},\ldots\right\rangle\right]\end{array}\right]\right]$$

b.
$$\rightsquigarrow\left[\text{PHON}\ \left[\begin{array}{l}\text{SEGS}\ \ \boxed{2}\\ \text{P-CTXT}\ \ \textit{nil}\vee\left[\text{SEGS}\ \left\langle \textit{cons},\ldots\right\rangle\right]\end{array}\right]\vee\left[\begin{array}{l}\text{SEGS}\ \ \textsf{liaison}(\boxed{2})\\ \text{P-CTXT}\ \ \left[\text{SEGS}\ \left\langle \textit{vow},\ldots\right\rangle\right]\end{array}\right]\right]$$

Variation is introduced in the pre-pausal context—the leftmost disjunct in (9a). This form eventually aligns with the pre-C form 2, and this historically truncated form acquires the status of basic form, as explained above. The pre-V form 1 is consequently reinterpreted as a derived form, as indicated in (9b).

The function liaison cannot represent a simple phonological process. The relation between liaison forms and non-liaison forms is grammaticalized in the form of a two-slot paradigm, which is used in the analysis of all manifestations of liaison in ModFr, including those that have historical origins other than the final consonant deletion described in the previous section. The slots of the paradigm can be filled in in several different ways. In all of the examples considered up to now, the liaison form is derived from the non-liaison form by the addition of an extra final consonant. This "latent" consonant can correspond to an unpredictable (historical) root consonant (10a), or it can be systematically associated with the grammatical features of the word (b). The liaison form can be suppletive (10c,d), or it can be defective (e). And finally, words that show no liaison alternation in ModFr, for example as in (8) above, simply have identical forms in both slots of their paradigm (f).

(10)

	non-liaison form	liaison form	
a.	boku	bokup	*beaucoup* 'a lot'
b.	pəti	pətiz	*petits* 'small.*pl*'
c.	sə	sɛt	*ce* / *cet* 'this'
d.	nuvo	nuvɛl	*nouveau* / *nouvel* 'new' (prenominal)
e.	fʁɑ̃	*	*franc* 'frank' (prenominal)
f.	ku	ku	*coup* 'blow' / *cou* 'neck' / *coût* 'cost'

The increasing grammaticalization of the liaison alternation in the transition to ModFr also motivates a move away from the P-CTXT model assumed in (7) and (9). P-CTXT is appropriate for sandhi phenomena that are primarily phonologically conditioned, because it gives a word direct access to the PHON values of its neighbors. The conditions on liaison in ModFr are no longer primarily phonological. We assume that consonant-initial words in MidFr became associated with an abstract feature [−LIAISON-TRIGGER], encoding the fact that they could not license the appearance of a liaison form. The switch to a non-phonological feature is crucial for the class of "aspirated *h*" words, which lost their initial consonant in early ModFr period (e.g. *hache* 'axe': MidFr [haʃœ] ⇝ ModFr [aʃ]), but still fail to trigger liaison, despite being vowel-initial phonologically.

(11)
$$\left[\begin{array}{ll}\text{PHON} & \left[\text{SEGS}\ \left\langle \text{h},\ldots\right\rangle\right]\\ \text{LTRIG} & -\end{array}\right]\rightsquigarrow\left[\begin{array}{ll}\text{PHON} & \left[\text{SEGS}\ \left\langle \textit{vow},\ldots\right\rangle\right]\\ \text{LTRIG} & -\end{array}\right]$$

The constraints on liaison in ModFr refer to the value of the lexically-specified feature [±LTRIG], instead of directly inspecting the SEGMENTS list of the licensing word.

The strict association between liaison forms and liaison contexts expressed in (7) and (9) must also be relaxed, because in many syntactic environments in ModFr, liaison is optional. The only general constraint is that a liaison form must be immediately followed by a [+LTRIG] word:

(12) a. *beaucoup aimer* [bokueme] / [bokupeme] 'like a lot'
b. *beaucoup manger* [bokumɑ̃ʒe] / *[bokupmɑ̃ʒe] 'eat a lot'

It follows that liaison forms cannot appear in isolation or before a pause.

It should be mentioned, finally, that a small group of alternating words in ModFr have a pre-pausal form with a pronounced final consonant:

(13) a. *huit femmes* [ɥifam] '8 women', *huit hommes* [ɥitɔm] '8 men', *huit* [ɥit] '8'
b. *dix femmes* [difam] '10 women', *dix hommes* [dizɔm] '10 men', *dix* [dis] '10'

These can be considered to be remnants of the MidFr system and we can assume that their lexical entries incorporate contextual constraints that are exceptional with respect to the dominant system. In other words, the behavior of these words does not motivate the extension of the liaison paradigm to three slots for all ModFr words, or the modification of the general constraint illustrated in (12). These words exhibit a good deal of instability, in part as a result of pressure from the more prevalent pattern, but we cannot conclude that they consitute a completely non-productive class. The adverb *plus* 'more', for instance, seems to have re-developed the pre-pausal form [plys] in the latter half of the 20th century. This and other cases of "resurrected" final consonants call for further investigation into the interaction between the competing, partially overlapping analyses in Modern French.

References

Asudeh, A. and Klein, E. (2002). Shape conditions and phonological context. In F. van Eynde, L. Hellan, and D. Beermann, editors, *Proceedings of the 8th International HPSG Conference*, pages 20–30, Stanford, CA. CSLI Publications.

Bonami, O. and Boyé, G. (2003). La nature morphologique des allomorphies conditionnées. Les formes de liaison des adjectifs en français. In B. Fradin, G. Dal, F. Kerleroux, N. Hathout, M. Plénat, and M. Roché, editors, *Les Unités morphologiques / Morphological Units*. Silex, Villeneuve d'Ascq.

Bonami, O., Boyé, G., and Tseng, J. (2004). An integrated approach to French liaison. In G. Jäger, P. Monachesi, G. Penn, and S. Wintner, editors, *Proceedings of Formal Grammar 2004*, pages 29–45. Loria, Nancy.

Bourciez, É. and Bourciez, J. (1967). *Phonétique française. Étude historique*. Klincksieck, Paris.

Casagrande, J. (1984). *The Sound System of French*. Georgetown University Press, Washington, DC.

Harris, A. C. and Campbell, L. (1995). *Historical syntax in cross-linguistic perspective*. Cambridge University Press, Cambridge.

Thurot, C. (1881–1883). *De la prononciation française depuis le commencement du XVIe siècle d'après les témoignages des grammairiens*. Imprimerie nationale, Paris.

Zink, G. (1986). *Phonétique historique du français*. Presses Universitaires de France, Paris, 6th (1999) edition.

On the copula: from a Fregean to a Montagovian treatment

Frank Van Eynde
University of Leuven - Belgium

Abstract

The analysis of the copula as a semantically vacuous word in mainstream HPSG is appropriate for some of its uses, such as the progressive and the passive, but not for its use in clauses with a predicate complement. In such clauses the copula denotes a relation of coreference between the indices of the subject and the predicate complement.

1 The Fregean treatment

In HPSG the copula is commonly treated as semantically vacuous. In Pollard and Sag (1994), for instance, its CONTENT value is identified with that of its predicate complement.

