LFG 2010 ABSTRACTS

date

JUNE 18-20, 2010

location

TORY BUILDING, CARLETON UNIVERSITY, OTTAWA, CANADA

Agreement versus Pronominal Incorporation in Eurasian Relative Clauses

Farrell Ackerman (UCSD) & Irina Nikolaeva (SOAS)

There is an externally headed prenominal non-subject relative clause type that has been largely neglected in both the typological and theoretical literature. Moreover, it is attested only in genetically related and unrelated languages of Eurasia. It has the following surprising property: *The subject's person/number and pronominal properties are expressed by markers from the nominal possessive paradigm which appear on the external head, i.e., the modified relativized nominal, of the relative clause.* That is, the person number markers (PNM) bear properties of the SUBJ, but these markers appear not within the local domain of the verb with which they are associated, but on the external head of the prenominal relative. This is illustrated schematically in Figure 1:

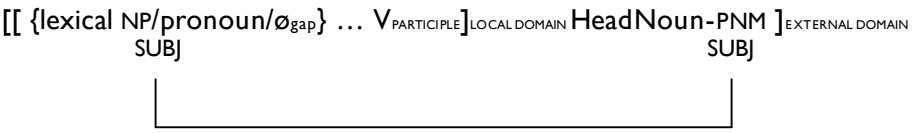


Figure 1

The parellelisms in marking between these relatives and nominal possessive clauses led Ackerman and Nikolaeva 1997, 1998 and Ackerman, Nikolaeva, Malouf 2008 to refer to them as Possessive Relatives.

Over the past few years several analyses from different theoretical traditions have been proposed to address possessive relatives. In this talk we focus on an empirical issue that has been ignored in this literature, but is consequential for any credible analysis. This concerns the status of the PNMs as SUBJ agreement markers versus (incorporated) SUBJ pronominals. Within the syntactocentric perspective of grammar analysis typified by Hale 2002, Kornfilt 2005, 2008, Baker and Vinokurova 2008, Baker 2009 have focused on languages where an independent lexical NP obligatorily co-occurs with a PNM on the relativized head. Consider the examples in (1) and (2) from Sakha (Turkic):

1. *[Masha cej ih-er] caakky. *Absence of PNM
   Masha tea drink-AOR cup
   'a cup that Masha drinks tea from'

2. [**Masha**$_i$ cej ih-er] caakky**-ta**$_i$. Presence of PNM = Agreement
   Masha tea drink-AOR cup-3sg
   'a cup that Masha drinks tea from'

This contrast in grammaticality has led Baker 2009:13 to observe about Sakha that "a participial phrase with an overt subject in argument position is ungrammatical if it lacks possessive agreement. While this appears to be true for Sakha it not true for Tundra Nenets, where a lexical NP cannot co-occur with PNMs, unless the lexical NP functions as a TOPIC anaphorically bound to the PNM interpreted as a pronominal.[1] This can be seen in (3) and (4) below:

3. [Watah ta-wi$_o$] ti Absence of PNM
   Watah-GEN give-PART deer
   `the deer Wata gave'

4. [**Watah** ta-wi$_o$] te-**da** Presence of PNM = Anaphoric binding
Watah-GEN give-PART deer-3SG
`Wata$_i$, the deer he$_i$ gave'

The basic contrast between 3rd SG as an agreement marker versus pronominal should be evident from the contrast between (1) & (2) versus (3) & (4).

With respect to pronominal subjects, it appears that there are three logical options: languages where (i) independent pronouns and PNMs are both obligatory, (ii) independent pronouns are obligatory, while PNMs are optional (Dagur, Yukaghir), and (iii) independent pronouns are optional, while PNMs are obligatory (Tundra Nenets, Ostyak, Evenki). In the domain of Possessive Relatives we are only aware of the existence of (ii) and (iii). On the other hand, glossing conventions can sometimes be misleading in languages for which we have no first-hand evidence. (Kornfilt on Uighur).

Though the distributions and interpretation of PNMs in languages with Possessive Relatives clearly differs, it appears that within a given language their distribution and interpretation are consistent across nominal possessive and Possessive Relative constructions, i.e., whatever behavior is evident for lexical NP/pronoun in one construction, it is the same in the other.

These varying interpretations of the function of PNMs as agreement markers or pronominals clearly recalls the functions of similar markers on predicates in matrix clauses, as observed in Bresnan and Mchombo 1987 and attested in numerous languages since. We argue that all of the principled cross-linguistic flexibility and empirical coverage derived from optional PRED = pro, extends to explain the varying values of PNMs in Possessive Relatives.