(1) $\begin{bmatrix} \text{CAT} \mid \text{SUBCAT} & \left\langle \text{NP} , \text{XP} [+ \text{PRD}] : \boxed{1} \right\rangle \\ \text{CONTENT} & \boxed{1} \end{bmatrix}$

This treatment can be called Fregean, since it mirrors the practice in predicate logic to omit the copula from the PL formulae.

(2) a. John is a teacher.
b. *teacher*(j)

The usual argument for this treatment is the observation that the omission of the copula does not affect the meaning of the clause, as illustrated in (3). In Russian this also holds for the finite forms, more specifically the present tense, as in (4).[1]

(3) a. John seems (to be) a nice guy.
b. With John (being) ill we cannot go on holiday.

(4) Ona xorosij vrac.
she good doctor
'She is a good doctor.'

At the same time, the treatment of the copula as semantically vacuous creates a number of problems. To spell them out I employ the HPSG style CONTENT values, as defined in Ginzburg and Sag (2000).

[1] Similar observations have been made for Japanese, Hungarian, Arabic, Mauritian Creole and African American Vernacular English, see a.o. Bender (2001), Dalrymple, Dyvik and Holloway King (2004) and Henri and Abeillé (2007).

1. Since the combination of the copula with its predicate complement in (2a) is a VP and since the CONTENT value of a VP is of type *state-of-affairs*, it follows, given the structure sharing in (1), that the predicate complement must denote a state of affairs as well. This is a problem for the nominal predicates, since NPs have a CONTENT value of type *scope-object*.
2. As illustrated by the contrast with (5), it is not only the verb that lacks a semantic counterpart in (2), but also the indefinite article.

(5) a. John knows a teacher.
b. ∃x [*teacher*(x) & *know*(j,x)]

In other words, the article has to be treated differently in predicative and nonpredicative NPs.
3. Since the proper noun and the personal pronoun in (6) denote an entity rather than a state of affairs (or a property), they cannot be treated as predicating over the subject, as in (2b).

(6) a. The winner is Jimmy Logan.
b. That must be her.

The usual way out is to claim that the copula is ambiguous between a predicating sense, as in (2), and an identifying sense, as in (6), and to restrict the Fregean treatment to the former. The problem with this claim, though, is that nobody has taken up the challenge of giving it some substance and empirical bite: "there are a number of delicate semantic issues to be teased out, e.g. the proper analysis of the distinction between the "*be* of identity" and the "*be* of predication". ... We will not attempt to settle these questions here." (Pollard and Sag 1987, 66). The issue is mentioned again in Pollard and Sag (1994, 360), but left unresolved, and it is obliquely referred to in Ginzburg and Sag (2000, 195), curiously enough in a context where the authors plead for a uniform treatment of the predicating and identifying senses: "Note that this analysis [the analysis of pied piping in NPs – A.A.] provides an account of examples like *I wondered [whose cousin] she was pretending to be* __, if we assume that complements of the identity copula are also predicative NPs." In other words, the distinction is not only hard to define, it also gets in the way of a uniform treatment of pied piping.
4. The omission of the copula complicates the treatment of conjunction. Since the semantic counterpart of the conjunction requires (at least) two conjuncts, it has to be omitted, when one of the conjuncts happens to be the copula, as in (7).

(7) John is and remains a crook.

5. If there is no semantic counterpart for the copula, then it is not clear how the assignment of a semantic role to the experiencer denoting pronouns in the Dutch clauses below can be modeled.

(8) a. Wat dit betekent is me nog steeds een raadsel.
what this means is to-me still always a puzzle
'What this means is still a puzzle to me.'

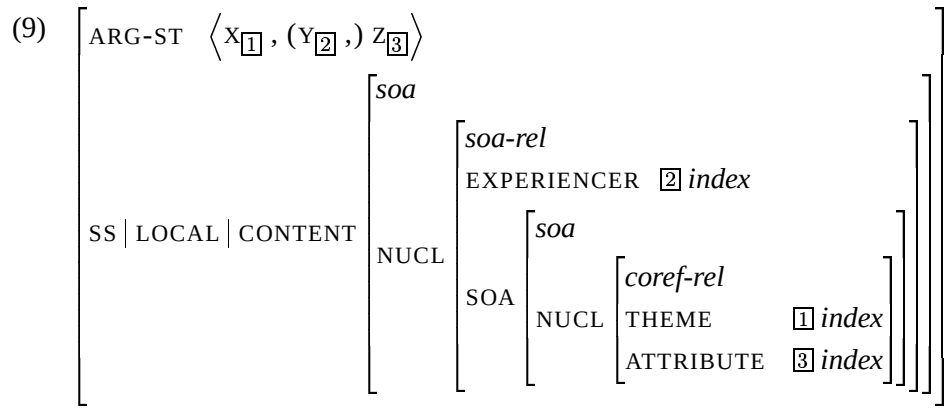
b. Dat kereltje is ons tot last.
that guy-DIM is us to burden
'That little guy is a real burden for us.'

Barring the copula, the only candidate for assigning the role is the predicate complement, but this is not very plausible, since there is no independent evidence for treating nouns like *raadsel* 'puzzle' and prepositions like *tot* 'to' as assigners of the EXPERIENCER role. Moreover, it is at odds with the widely shared assumption that the experiencer receives its role from the verb in *het lijkt me een raadsel* 'it seems a riddle to me'.

Summing up, the omission of the copula from the semantic representations complicates the treatment of the predicate nominals, the indefinite article and the conjunction, it presupposes an as yet unsubstantiated distinction between predicating and identifying uses, and it does not account for the assignment of the EXPERIENCER role.

2 The Montagovian treatment

Building on a suggestion in Quine (1960, 114–118), Richard Montague treats the copula as a two-place relation. More specifically, in the PTQ model (Montague 1974, 247–270), with its distinct representations for disambiguated English and intensional logic (IL), he treats the copula as a transitive verb in disambiguated English and as the relation of identity in intensional logic. The mapping between disambiguated English and intensional logic is slightly different for the copula than for the other transitive verbs, but it does not result in the omission of the copula from the IL representations. Translated in terms of the HPSG style CONTENT values, this means that the copula is not semantically vacuous. Employing the semantic ontology and representation format of Ginzburg and Sag (2000), on the one hand, and the analysis of predicate complements in Van Eynde (2008), on the other hand, I propose the following AVM for the copula.

(9)


In words, the copula takes three syntactic arguments which each have a CONTENT value of type *scope-object*, and introduces a state of affairs whose nucleus is a relation of type *soa-rel*. As in Müller's analysis of the German *erscheinen* 'seem' (Müller 2002, 104), this relation contains an EXPERIENCER attribute and an SOA attribute. The value of the latter is again a state of affairs and has as its nucleus the relation of co-reference, which holds between the indices of the subject and the predicate complement.[2] These indices are co-referent but not token-identical. Token-identity would be too strong a requirement, since the presence of PERSON, NUMBER and GENDER features in the HPSG indices would then impose agreement for these features between the subject and the predicate complement, thus erroneously excluding (10).

(10) a. If I were you, ...
b. We are a good team.