We utilize PRED = pro in a constraint-based construction theoretic analysis of Possessive Relatives that accounts in a direct way for the parametric difference between the contrasting functional values of PNMs in Possessive Relatives. On this analysis, the entire Possessive Relative is interpreted formally as a "possessive" construction: this is 2-place relation which is semantically vague with respect to the relation between e.g., the SUBJ of the participle and the relativized head. The participle enters into a modification relation with the head and this has two relevant consequences: (i) the semantics of the verb serves as a restrictor on the interpetation of the otherwise vague semantics associated with the possessive construction, i.e., the relation between the SUBJ of the participle and the relativized is defined by the semantics of the participle and (2) the value of the PNM on the head is identified as the value of the SUBJ of the participle. If a language (or a construction in a language) has an agreement function for the PNM, then only the person/number values are identified (agree) with those of the expressed SUBJ in the relative. If, in contrast, the PNM has a pronominal function, the same mechanism that identifies person/number features with the SUBJ requirement in the former language, now also provides a pronominal value for that SUBJ. As in languages where this latter stategy entails an anaphoric relation between an overt element and a pronominal marker in matrix clauses, the same construal occurs with the co-occurrence of these two elements in Possessive Relatives. Given the pervasive parallelism between nominal possessive constructions and Possessive Relatives, this analysis also extends to the difference in the functional status of PNMs in these constructions as well, where the PNM is sometime a pronominal possesor and sometimes reflects agreement with a possesor. In sum, we provide a minimal parametric difference between the two behaviors of PNMS in Possessive Relatives that is consistent with what has been previously proposed for matrix clauses.

References: Ackerman and Nikolaeva 1997 *Identity in form, difference in function* LFG97; Ackerman 1998 *Construction and mixed categories* LFG98; Ackerman, Nikolaeva, Malouf 2004 *Possessive relatives and cooperating constructions*, HPSG 2004 Proceedings; Hale 2002 *On the Dagur object relative* Journal of East Asian Linguistics; Kornfilt 2005 *Agreement and it placement in nonsubject relative clauses,* Oxford Handbook of Comparative Syntax; Kornfilt 2008 *Subj case and agreement in two types of Turkic RCs,* Leipzig Spring on Language Diversity; Baker and Vinokurova 2008 *Two modalities of case assignment: case in Sakha,* Rutgers ms.; Baker 2009 *Degress of nominalization: Clause-like constituents in Sakha, Rutger ms.*

The Catalan definite article as lexical sharing

Alex Alsina
Pompeu Fabra University

The goal of this paper is to explain a set of puzzles posed by the Catalan definite article (d-article). The main claim of the analysis is (a) that the Catalan d-article is not itself a word, but a word part, a bound morpheme that forms a compound word with a following X^0 and (b) that the derived word has the grammatical category of both component elements: D (determiner), for the d-article, and the category (N, A, C, or P) of the second element. It is thus an instance of lexical sharing (Wescoat 2002, 2005, 2007), whereby a single word instantiates two (or more) X^0 nodes in the c-structure.

The d-article (*el*, *la*, *l'*, etc) has the same broad distribution as Ds such as *aquest* 'this': they are both initial in the NP, (1a), and may immediately precede a variety of categories in the NP, including Ns, postnominal As, PPs, and relative clauses, (1b):

(1) a. *el/aquest noi* 'the/this boy', **noi el/aquest* 'boy the/this'

b. *el/aquest (noi) dolent* 'the/this (boy) bad', *el/aquest que veus* 'the/this that you-see'

However, there are several distributional facts that distinguish the d-article from other Ds such as *aquest*. The d-article, unlike *aquest*, (i) requires a host (i.e., cannot be the sole element of its NP), (2a); (ii) if immediately followed by a PP, this PP must be headed by *de* 'of', (2b); (iii) if immediately followed by a relative clause, this clause must be introduced by *que* 'that', (2c) (cf. (1b)):

(2) a. **Veus el.* 'you-see the' vs. *Veus aquest.* 'you-see this'

b. *el *(de pell) sense caputxa* 'the (leather) one without a hood' (Martí 2002: 1286)
aquest (de pell) sense caputxa 'this (leather) one without a hood'

c. *{aquests/*els} dels quals t'he parlat* '{those/*the} of which I have spoken to you'
*la *(primera) on vaig viure* 'the *(first one) where I lived'

It has other features not found in other Ds: A) it is stressless, as evidenced by the schwa in these forms (əl, lə), and forms a phonological word with the following word (*el nus* [əlnús] 'the knot', *la cua* [ləkúə] 'the tail'); B) the masculine sg. form exhibits the following alternation: *el* before stems beginning by a consonant and *l'* before stems beginning by a vowel, (*el noi*/**l'noi* 'the boy' vs. *l'amic*/**el amic* 'the friend'). This alternation cannot be handled by a general rule, but is restricted to a set of elements consisting of the d-article and verbal clitics, such as *em*/*m'* (1st p.sg.), *et*/*t'* (2nd p.sg.), *es*/*s'* (3rd refl.). In addition, the d-article is homophonous with the 3rd person verbal clitic, giving rise to NP/VP ambiguities such as *la veu* 'the voice' and '(s)he sees her'.

These phonological and morphological facts strongly suggest that the d-article should be given an analysis analogous to that of verbal clitics, which have been argued to be bound morphemes in most of the Romance languages (Miller 1992 and Miller and Sag 1997 for French, and others for Spanish, Italian, and Catalan). By extending this analysis to the d-article (also defended for French by Miller 1992:274–279), we can provide a natural explanation for the *el/l'* alternation, the fact that it is phonologically integrated in the following word, and the fact in (2a) that it needs a host.