It is easy to show that the Montagovian treatment solves the problems with the Fregean one. First, since the CONTENT value of the VP is now provided by the copula, rather than by its complement, there are no complications with the predicate nominals: their CONTENT value is simply of the same type as that of the other NPs, i.e. *scope-object*. Second, the indefinite article can be treated in the same way in predicative and nonpredicative NPs, as pointed out in Quine (1960, 118) and echoed in Montague (1974, 213). Third, there is no need to differentiate between the predicating and identifying senses: "our uniform symbolization of *be* will adequately cover both the *is* of identity and the *is* of predication." (Montague 1974, 267). Fourth, we no longer need an exceptional treatment for the conjunction when one of the conjuncts is the copula. Fifth, the optional second argument can be assigned the EXPERIENCER role in the same way as the second argument of a verb like *seem*.

[2] If the subject has a non-referential index, as in *it is Friday*, the THEME role is left unassigned. The same holds for the EXPERIENCER role, if there is no constituent which expresses it. Instead of separating the EXPERIENCER role from the two other ones, as in (9), one could treat it as an attribute of the coreference relation and eliminate the *soa-rel* object. This yields a simpler AVM, but an advantage of the more articulate analysis in (9) is that it captures the intuition that the predicate complement and its target form a semantic unit.

3 Non-copular uses of *be*

A defining characteristic of the copula, as represented in (9), is that the predicate complement has a CONTENT value of type *scope-object*. This not only subsumes nominal predicates, but also adjectival, prepositional, gerundial and clausal predicates, as demonstrated in Van Eynde (2008). What it does not subsume, however, are the combinations of *be* with a VP complement that denotes a state-of-affairs, as in (11).

(11) a. They are going home.
b. She was bitten by a big black dog.
c. You are to leave this room at once.

The progressive and the passive *be*, as used in (11a) and (11b), do not introduce a new state of affairs, but inherit the one of their participial complement, as spelled out in (12).

(12) $\begin{bmatrix} \text{ARG-ST} & \left\langle \text{NP} , \text{VP}[\textit{ptc}] : \boxed{2} \right\rangle \\ \text{SS} \mid \text{LOCAL} \mid \text{CONTENT} & \boxed{2}\ \textit{soa} \end{bmatrix}$

By contrast, the modal use of *be* introduces a state of affairs which is distinct from the one of its infinitival complement; it takes the latter as the value of its SOA argument, just like the other modals. This, admittedly, results in a modicum of lexical ambiguity, but as compared to the distinction between the predicating and identifying uses of the copula, the distinctions between copular *be*, progressive *be*, passive *be* and modal *be* are easy to capture and resolve. Moreover, they are independently motivated by the fact that the copular *be* corresponds to the most commonly used copular verbs of other languages, such as *zijn* in Dutch, *sein* in German, and *être* in French, whereas the progressive, passive and modal *be* either have no translational equivalent or one that differs from the copula. The Dutch equivalent of the passive *be*, for instance, is *worden*, rather than *zijn*, the one of the modal *be* is *moeten* or *hebben*, and the progressive *be* has no equivalent in Dutch.[3]

4 Conclusion

The anaylsis of the copula as a semantically vacuous word is appropriate for some of its uses, such as the progressive and the passive, but not for its use in clauses with a predicate complement. In such clauses, it denotes a state of affairs, more specifically a relation of co-reference between the indices of the subject and the predicate complement.

[3]The combination of the copula with the Dutch *aan het* cannot be considered a translational equivalent of the English progressive.

References

Bender, E.(2001), *Syntactic variation and linguistic competence: The case of AAVE copula absence*, PhD thesis, Stanford University.

Dalrymple, M., Dyvik, H. and Holloway King, T.(2004), Copular complements: closed or open?, *in* M. Butt and T. Holloway King (eds), *Proceedings of the LFG04 Conference*, CSLI Publications, Stanford, pp. 188–198.

Ginzburg, J. and Sag, I.(2000), *Interrogative Investigations*, CSLI, Stanford.

Henri, F. and Abeillé, A.(2007), The syntax of copular construction in Mauritian, *in* S. Müller (ed.), *On-Line Proceedings of HPSG 2007*, CSLI Publications, Stanford University, pp. 130–148.

Montague, R.(1974), *Formal philosophy*, Yale University Press, New Haven.

Müller, S.(2002), *Complex predicates*, CSLI, Stanford.

Pollard, C. and Sag, I.(1987), *Information-based Syntax and Semantics*, CSLI Publications, Stanford.

Pollard, C. and Sag, I.(1994), *Head-driven Phrase Structure Grammar*, CSLI Publications and University of Chicago Press, Stanford/Chicago.

Quine, W.(1960), *Word and object*, MIT Press, Cambridge, Massachusetts.

Van Eynde, F.(2008), Predicate complements, *in* S. Müller (ed.), *On-Line Proceedings of HPSG 2008*, CSLI Publications, Stanford University, pp. 253–273.

Abstracts of Reserve Papers

Non-Verbal Predicates in Tongan

Douglas Ball
Stanford University

1 Introduction

In any head-driven theory of clause structure, where clauses are some sort of verbal category, the presence of clauses without an overt verb poses an interesting puzzle. Should or can the non-verbal clauses be assimilated to verbal clauses via a phonologically-null verb, or should they be analyzed in another way? This issue has been the subject of a fair amount of research in HPSG over the last 10 years (Sag and Wasow 1999 and Bender 2001 looking at AAVE, Avgustinova 2006 looking at Russian, and Henri and Abeillé 2007 looking at Mauritian Creole). In this paper, I add another language – the Polynesian language Tongan – to the discussion, both to better understand this kind of construction cross-linguistically and to better understand the nature of Tongan clause structure. After illustrating the differences between verbal and non-verbal clauses in Tongan, I consider both constructional (phrase structure-based) and head-driven (lexical item-based) analyses for the latter kind of clause. I argue that a head-driven approach, where an element within the predicate is the head, is the best solution. However, such an analysis has to split the head's arguments so they do not combine with the head simultaneously.

2 Basic Data

All clauses in Tongan share some properties: the predicational elements appear near the left-edge and the two clause types share some argument realization patterns. Thus, in the verbal clause in (1), the predicate – a verb (bracketed) – appears after a Tense–Aspect–Mood (TAM) word and an absolutive-marked NP (among other arguments) appears:

(1) Na‘e [tō] ‘e Sione ‘a e manioke.
PST plant ERG (name) ABS DET cassava
TAM V Erg NP Abs NP
‘Sione planted the cassava.’

In the non-verbal clause in (2), the predicate (bracketed) also appears early in the clause and there is an absolutive-marked NP (the predicate is initial in (2) because this kind of non-verbal predicate cannot appear with a TAM; others can, in which case the predicate does follow the TAM):

(2) [Ko e faiako] ia.
ESS DET teacher 3SG
Predicate Abs NP
‘He is a teacher.’ (Churchward 1953, 25)

However, the two different types of clauses differ in a number of respects.

The first area of difference is, somewhat obviously, in the presence of a verb. Tongan, unlike many other languages with non-verbal clauses, lacks an alternation between verbless and verb-ful constructions. Instead, all predicate locatives and nominals appear with an initial preposition (like *ko* in (2)) while simplex eventive and property predicates are realized as single words: verbs (like *tō* in (1)).