Before adopting the lexical sharing approach to the d-article-host compound, we explore the more standard position that it has only one category: either (A) that of the host or (B) that of the d-article. However, both alternatives either fail to explain some facts or do so by complicating the description considerably. Option A fails to explain, without additional machinery, why the article-host compound, whose category is that of the host—spec-A (specificational A or postdeterminer), preN-A (prenominal A), postN-A (postnominal A), N, P, etc.—cannot occur with a D in the same NP and must be initial in the NP. Alternative B, on the other hand, does not face these problems, but it does not explain (i) why, if the host in the article-host compound is an N, the compound D cannot be followed by a spec-A (**el gat altre* 'the cat other'); (ii) why, if the host is a preN-A, the compound D needs to be followed by a noun (**el bon* 'the good', cf. *el bon amic* 'the good friend'); (iii) why, if the host is a preposition, an NP must follow (**el de* 'the of' cf. *el de la pau* 'that of the peace'), or (iv) why, if the host is a C, a clause must follow (**el que* 'the that', cf. *el que vols* 'the one you want). (Miller 1992 and Tseng 2003, among others, adopt alternative A, but, in order to make it work, introduce a set of features (EDGE features) and principles (specific LP constraints and others) that are otherwise not needed and

view it as accidental that the article-host compound appears in the position that corresponds to a D).

The problems just noted do not arise in the approach taken here, where the compound containing the d-article instantiates both a D and the X^0 category corresponding to the host (N, A, P, or C). And the facts are explained by appealing to independently needed principles. On the assumption that there is no more than one D per noun phrase, which follows from the DP hypothesis, and that the head precedes its complements, we explain that, since the article-host compound instantiates the D, there cannot be another D in the same DP and it precedes its NP complement and all words that correspond to it. Problems (i)–(iv) are explained because the article-host compound has the c-structure distribution of both the D and the category of the host: if the host is an N, it cannot be followed by a spec-A, because spec-A occupy the Spec of N and, therefore, precede the N, etc.

The restrictions illustrated in (2b,c) can be explained by imposing a constraint on the grammatical category of the host, which is to be expected in morphological operations. We assume, as a general constraint, that the host that the d-article combines with is of category N or A (or A-spec—specifiers of adjectives) mapping onto the same f-structure; this gives combinations like *el-noi*, *l'altre*, *el-bon*, etc. In addition, there are two lexically listed compounds whose first member is the d-article: d-article+*de* (such as *el de*, *la de*, etc.) and d-article+*que* (such as *el que*, *la que*, etc.); the second member of these compounds maps onto an adjunct of the f-structure of the d-article. Combinations involving the d-article that are not licensed by the general constraint or lexically listed are ill-formed, such as those shown in (2b,c). (See Brucart and Gràcia 1984 for a different explanation for these facts.)

Lexically listing compounds whose first member is the d-article allows us to explain certain combinatorial restrictions. The relative pronoun *qual* can only occur immediately preceded by the d-article, as in *el/la qual*, *els/les quals*. Assuming that *qual* is a spec-A and the second member of a compound with the d-article ensures that these two elements occur together and are initial in their noun phrase. (Note the spelling of the cognate forms in French: *lequel*, *laquelle*, *lesquels*, *duquel*, *desquels*, etc.). The so-called strong possessives *seu*, *meu*, *nostre*, etc. occur either immediately following the d-article or postnominally, in the standard variety: *el seu amic* 'his friend' vs. *cap amic seu* 'no friend his', but **cap seu amic* 'no his friend'. These words are alternatively postN-A or spec-A and, if the latter, part of a compound with the d-article.

An interesting consequence of the proposed treatment of the d-article is that multiple lexical sharing arises when weak prepositions such as *per*, *de* or *a* immediately precede a d-article. Contracted forms such as *pel*, *del* or *al* are argued by Wescoat 2007 to involve lexical sharing (instantiating P and D), but there are reasons to believe that these Ps are always bound morphemes, even when not contracted. Thus, a sequence such as *al peu* 'at the foot' (or *pel camí* or *dels quals*) would be a single word with three categories: P D N (or P D A and others). More complex cases of multiple lexical sharing include examples like *la dels nens* (D P D N) 'the children's' and *pels del teu* (P D P D A) 'by those of yours'. These complex cases, involving the sequence d-article–*de* (which is not possible in French or Italian), constitute a strong argument for a lexical sharing analysis of the d-article: they cannot be accounted for by an extension of the Miller/Tseng approach, where d-article and weak Ps correspond to features of the host's NP. A single word, such as *pel de la noia* 'by that of the girl' (P D P D N), may correspond to a c-structure in which a DP contains another DP. A formal proposal is presented to deal with such complex cases and ensure a correct mapping between word string, c-structure, and f-structure. The arguments for adopting a lexical sharing analysis of the d-article in Catalan indicate that, if we aim for an explanatory theory, this is the correct analysis for the d-article in French and Italian as well.