The second area of difference is in the amount of ordering freedom with nominal expressions. In verbal clauses, the nominal expressions are free to appear in any order after the predicates. So, in addition to the ergative NP $\prec$ absolutive NP order in (1), the reverse order is also possible, as shown in (3):

(3) Na‘e tō ‘a e manioke ‘e Sione.
PST plant ABS DET cassava ERG (name)
TAM V Abs NP Erg NP
‘Sione planted the cassava.’

However, the order of nominal expressions is not free in non-verbal clauses, as shown in with the predicate nominal in (4):

(4) a. Ko e faiako ‘a Sione.
ESS DET teacher ABS (name)
‘Sione is a teacher.’

b. *Ko ‘a Sione e faiako
ESS ABS (name) DET teacher

In fact, the actual predicate nominal must immediately appear after the preposition *ko*.

Furthermore, there are differences in adverbial positioning. A certain class of adverbials canonically appears after the verb in verbal clauses, as shown in (5):

(5) Na‘e tō *foki* ‘e Sione e manioke.
PST plant also ERG (name) ABS.DET cassava
‘Sione also planted the cassava.’

However, in predicate nominals, this kind of adverbial must appear after the predicate, and cannot appear immediately after *ko*, as shown in (6):

(6) a. Ko e faiako *foki* au.
ESS DET teacher also 1SG
‘I am also a teacher.’

b. *Ko *foki* e faiako au
ESS also DET teacher 1SG

Finally, there is a difference in the patterns with coordination. A subject cannot distribute over putative coordinated head + complement units in a verbal clause (7), but a subject can distribute over coordinated predicate nominals (8):

(7) *Na‘e [kai e ika] pea [fufulu e ngaahi tisi] ‘e Sione
PST eat ABS.DET fish and wash ABS.DET PL dish ERG (name)
Intended: ‘Sione ate the fish and washed the dishes.’

(8) [Ko e faiako] pea [ko e tangata fa‘a] ‘a Sione.
ESS DET teacher and ESS DET man farming ABS (name)
‘Sione is a teacher and is a farmer.’

The nominal and adverbial ordering data, plus the coordination data, suggest that the predicate in non-verbal clauses is a phrasal constituent; thus, the principal difference between verbal and non-verbal clauses that the latter has a phrasal predicate constituent. The question then becomes how to ensure that the predicate is, in fact, a phrase in non-verbal clauses.

I assume, following Dukes (2001) and Ball (2008, ch. 3), that verbal clauses (the TAM notwithstanding) form a flat head + arguments structure. So, the verb and its arguments combine simultaneously using Schema 3 from Pollard and Sag 1994, 40, given in (9) using the constraint language and feature geometry of Sign-Based Construction Grammar (Sag 2007):

(9) *head-all-valents-cxt* ⇒

$$\begin{bmatrix} \textit{phrase} \\ \text{SYN} \quad \begin{bmatrix} \text{CAT} & \boxed{0} \\ \text{VAL} & \langle\ \rangle \end{bmatrix} \end{bmatrix} \rightarrow \mathbf{H} \left[\text{SYN} \quad \begin{bmatrix} \text{CAT} & \boxed{0} \\ \text{VAL} & \langle\ \boxed{1}, \ldots, \boxed{n}\ \rangle \end{bmatrix} \right] \boxed{1}, \ldots, \boxed{n}$$

The question then becomes how to deal with non-verbal clauses within this view of verbal clauses. I will consider the same kinds of analyses that Bender (2001, ch. 3) discusses for possible analyses of the verbless construction in AAVE.

3 Zero Copula and Constructional Analyses

An obvious approach to non-verbal clauses is to suppose that non-verbal clauses do have a verb, albeit one with no phonological content (such an analysis is proposed for Tongan's relative Niuean by Massam, Lee, and Rolle (2006)). For Tongan predicate nominals, such a verb would have to select for a PP headed by *ko* and an absolutive-marked NP. Such a verb could give rise to the structure in Figure 1 via (9). While this analysis provides a rationale for why the prepositional phrase behaves as a unit, it suffers

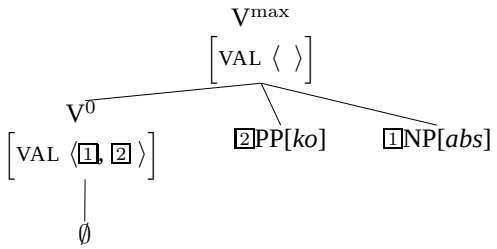


Figure 1: A Zero Copular Analysis of Tongan Predicate Nominals

from several problems. First, given the general syntactic flexibility of verbal arguments in Tongan, there seems no principled reason on the zero copula analysis why the NP[*abs*] and PP[*ko*] could not occur in a different order. Yet they cannot, as shown in (10):

(10) *∅ 'A Sione ko e faiako
(be) ABS (name) ESS DET teacher
Intended: 'Sione is a teacher.'

Furthermore, if postverbal adverbials come after verbs, the predicted slot for adverbials is before the PP[*ko*]. Yet, this order, too, is impossible, as in (11):

(11) *∅ Foki ko e faiako 'a Sione
(be) also ESS DET teacher ABS (name)
Intended: 'Sione is also a teacher.'

Constructional analyses, like those shown in Figure 2, are also problematic. The central problem for

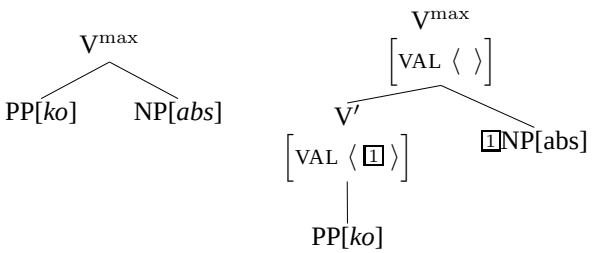


Figure 2: Constructional Analyses of Tongan Predicate Nominals

these accounts is of the same kind that Bender (2001) points out as problematic for the constructional accounts of AAVE zero copula constructions: copular clauses in Tongan do not obligatorily require an subject, as shown in (12):

(12) Ko e tangata faiako foki.
ESS DET male teacher too
'He is a teacher, too.'

The fact illustrated in (12) requires either approach given in Figure 2 to posit either multiple constructions for the full range of data or a phonologically-null *pro* in the syntax, lessening the appeal of either approach.

4 Head-driven Approaches

A simple head-driven solution (extending (9)) where non-verbal clauses are treated exactly like verbal clauses is also not possible. If *ko* is a head with both the NP[*abs*] and NP[*det*] on its VAL list, (9) could license a structure with *ko*, the NP[*abs*], and the NP[*det*] as sisters. However, as long as (9) allows the valents in any order after the head (needed for the 'scrambling' facts illustrated in (1) and (3)), this account would allow for the unacceptable subject-complement and predicate-adverbial orders previously shown in (4b) and (6b). Thus, it appears that, for a head-driven analysis to work, the valents cannot combine with their head simultaneously, but instead have to combine separately in some way.