References

Brucart, José M. and Lluïsa Gràcia. 1984. I sintagmi nominali sense testa: Uno studio comparato. *Rivista di Grammatica Generativa* 11:3–32. **Martí i Girbau, Núria**. 2002. El SN: els noms, ed. Joan Solà et al., *Gramàtica del català contemporani*, vol. 2, 1281–1339. Barcelona: Empúries. **Miller, Philip H**. 1992. *Clitics and constituents in phrase structure grammar*. New York: Garland. **Miller, Philip H. and Ivan A. Sag**. 1997. French clitic movement without clitics or movement. *Natural Language and Linguistic Theory* 15:573–639. **Tseng, Jesse.** 2003. Phrasal affixes and French morphosyntax. *Proceedings of Formal Grammar 2003*, 177–188. Technische Universität Wien. **Wescoat, Michael T**. 2002. *On Lexical Sharing*. Ph.D. dissertation, Stanford University. **Wescoat, Michael T**. 2005. English nonsyllabic auxiliary contractions: an analysis in LFG with lexical sharing. *Proceedings of LFG'05*, 468–486. CSLI Publications. **Wescoat, Michael T**. 2007. Preposition-determiner contractions: An analysis in Optimality-Theoretic Lexical-Functional Grammar with lexical sharing. *Proceedings of LFG'07*, 439–459. CSLI Publications.

Pronominal object shift in the light of object placement in general

Maia Andréasson
Department of Swedish, University of Gothenburg, Sweden

This paper addresses the phenomenon of pronominal object shift in Danish and Swedish and to which extent it is relevant to analyse object shift as an isolated choice between two positions. An extensive corpus investigation shows that there are in these languages distinct strategies for the placement of pronominal objects, that has not previously been discussed in the object shift literature. On the basis of this investigation, we argue that it is necessary to explore all positions for pronominal objects to be able to analyse the underlying causes of pronominal object shift.

Most previous analyses of pronominal object shift in Scandinavian languages focus on pronouns with NP antecedents and on the two positions preceding and following a sentence adverbial, hence the notion object *shift* (cf. Holmberg 1986, 1999; Hellan & Platzack 1995; Josefsson 1992, 2003; Sells 2001; Svenonius 2002; and Vikner 1994, 1997). In (1 a) the pronoun *henne* ('her') is placed in the shifted position preceding the negation *inte*, and in (1 b) in the in situ position following the negation.

(1) a. Jag såg henne inte.[SW]
I saw her not
'I didn't see her.'
b. Jag såg inte henne.[SW]
I saw not her
'I didn't see her.'

The standard assumption is that it is obligatory for weak pronominal objects to shift and that only contrasted or focussed pronominal objects appear in situ in standard Danish. In Swedish, it has been noted that also weak objects appear in situ to some extent.

However, recent research has shown that there is a significant difference in distribution between object pronouns with NP antecedents (here PRON$_{np}$, eg. *henne*, above) and object pronouns with sentence and VP antecedents (here *det*$_{s/vp}$, eg. *det* in example (2) below). In a comparison of the shifted and the in situ position, more *det*$_{s/vp}$ appear in situ in both languages.

(2) a. Jag tror/kan inte det.[SW]
I think/can not that
'I don't think so.'/I can't.'

Furthermore, it has been shown in recent research that while a PRON$_{np}$ is generally contrasted or focussed in situ, a *det*$_{s/vp}$ is generally not both in Swedish and in Danish. For *det*$_{s/vp}$ it is instead the factivity of the matrix verb (cf. Karttunen 1971) that affects the object position. *Det*$_{s/vp}$ with factive matrix verbs have a higher cognitive status (just as PRON$_{np}$, cf. Gundel, Hedberg & Zacharski 1993; Gundel, Hegarty & Borthen 2003; Borthen & Fretheim 1999) and are licensed in the shifted position more frequently than *det*$_{s/vp}$ with non-factive matrix verbs, that appear in situ to a greater extent.

The corpus study of this paper shows that there are significantly different strategies for the placement of pronominal objects with different cognitive status in Danish and in Swedish, and it is not until we include all word order options for pronominal objects that the patterns emerge. Consequently, we argue that – contrary to what have been assumed in previous studies of pronominal object shift – it is not feasible to analyse object shift as a choice between two positions, the in situ position, following the negation and the shifted position preceding it. Both the initial position, and the possibility of omitting the pronoun altogether must be considered in an analysis of object shift.

The new data strongly supports the assumption that factive verbs take cognitively highly accessible pronominal objects. In Swedish, the unmarked option for the factive verbs *förstå* ('understand') and *veta*

('know') turns out to be to leave out the pronominal object entirely in declarative clauses (59–67%), see (3). This indicates that information related to the pronoun is cognitively highly accessible.

(3) Jag vet inte. Jag förstår inte. [SW]
I know not I understand not
'I don't know. I don't understand.'

In Danish pronominal complements to the counterparts of these verbs, *vide* and *forstå*, are generally not left out, but mostly realised in the shifted position (39–56%) see (4), where only weak – and hence accessible – objects appear, or in the initial position (28–38%).

(4) Jeg ved det ikke. Jag forstår det ikke [DA]
I know it not I understand it not
'I don't know. I don't understand.'

When the matrix verb is non-factive, i.e. *tro* and *tycka/synes* ('think, believe'), the objects are not left out, and they are rare in the shifted position in Danish (10–27%), and even more so in Swedish (1–10%). For non-factive verbs it is instead the initial position that is the unmarked option in declarative clauses (for Danish 73–88% and for Swedish 64–84%). In questions, where there is no possibility of placing an object in the initial position, the in situ placement dominates both in Danish and in Swedish for non-factive verbs.

An investigation of only two positions, shifted or in situ, would here for example wrongly lead to the conclusion that pronominal objects to factive verbs in Swedish appear in situ in up to 50% of the cases, when the real number is in fact that only 3% appear in this position. The corpus data including all possible positions instead leads to an analysis where a pronominal object to a factive predicate gets the value 0 for the ACTVN feature in the i-structure (cf. O'Connor 2006). In Swedish, these will generally be linked to a c-structure with a zero instantiation of the object. In Danish, these objects are instead normally linked to a c-structure where the object is in the shifted position. Pronominal objects to non-factive predicates will normally get a higher ACTVN value, due to them being less accessible, and they will normally be linked to the in situ position or the initial position, depending on sentence type and other factors.

Selected references

Gundel, J., N. Hedberg & R. Zacharski 1993. Cognitive Status and the Form of Referring Expressions in Discourse. *Language*, Vol. 69, No. 2 (Jun., 1993), p. 274–307.

Gundel, J., K. Borthen & T. Fretheim 1999. The Role of Context in Pronominal Reference to Higher Order Entities in English and Norwegian. In: P. Bouquet et al. (Eds.): *CONTEXT'99*, LNAI 1688, p. 475–478.

Gundel, J., M. Hegarty & K. Borthen 2003. Cognitive Status, Information Structure, and Pronominal Reference to Clausally Introduced Entities. *Journal of Logic, Language and Information* 12, p. 281–299.

Hellan, L., & C. Platzack 1995. Pronouns in Scandinavian languages: an overview. *Working papers in Scandinavian Syntax* 56, p. 47–69.

Holmberg, A. 1986. *Word Order and Syntactic Features in the Scandinavian Languages and English*. Stockholm: Institutionen för lingvistik, Stockholms universitet.

Holmberg, A. 1999. Remarks on Holmberg's Generalization. *Studia Linguistica* 53, p. 1–39.

Josefsson, G. 1992. Object shift and weak pronominals in Swedish. *Working Papers in Scandinavian Syntax* 52, p. 1–28.

Josefsson, G. 2003. Four Myths about Object Shift in Swedish – and the Truth... In: *Grammar in focus: festschrift for Christer Platzack 18 November 2003*. Vol. II. Lund: Institutionen för nordiska språk, Lunds universitet, p. 199–207.

Karttunen, L. 1971 Some observations on factivity. In: *Research on Language & Social Interaction*, 4:1, p. 55–69.

O'Connor, Robert 2006. *Information Structure in Lexical-Functional Grammar: The Discourse-Prosody Correspondence in English and Serbo-Croatian.* Diss. Manchester: University of Manchester.

Sells, P. 2001. *Structure, Alignment and Optimality in Swedish*. Stanford: CSLI.

Svenonius, P. 2002. Subject positions and the placement of adverbials. In: P. Svenonius (ed.), *Subject, predicates and the EPP*. Oxford: Oxford University Press, p. 199–240.

Vikner, S. 1994. Scandinavian object shift and West Germanic scrambling. In: N. Corver & H. van Riemsdijk (eds.), *Studies on scrambling: movement and non-movement approaches to free word order phenomena*. Berlin: Mouton de Gryuter, p. 487–517.

Vikner, S.. 1997. The interpretation of object shift, Optimality Theory, and Minimalism. *Working Papers in Scandinavian Syntax* 60, p. 1–24.

Pottsian LFG

Doug Arnold & Louisa Sadler
University of Essex

Potts (2005) provides a very appealing account of the semantics of (*inter alia*) 'supplementals' such as appositive relative clauses (ARCs). This paper considers how the approach can be implemented in the architecture of LFG, with ARCs as the focus for exemplification (a side effect is thus to provide a novel account of ARCs in LFG).[1]

Potts' central idea is that the interpretation of every expression involves two dimension: an *at-issue* dimension of normal truth-conditional content, and a 'conventional implicature' *ci-dimension*. In addition to the normal logical types e, t, $\langle e, t\rangle$, etc, his type theory includes *ci-types*, such as $\langle e, t^c\rangle$, which takes a normal 'at-issue' entity to a ci-proposition. In the case of an ARC such as (1) the at-issue content of the subject NP will just correspond to *Kim* (type e), and the ci-content will be the proposition that Sam dislikes Kim (type t^c).

(1) *Kim, who Sam dislikes*, will not come.

Since Potts stipulates that there are no functions of type $\langle a^c, b\rangle$ for any types a, b (that is, no functions from the ci-dimension), this provides an appealing account of the way supplementals are interpreted, e.g. the familiar 'wide scope' behaviour of ARCs.

Potts, (p85ff) notes that the approach seems problematic from a resource sensitivity perspective, since it seems that part of the content is consumed twice (e.g. the content of the host NP contributes to both at-issue and ci-dimensions, so it is consumed twice, once in each dimension). He sketches a solution whereby an emotive adjective like *damn* is associated lexically with a resource like (2).