Any approach that combinatorically separates the 'subject of predication' (NP[*abs*]) argument from the argument involved with the predication (NP[*det*]) can easily derive both the phrasality of the predicate and, with the appropriate general constraint on the linear ordering of heads, also the requisite ordering. Furthermore, by using the head preposition as the locus for stating the possible arguments within the clause (the argument structure) – as any head-driven approach would – also allows for a conventional argument realization theory to be used (such as the one discussed by Ginzburg and Sag (2000, 170fn, 171)), straightforwardly allowing for non-verbal clauses without a overt subject, such as (12).

Deciding precisely how to split the valents thus depends on which account can be best integrated into a larger theory of Tongan clause structure. To this end, I propose that the split is into two combinatoric attributes: VAL and GOV. (The use of the feature GOV follows work on complex predicates by Chung 1998 and others; the instance here, like these complex predicates, also seems to require that a special predicative constituent be built before the arguments can combine with it.) Prepositions would have the combinatoric values in (13):

(13)
$$\begin{bmatrix} \text{VAL} & \langle\ \text{NP}[\textit{abs}]\ \rangle \\ \text{GOV} & \langle\ \text{NP}[\textit{det}]\ \rangle \end{bmatrix}$$

This amounts to dividing the arguments of a predicational preposition into sententially-relevant (VAL) and locally-relevant (GOV) groupings. The addition of the GOV feature also requires that an additional phrase structure schema (construction): one that allows a head to combine with the governed argument(s). And to achieve the desired interaction between the new schema and the earlier schema (9), the head-daughter in (9) must be specified to have an empty GOV list. A canonical predicate nominal would thus be licensed through these two schemata (constructions), as shown in Figure 3.

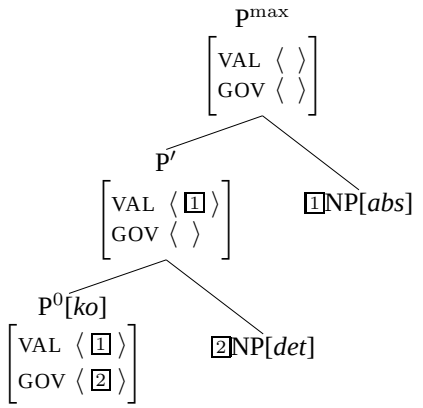


Figure 3: GOV/VAL Split Analysis of Tongan Predicate Nominals

This approach has at least two advantages over the perhaps more obvious split of the VAL list into SUBJ and COMPS features (Pollard and Sag 1994, ch. 9). First, the schema (construction) in (9) can both continue to be utilized for verbal clauses, with only a minimal change, and is extended to cover the final 'level' of the formation of non-verbal clauses as well. On an analysis with a SUBJ/COMPS split, in contrast, how verbal clauses are analyzed would have be re-worked, forcing a decision about what, if anything, is the 'subject' in Tongan, a vexed question (Dukes 1998). Second, the VAL/GOV

split analysis offers a much cleaner analysis of the behavior of 'postverbal' adverbials. As noted in (5) and (6), these adverbials appear after the predicate. With the GOV feature, this generalization is easily captured, regardless of the precise analysis of the adverbials. The adverbials can either uniformly select for an expression with an empty GOV list (subject to some other semantic restrictions) or, taking the adverbials-as-complements approach (Bouma, Malouf, and Sag 2001, for example), these adverbials could be added members of VAL lists. On either approach, the VAL/GOV approach successfully predicts the location of the adverbials as part of general constraints on the interface between linear ordering and immediate dominance.

5 Conclusions

Non-verbal clauses in Tongan are uniformly comprised of a predicational PP + a 'subject of predication'. Trying to analyze this configuration in the same manner as verbal clauses is problematic, as is an analysis with either a zero copula or a special construction. A split-VAL approach solves the problems of these analyses; however, the split I propose is not the traditional SUBJ-COMPS one (as proposed in head-driven accounts of verbless constructions in Bender 2001; Avgustinova 2006; Henri and Abeillé 2007), but one that more directly splits the sentential and prepositional arguments into separate classes.

References

Avgustinova, Tania. 2006. "A Functional Typology of Copular "Be": Towards an HPSG Formalisation." In *Proceedings of 13th International Conference on Head-Driven Phrase Structure Grammar*, edited by Stefan Müller, 27–38. Stanford, CA: CSLI Publications.

Ball, Douglas L. 2008. "Clause Structure and Argument Realization in Tongan." Ph.D. diss., Stanford University.

Bender, Emily M. 2001. "Syntactic Variation and Linguistic Competence: The Case of AAVE Copula Absence." Ph.D. diss., Stanford University.

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

Chung, Chan. 1998. "Argument Composition and Long-Distance Scrambling in Korean: An Extension of the Complex Predicates Analysis." In *Complex Predicates in Nonderivational Syntax*, edited by Erhard W. Hinrichs, Andreas Kathol, and Tsuneko Nakazawa, 159–220. New York: Academic Press.

Churchward, C. Maxwell. 1953. *Tongan Grammar*. London: Oxford University Press.

Dukes, Michael. 1998. "Evidence for Grammatical Functions in Tongan." In *Proceedings of the LFG98 Conference (Workshop on Voice and Grammatical Functions in Austronesian)*, edited by Miriam Butt and Tracy Holloway King, 17 pp. Stanford, CA: CSLI Publications. `http://www.sultry.arts.usyd.edu.au/LFG98/austro/workshop.htm`.

———. 2001. "The Morphosyntax of 2P Pronouns in Tongan." In *The Proceedings of the 7th International Conference on Head-Driven Phrase Structure Grammar*, edited by Dan Flickinger and Andreas Kathol, 63–80. Stanford, CA: CSLI Publications. `http://cslipublications.stanford.edu/HPSG/1/`.

Ginzburg, Jonathan, and Ivan A. Sag. 2000. *Interrogative Investigations: The Form, Meaning, and Use of English Interrogatives*. Stanford, CA: CSLI Publications.

Henri, Fabiola, and Anne Abeillé. 2007. "The Syntax of Copular Constructions in Mauritian." In *Proceedings of the HPSG07 Conference, Stanford University*, edited by Stefan Müller, 130–149. Stanford, CA: CSLI Publications.

Massam, Diane, Josephine Lee, and Nicholas Rolle. 2006. "Still a Preposition: The Status of *Ko*." *Te Reo* 49:3–37.

Pollard, Carl J., and Ivan A. Sag. 1994. *Head-driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Sag, Ivan A. 2007. "Sign-Based Construction Grammar: An Informal Synopsis." Manuscript, Stanford University.

Sag, Ivan A., and Thomas Wasow. 1999. *Syntactic Theory: A Formal Introduction*. 1st Edition. Stanford, CA: CSLI Publications.

Serial Verb Constructions in Mandarin Chinese

Janna Lipenkova and Stefan Müller
Deutsche Grammatik
Institut für Deutsche und Niederländische Philologie
Fachbereich Philosophie und Geisteswissenschaften
FU Berlin
j_lipenkova@yahoo.de|Stefan.Mueller@fu-berlin.de

1 The Phenomenon

The serial verb construction (SVC) is a syntactic structure that exists in a number of languages which are therefore called serializing languages. These are usually languages spoken in areas forming geographic clusters, like West Africa, Oceania, Central America and South-East Asia. SVCs also appear in a great number of pidgins and creole languages. A common characteristic of these languages, many of which are isolating languages, is the lack of a well-developed prepositional and inflectional system. In the course of their evolution, the original SVCs often become subject to grammaticalization and lexicalization processes and develop to prepositional expressions or lexical compounds.