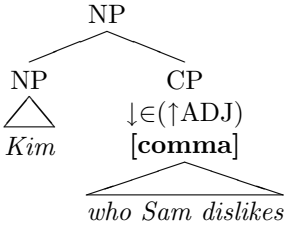
(2) $f \multimap [f \otimes p^c]$

Thus an emotive like *damn* will combine with a noun like *Republican* to produce two resources: an at issue resource (f) corresponding to the normal meaning of *Republican*, and a ci-resource (p^c) corresponding to something like $bad'(Republican')$, which expresses disapproval of Republicanism.

The suggestion is not developed beyond this description of a resource, and he does not consider whether the approach can be generalized to deal with ci-content that is not lexically based. We attempt to remedy this here.

We assume a rather conventional structure for ARCs, as in (3), where [**comma**] is a meaning constructor taking the normal RRC semantics of *who Sam dislikes* into ARC semantics.

(3)



In relation to (3) we propose that the meaning of *Kim, who Sam dislikes,* can be produced by consuming the resources corresponding to *Kim* and *who Sam dislikes*, contributing a 'tensor' resource $K_{\langle e\rangle} \otimes WhatSamDislikes_{\langle t^c\rangle}$, involving resources of types e and t^c.

We follow the analysis of restrictive relatives (RRCs) in Dalrymple (2001, 416ff): semantically, RRCs are functions from noun semantics to noun semantics (i.e. an RRC consumes, and produces, a resource of type $\langle e, t\rangle$ — from a Pottsian perspective, both resources are entirely within the at-issue dimension).

If we abbreviate to $\mathcal{P}$ the actual content of the relative clause, we have (4).

(4) $\lambda Q.\lambda X.\mathcal{P} \wedge Q(X) : [v_{\langle e\rangle} \multimap r_{\langle et\rangle}] \multimap [v_{\langle e\rangle} \multimap r_{\langle et\rangle}]$

Here $v_{\langle e\rangle}$and $r_{\langle et\rangle}$are abbreviations for $(ADJ \in \uparrow)_\sigma VAR)$ and $(ADJ \in \uparrow)_\sigma RESTR)$, which are the resources associated with VAR and RESTR of $(ADJ \in \uparrow)_\sigma$ — the resource corresponding to the modified noun (the noun of which the relative clause is an adjunct).

To make this more concrete, the semantics of the RRC *who Sam dislikes* will be as in (5).

(5) [**who Sam dislikes**] $= \lambda Q.\lambda X.person(X) \wedge\ dislikes(Sam, X) \wedge\ Q(X) : [v_{\langle e\rangle} \multimap r_{\langle et\rangle}] \multimap [v_{\langle e\rangle} \multimap r_{\langle et\rangle}]$

[1] Existing LFG work on related constructions includes Fortmann (2006), where the treatment involves f-structure 'orphans'. This does not deal with the semantics, which is our main focus here. In the full paper we will present evidence that such an orphan approach not appropriate for at least some supplementals, including ARCs.

We define [**comma**] as in (6), where h is an abbreviation for $(ADJ \in \uparrow)_\sigma$ (the resource associated with the host NP).

(6) $[\textbf{comma}] = \lambda P.\lambda Y.[Y, (P(\lambda Z.true))(Y)] : [[v_{\langle e\rangle} \multimap r_{\langle et\rangle}] \multimap [v_{\langle e\rangle} \multimap r_{\langle et\rangle}]] \multimap [h_{\langle e\rangle} \multimap [h_{\langle e\rangle} \otimes h_{\langle t^c\rangle}]]$

On the glue side, this consumes an RRC-like resource, and produces a resource of the kind Potts suggested in (2); on the meaning expression side, it is a function that applies to an RRC meaning expression, does some type lowering (cf. $\lambda Z.true$), and yields an expression $\lambda Y.[Y, T]$, a function from individuals to a pair of meaning expressions. If we abbreviate to $\mathcal{M}$ the restrictive meaning of *who Sam dislikes*, we have (7), expanding this abbreviation, we have (8).

(7) $\lambda P.\lambda Y.[Y, (P(\lambda Z.true))(Y)](\mathcal{M}) : h_{\langle e\rangle} \multimap [h_{\langle e\rangle} \otimes h_{\langle t^c\rangle}]$ =
$\lambda Y.[Y, (\mathcal{M}(\lambda Z.true))(Y)] : h_{\langle e\rangle} \multimap [h_{\langle e\rangle} \otimes h_{\langle t^c\rangle}]$

(8) $\lambda Y.[Y, (\lambda Q.\lambda X.person(X) \wedge dislikes(Sam, X) \wedge Q(X)(\lambda Z.true))(Y)] : h_{\langle e\rangle} \multimap [h_{\langle e\rangle} \otimes h_{\langle t^c\rangle}]$ =
$\lambda Y.[Y, (\lambda X.person(X) \wedge dislikes(Sam, X) \wedge \lambda Z.true(X))(Y)] : h_{\langle e\rangle} \multimap [h_{\langle e\rangle} \otimes h_{\langle t^c\rangle}]$ =
$\lambda Y.[Y, (\lambda X.person(X) \wedge dislikes(Sam, X) \wedge true)(Y)] : h_{\langle e\rangle} \multimap [h_{\langle e\rangle} \otimes h_{\langle t^c\rangle}]$

If the meaning constructor associated with *Kim* is $Kim{:}h_{\langle e\rangle}$, we can now produce (9).

(9) $\lambda Y.[Y, (\lambda X.person(X) \wedge dislikes(Sam, X) \wedge true)(Y)](Kim) : h_{\langle e\rangle} \otimes h_{\langle t^c\rangle}$ =
$[Kim, (\lambda X.person(X) \wedge dislikes(Sam, X) \wedge true)(Kim)] : h_{\langle e\rangle} \otimes h_{\langle t^c\rangle}$ =
$[Kim, (person(Kim) \wedge dislikes(Sam, Kim) \wedge true)] : h_{\langle e\rangle} \otimes h_{\langle t^c\rangle}$

Thus, corresponding to *Kim, who Sam dislikes,* we have, on the meaning side, a pair of meanings (corresponding to *Kim*, and the proposition that Sam dislikes Kim). On the glue side, we have two corresponding resources, one in the at-issue dimension, and one in the ci-dimension.

In order to deal with these resources separately, we will need a new inference rule, as in (10) (inspired by the Context Split rule of Dalrymple (2001, p297)):

(10) $\dfrac{[M, M'] : R_e \otimes R_{t^c}}{M : R_e \quad M' : R_{t^c}}$

The resource corresponding to *Kim* can now be consumed by the main verb, in the normal way. If – in Pottsian fashion – we assume that there are no meaning constructors that consume ci-resources, the end result will be two resources: an at-issue resource corresponding to *Kim will not come*, and a ci-resource expressing the proposition that Sam dislikes Kim. The ci-resource will remain entirely separate from the at-issue content, accounting for the wide-scope interpretation of the ARC.

This is a promising result – it suggests that we can directly incorporate Potts's approach, and his analyses of particular phenomena, into the LFG framework. It raises two kinds of question, which are addressed in the full paper.

First, there are questions of empirical adequacy: does the account capture/explain the well-known idiosyncrasies of the construction in English, and does it generalize to other kinds of supplemental, e.g. appositive NPs (*Kim, a cyclist*)?

Second, it is interesting to ask whether there are alternative implementations of Potts ideas, and whether they involve substantive differences and/or empirical advantages. For example, can the projection architecture of LFG be exploited to dispense with Potts' non-standard ci-types? Suppose we introduce an additional semantic projection, *ci*, separate from, but similar to the normal σ-projection. In place of the glue expression in (8), we might have something like $h_{\langle e\rangle} \multimap [h_{\langle e\rangle}, \uparrow_{ci}]$, where the resource associated with the antecedent NP belongs to the σ projection (as usual), but the resource associated with the ARC belongs to this *ci*-projection. A further, even simpler, possibility would be to directly associate the content of the ARC with the root clause, i.e. to replace the glue in (8) with something like $h_{\langle e\rangle} \multimap [h_{\langle e\rangle}, \uparrow^*_\sigma]$, where $\uparrow^*$ abbreviates an inside-out functional uncertainty expression that denotes the root f-structure.

References

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

Christian Fortmann. The complement of *verba dicendi* parentheticals. In Miriam Butt and Tracy Holloway King, editors, *The Proceedings of the LFG '06 Conference*, University of Konstanz, Germany, 2006. URL `http://csli-publications.stanford.edu/LFG/11/lfg06fortmann.pdf`.

Christopher Potts. *The Logic of Conventional Implicatures*. Oxford University Press, Oxford, 2005.

Putting it all together: Agreement, incorporation, coordination and external possession in Wubuy

Brett Baker (New England), Kate Horrack (New England), Rachel Nordlinger (Melbourne), Louisa Sadler (Essex)

In this paper we examine the interaction of a number of grammatical phenomena in Wubuy, a polysynthetic language from northern Australia, and show how they can be given a comprehensive analysis within the framework of LFG. While each of these phenomena – noun incorporation, verbal agreement, coordination and external possession – has received various treatments within the LFG literature, no one study has addressed the compatibility of these analyses under interaction, despite the fact that they frequently co-occur in the world's languages. In this paper we use data from Wubuy to showcase the effects of this interaction, and investigate the implications for LFG. We show how standard LFG treatments of agreement and coordination combine effortlessly with the analysis of incorporation presented in Nordlinger and Sadler (2008) (henceforth NS08) to account for the complex Wubuy data. We also provide an analysis of the external possession construction (building on earlier work in LFG, e.g. Schrock 2007, Lødrup 2009) that can likewise interact appropriately with the rest of the grammar, providing a single unified account of a range of empirical facts. As well as accounting for the Wubuy data, this work has implications for LFG analyses of polysynthetic languages more generally.