Currently there exists no unambiguous, complete definition of the SVC which could be applied to all serializing languages. However, SVCs are usually considered to have the following properties (Baker, 1989; Chang, 1990; Wälchli, 2009):

- SVCs are composed of two or more verbal phrases following each other without an overt syntactic marking.
- An SVC is used for the description of a single event, which is depicted in two or more sub-events.
- The verbs forming an SVC have a common external argument (subject).
- Additionally, the verbs can also share internal arguments.

Chinese SVCs show all the same syntactic structure: they are composed of two or more VPs consisting each of a verb and at least one object, which stays unrealized in the case of SVCs with common internal arguments (*shared-object-svcs*).[1] Each of the VPs can be marked by the perfective aspect marker *le*. The first VP can be marked by the durative aspect marker *zhe*. Albeit this homogeneous syntactic structure, the Chinese SVC shows a diversity of meanings which constitute a challenge for a formal analysis. The variations are determined by the semantic relation between the events described in the individual VPs. The SVC always denotes a temporal relation (simultaneous or consecutive) between the events. Furthermore, SVCs without aspect marking allow for different causative interpretations. The preferred interpretation is usually chosen with respect to linguistic context and world knowledge. The following examples illustrate the possible semantic constellations.

Example (1) demonstrates a temporal relation between the two VPs. The sentence is an instance of coordination, in which the two constituents cannot be reversed without affecting the semantics of the expression:

(1) Ta1 qi3chuang2 chuan1 yi1fu.
he get.up put.on clothes
'He gets up and (then) puts on his clothes.'

If there is no context presupposing a different relation between the events, they would be interpreted in a consecutive manner.

The examples in (2)–(5) are SVCs with a temporal and functional relation (subordination) between the parts of the SVC. (2) shows examples in which the second VP specifies the purpose of the first one:

(2) a. Ta1 qu3 qian2 qu4 guang4jie4.
he withdraw money go shopping
'He withdraws money and goes shopping.'

[1] An exception is the first VP with durative aspect marking, which can consist of an intransitive verb without object.

b. Ta1 qu3 qian2 hua4.
he withdraw money spend
'He withdraws money to spend it.'

(2a) is an example without a shared object and in (2b) the object of the two verbs is shared, that is, although *qian2* ('money') is the object of both *qu3* ('to withdraw') and *hua4* ('to spend') it is realized only as the object of the first verb.

In (3) the first VP causes the second one.

(3) Ta1 xie3 zi4 ai4 ma3.
he write character suffer critics
'He was criticized for writing.'

The first VP specifies an instrument for the second one in (4):

(4) Ta1 na2 bi3 xie3 zi4.
he take pen write characters
'He writes characters with a pen.'

In (5) the first VP describes the manner in which the action of the second VP takes place:

(5) Ta1 zuo4 qi4che1 qu4 xue2xiao4.
he sit bus go school
'He takes a bus to go to school.'

2 Previous Analyses

An influential SVC-analysis is proposed by Baker (1989); Baker investigates SVCs in Sranan and Yoruba in the framework of GB and makes use of a narrow definition of SVCs. He eliminates coordinative constructions, embedded clauses and small clause predicates from the analysis and considers only those serial constructions which fulfill the criterion of a shared object. He licenses the SVC with the General Serialization Parameter, which states that double-headed constructions are allowed in some languages (serializing languages). The shared internal argument is assigned theta-roles by both verbs. Figure 1 shows the analysis of the Sranan example in (6).

(6) Kofi naki Amba kiri.
Kofi hit Amba die
'Kofi struck Amba dead.'

A HPSG analysis for SVCs in Ga is proposed by Kropp Dakubu, Hellan and Beermann (2007). Ga has two sorts of SVCs: on the one hand, there are lexically fully productive SVCs which show a temporal relation of succession between the two verbal phrases and which are analyzed as instances of modification. On the other hand, Extended Verbal Complexes (EVCs) are

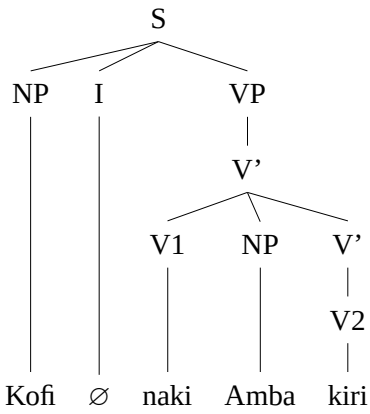


Figure 1: Baker's Analysis of Serial Verb Constructions with Missing Objects

also composed of consecutive verbal phrases; at least one of these phrases contains a so-called preverb subcategorizing for a VP-complement. In this case, the verbs share common arguments. To establish the sharing of an argument inside a VP with a verb that lacks the argument, the authors suggest projecting information about the arguments contained inside the VP. The authors use a feature called QVAL the value of which is a feature structure containing the features SUBJECT and DOBJECT. The values of SUBJECT and DOBJECT are *synsem* objects which are identical to the subject and the object on the SUBJ list and the COMPS list respectively. While the subjects and complements that are represented under VAL|SUBJ and VAL|COMPS are canceled off as usual in HPSG, the elements under QVAL get projected in any case. They can be used to establish the identity between an object that is realized in a VP and a missing object in a second VP.

The features SUBJECT and DOBJECT are reminiscent of grammatical functions in the LFG framework. In what follows it will be shown that LFG like grammatical functions are not needed for SVCs in Mandarin Chinese and we also believe that our account extends to serial verb constructions in Ga.

3 The Analysis

As mentioned, there is yet no unambiguous definition of the SVC in Mandarin. Before proposing an approach for the analysis, we therefore first have to delimit the field of considered constructions by positing further distinctive conditions. Additionally to the general criteria cited above (lack of overt syntactic marking between the VPs, description of a single overall event, common arguments), we assume that the con-

stituents of an SVC can appear on their own as predicative expressions, therefore excluding coverb phrases in which the coverbial expressions subcategorize for a main verb:

(7) a. Ta1 gen1 wo3 qu4 Shang4hai3.
he COV:with/follow me go Shanghai
'He goes with me to Shanghai.'

b. * Ta1 gen1 wo3.
he COV:with/follow me

Furthermore, a change of the positions of SVC constituents leads to an alternation in meaning; this way, we exclude coordinate constructions which show the same syntactic structure, but are semantically different from SVCs in not bearing a definite temporal or functional relationship between the events:

(8) Ta1 tian1tian1 xie3 xin4 da3 dian4hua4.
he everyday write letter make phone.call
'Everyday he writes letters and makes phone calls.'

The two VPs in (8) can be converted without basic change to the meaning: there is no functional relationship between the two events; a temporal relationship (simultaneous, consecutive, alternating) can only be inferred in a specific context.

3.1 Classification of SVCs

We propose an analysis which makes use of both the coordination and the subordination structure. We assume the general constraints in Figure 2 for the type *svc*, which is a supertype of all serial verb constructions in Mandarin. We assume that the semantic contribution of SVCs is introduced on the phrasal level along the lines suggested in Copestake et al., 2005. The constraint in Figure 2 does not restrict the number of daughters and hence can also account for iterative SVCs with several connected VPs like the one in (9):

(9) Ta1 qi3chuang2 chuan1 yi1fu qu4 guang1jie1.
he get.up put.on clothes go shopping
'He gets up, puts on his clothes and goes shopping.'

We then assume the type hierarchy for SVCs that is shown in Figure 3. Iterative SVCs can consist of an unlimited number of VPs denoting a sequence of events. Binary SVCs consists of two VPs which denote events standing in a subordinate semantic relation to each other. An unmarked binary SVC allows for different interpretations of this relation. In our analysis, we concentrate on SVCs with an interpretation that is not context dependent, that is SVCs with aspect marking and SVCs with a shared object. In aspect marked SVCs, the aspect marker semantically acts as a pseudo-conjunction rather then as a real aspect particle and therefore allows for an unambiguous interpretation of the construction. The leaf types of the aspect-marked SVC are differentiated on the basis of the semantic relation between the two constituents, which is determined by the positions of the aspect particles. The SVC with object-sharing described in Baker's syntactic analysis is considered a special type of the binary SVC; in contrast to the African languages Baker concentrates on, Chinese SVCs with shared objects only can bear a purpose relation between the two constituents. There are shared object SVCs that contain an aspect marker in the first VP (*marked-shared-obj-svc*) and shared object SVCs that do not (*unmarked-shared-obj-svc*).

Due to space limitations we cannot discuss the constraints on all these SVCs, but we explain the analysis of unmarked object sharing SVCs in more detail in the next section.

3.2 Analysis of Object Sharing

For the analysis of the shared object construction we assume that arguments are not removed from the SUBCAT list once they are saturated, but rather they are projected and marked as realized. Such a proposal was first suggested in the framework of HPSG by Meurers (1999) and Przepiórkowski (1999) in order to treat case assignment in German. Later Müller (2008) used such a non-cancellation approach to account for depictives in German and English, and Bender (2008) analyzed the rather free constituent order of Wambaya using the information about already realized arguments to allow case agreeing modifiers to refer to their heads which were realized in a different position inside the clause. Instead of listing *synsem* objects as members of the SUBCAT list, a more elaborated data structure is used:

(10) $\begin{bmatrix} \text{ARGUMENT} & \textit{synsem} \\ \text{REALIZED} & \textit{bool} \\ \textit{argument} & \end{bmatrix}$

svc →

$$\left[\begin{array}{l}
\text{SS|LOC|CAT}\ \left[\begin{array}{ll}\text{HEAD} & \left[\begin{array}{l}\text{VFORM}\ \textit{fin}\\ \textit{verb}\end{array}\right]\\ \text{SPR} & \langle\, \boxed{1}\,\text{NP}[\textit{nom}]\,\rangle\end{array}\right]\\
\text{C-CONT}\ \left[\begin{array}{ll}\text{IND} & e\\ \text{RELS} & \left\langle\left[\begin{array}{l}\text{EVENT}\ e\\ \textit{svc-reln}\end{array}\right]\right\rangle\end{array}\right]\\
\text{NON-HEAD-DTRS}\ \left\langle\left[\text{SS|LOC|CAT}\ \left[\begin{array}{ll}\text{HEAD} & \left[\begin{array}{l}\text{VFORM}\ \textit{fin}\\ \textit{verb}\end{array}\right]\\ \text{SPR} & \langle\, \boxed{1}\,\text{NP}[\textit{nom}]\,\rangle\end{array}\right]\right], \ldots, \left[\text{SS|LOC|CAT}\ \left[\begin{array}{ll}\text{HEAD} & \left[\begin{array}{l}\text{VFORM}\ \textit{fin}\\ \textit{verb}\end{array}\right]\\ \text{SPR} & \langle\, \boxed{1}\,\text{NP}[\textit{nom}]\,\rangle\end{array}\right]\right]\right\rangle
\end{array}\right]$$

Figure 2: Constraints on SVCs

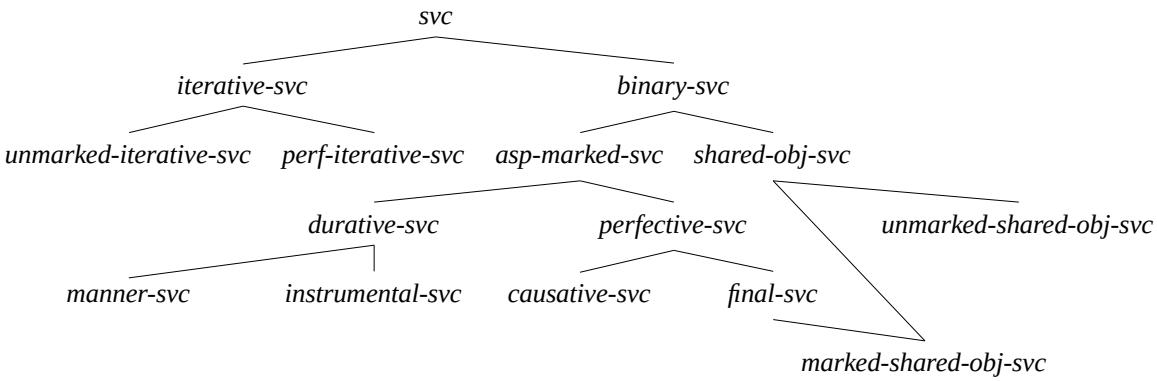


Figure 3: Classification of SVCs in Mandarin Chinese

The value of ARGUMENT is a *synsem* object and the value of REALIZED is '+' if the argument was realized in a head-argument-phrase and '−' if it is still to be saturated. With this change in feature geometry we can formulate the constraint on *shared-obj-svc* as in Figure 4. The object of the first daughter has to be realized. Its *synsem* properties (2) are identified with the *synsem* properties of the unrealized argument of the second daughter. In examples like (11) that involve ditransitive verbs, only the direct object maybe shared.[2]

(11) Ta1 mai3 yi1 jian4 li3wu4 song4 gei3
he buy one CL present offer for/to
wo3.
me
'He buys a present to offer it to me.'

This is captured by the constraints in Figure 4 since it is the first argument on SUBCAT that is shared. The remainder of the SUBCAT list has to be a list of spirits (*los*). This ensures that a VP is connected with an almost saturated VP and excludes lexical verbs with more than one unrealized argument in the second slot.

The semantic constraints on constructions of type *binary-svc* and *final-svc* are shown in Figure 5. The constraint on binary SVCs simply es-

binary-svc →

$$\left[\begin{array}{l}
\text{C-CONT}\ \left[\text{RELS}\ \left\langle\left[\begin{array}{ll}\text{ARG1} & e_1\\ \text{ARG2} & e_2\end{array}\right]\right\rangle\right]\\
\text{NON-HEAD-DTRS}\ \left\langle\left[\text{SS|LOC|CONT|IND}\ e_1\right], \left[\text{SS|LOC|CONT|IND}\ e_2\right]\right\rangle
\end{array}\right]$$

unmarked-shared-obj-svc →

$$\left[\text{C-CONT}\ \left[\text{RELS}\ \langle \textit{final}\rangle\right]\right]$$

Figure 5: Constraints on final SVCs

tablishes an underspecified relation between the subevents of the daughter VPs and the one on final SVCs adds the respective relation.

Figure 6 shows the analysis of example (2b). The verb *qu3* ('to withdraw') is combined with *qian2* ('money'), hence its object requirement 2 is satified and the respective argument continues to be present as a spirit (2) at the VP node. The

[2] Throughout the paper we use the following abbreviations: CL = classifier, COV = coverb, perfASP = perfective aspect particle.

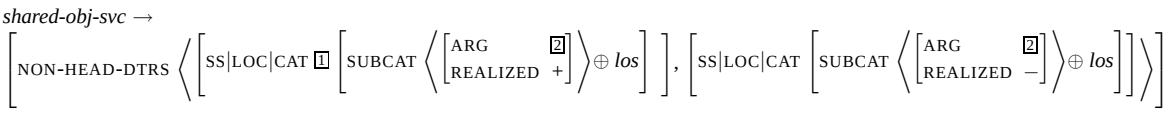


Figure 4: Constraints on shared object SVCs

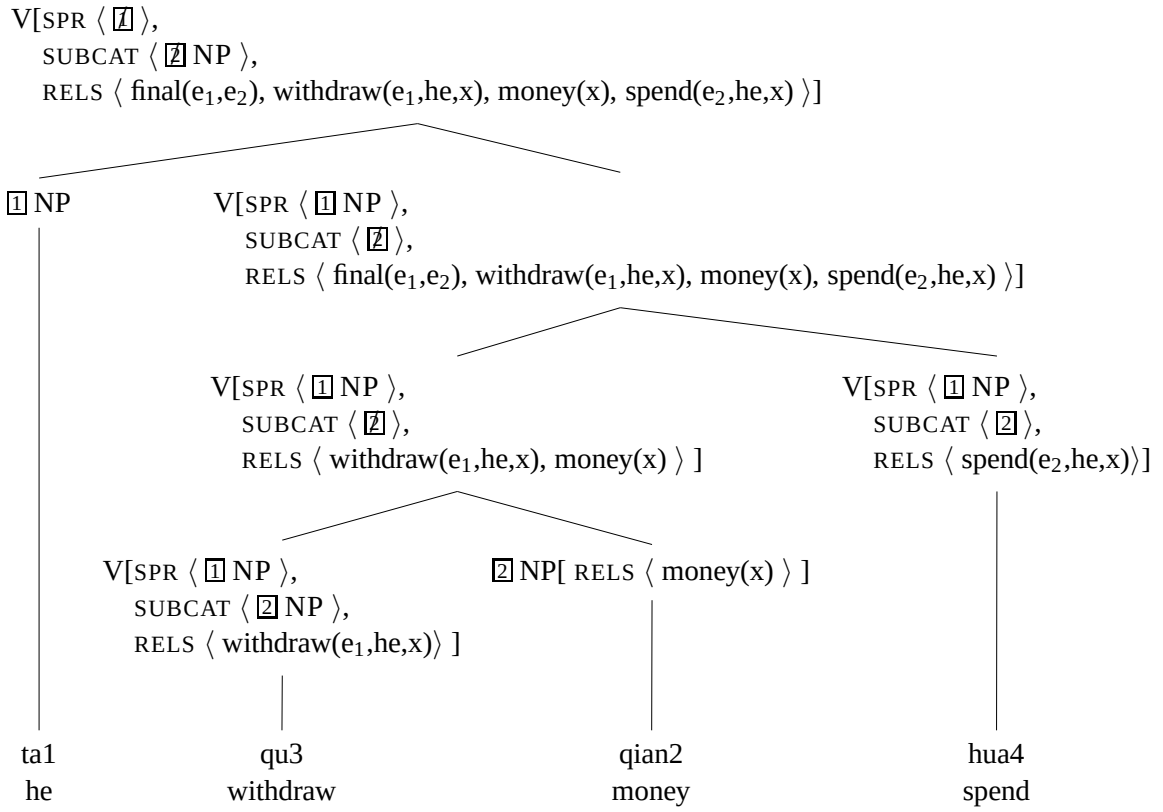


Figure 6: Analysis of SVC with Object Sharing

Final SVC Construction combines the VP with the verb (*hua4* ('to spend'), which is still looking for an argument. The *synsem* information of the missing argument is identified with the *synsem* information of the argument of the first verb which is already realized (2). The result of the combination is a VP that is saturated as far as complements are concerned and that also contributes the relation that describes the relation between the events contributed by the parts of the SVC (*final*(e_1,e_2)). In the last step the complex VP is combined with the subject *ta1* ('he').

4 Conclusion

In this paper we proposed for the first time a classification of SVCs in Mandarin Chinese. We gave an explicit account of argument sharing constructions that relies on the projection of the complete valence information of lexical items, which was independently found necessary for the analysis of constituent order in Wambaya, case assignment in German and depictive secondary predicates in German and English. The analysis does not use empty elements or unary branching trees. The introduction of grammatical functions like subject and object into HPSG was shown to be unnecessary.

The analysis is part of an implemented fragment of Mandarin Chinese.

References

Baker, M. C. 1989. Object Sharing and Projection in Serial Verb Constructions. *Linguistic Inquiry* 20, 513–553.

Bender, E. M. 2008. Radical Non-Configurationality without Shuffle Operators: An Analysis of Wambaya. In S. Müller (ed.), *Proceedings of the 15th International Conference on Head-Driven Phrase Structure Grammar, NICT, Keihanna, Japan*, pp 6–24, Stanford: CSLI Publications.

Chang, C. H. 1990. On Serial Verbs in Mandarin Chinese: VV Compounds and Coverbial Phrases in Chinese. Working Papers in Linguistics 39, Ohio State University.

Copestake, A., Flickinger, D. P., Pollard, C. J. and Sag, I. A. 2005. Minimal Recursion Semantics: an Introduction. *Research on Language and Computation* 4(3), 281–332.

Kropp Dakubu, M. E., Hellan, L. and Beermann, D. 2007. Verb Sequencing Constraints in Ga: Serial Verb Constructions and the Extended Verb Complex. In S. Müller (ed.), *Proceedings of HPSG 07*, pp 99–119, Stanford: CSLI Publications.

Meurers, W. D. 1999. Raising Spirits (and Assigning Them Case). *Groninger Arbeiten zur Germanistischen Linguistik (GAGL)* 43, 173–226.

Müller, S. 2008. Depictive Secondary Predicates in German and English. In C. Schroeder, G. Hentschel and W. Boeder (eds.), *Secondary Predicates in Eastern European Languages and Beyond*, pp 255–273, Oldenburg: BIS-Verlag.

Przepiórkowski, A. 1999. On Case Assignment and "Adjuncts as Complements". In G. Webelhuth, J.-P. Koenig and A. Kathol (eds.), *Lexical and Constructional Aspects of Linguistic Explanation*, pp 231–245, Stanford: CSLI Publications.

Wälchli, B. 2009. *Co-Compounds and Natural Coordination*. Oxford University Press.