Wubuy, like many polysynthetic languages, allows for productive incorporation of body parts, as shown in the following examples in which we see *-yarrga-* 'flipper' (1) and *-l̲anarr-* 'nail' (2) incorporated into the verbal word:[1]

(1) *nga-ngu-yarrga-gambana (ngarra-ngarruga-l̲ij)*
1SG-3FEM-flipper-roast.PR FEM.TOP-dugong
'I'm roasting the dugong's (FEM) flipper (NEUT)'

(2) *na-l̲anarr ngayawinyinyung*
MASC.TOP-nail 1sg.GEN
nga-ni-l̲anarr-wawayuwaa
1SG-3MASC-nail-cut.PC
'I was cutting off my nails (MASC)'

As is clear from examination of these two examples, however, incorporated body parts participate in two different construction types. (1) is an External Possession Construction (EPC), in which the whole (or possessor) is encoded as direct object. This is evidenced by (i) the fact that the object verb agreement (here, *-ngu-*) shows noun class agreement with 'dugong' (i.e. FEM) and not 'flipper' (NEUT); and (ii) the lack of genitive/oblique case marking on the external possessor NP, which shows it to be a core argument of the verb. The incorporated body part may be doubled by an external NP, which must appear in oblique case (3) showing it *not* to be a core argument of the verb. Example (4) shows that incorporation of the body part is not obligatory in EPC constructions – but that the external NP expressing the part remains in oblique case irrespective of whether or not it is doubled by an incorporated nominal.

(3) *ngaya nga-l̲aan-barrlhiyn yii-l̲aan-duj*
1SG 1SG-knee-sore.REFL.PP MASC.OBL-knee-LOC
'I have sore knee(s)/I am sore in the knee(s)/my knee(s) is/are sore'

(4) *ana-ngarrgu nga-rang*
RESID.TOP-'roo 1SG/RESID-spear.PP
a-lhuganda-rruj
NEUT.OBL-shin-LOC
'I speared the kangaroo in the lower leg'

In (2), the Internal Possession Construction (IPC), the incorporated body part is itself the direct object argument: the verb agrees with it directly (showing MASC object agreement in this case), and a doubled external NP appears in direct (unmarked) case. In the IPC, the possessor must be marked with the genitive case, as (2) also demonstrates.

Despite the difference in predicate-argument relations, and the morphosyntactic reflexes of this, incorporation in both cases can be clearly shown to be of the classifier type (Rosen 1989), since doubling of the incorporated body part is grammatical, and there is no reduction in valency. Furthermore, in both types of construction, the incorporated body part can be coordinated with an external NP, as shown in (5) and (6). In the EPC construction in (5), the part is an OBL and so coordination must be with other oblique NPs for the construction to be grammatical.[2]

(5) *man'-aa-l̲burrunggu, nga-m'-a-n̲ja-wagiwaa marri*
VEG.TOP-turkey, 1SG-VEG-arm-break.PC and
mana-ma-l̲aga
VEG.TOP-VEG.REL-leg
'I broke the wings (lit. 'arms') and the legs of the turkey'

In (6) [the IPC], on the other hand, the part is the object argument and so coordinates with other direct (unmarked) NPs, despite being incorporated:

(6) *wirri-wudu-miyn, marri andhiri, marri*
3PL/3NEUT-liver-get.PP and heart and
baga-l̲ang wirri-ma-ngarrgiwayn
eye 3PL-3VEG-cut.out.PP
'They got the liver (NEUT), and heart (NEUT), and the eye (VEG) they cut out.'

[1] All the examples cited here come from (a subset of) the authors' fieldnotes.

[2] Note that the external part nouns are in the 'relational' noun class form here, in which part nouns take double noun class prefixation to agree with the noun class of the possessor. We regard these forms as obliques since part nouns in relational noun class cannot control verb agreement (like part nouns in the IPC construction do), although they need not take an overt oblique case suffix.

The coordination of incorporated body parts with external NPs has received almost no mention in the literature, and would seem to violate many standard accounts of coordination based on constituent structure. However, as we illustrate below and more extensively in our paper, it follows directly from the interaction of NS08's analysis of nominal incorporation, and standard LFG analyses of coordination (e.g. Dalrymple 2001). (7) provides the lexical entry for the (first) IPC verb in (6), showing the analysis of the incorporated body part as projecting either the OBJ or a member of the OBJ (NS08). External NPs are also annotated with $(\uparrow \text{OBJ} \quad (\in)) = \downarrow$ in the c-structure, resulting in the (partial) f-structure in (8) for the first clause in (6). Note that case agreement amongst the coordinands is enforced by the fact that case is a distributive feature.[3]

(7) *wirri-wudu-miyn*
$(\uparrow \text{PRED}) = \text{'get}< (\text{SUBJ})(\text{OBJ}) >\text{'}$
$(\uparrow \text{OBJ} \quad (\in)) = \downarrow$
$(\downarrow \text{PRED}) = \text{'liver'}$
$(\downarrow \text{INDEX PERS}) = 3$
$(\downarrow \text{INDEX NUM}) = \text{SG}$
$(\downarrow \text{INDEX GEND}) = \text{NEUT}$
$(\uparrow \text{OBJ INDEX GEND}) = \text{NEUT}$

(8)
$$\begin{bmatrix} \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{INDEX} & \begin{bmatrix} \text{GEND} & \text{MASC} \\ \text{PERS} & 3 \\ \text{NUM} & \text{PL} \end{bmatrix} \end{bmatrix} \\ \text{PRED} & \text{'GET}< (\text{SUBJ}) \ (\text{OBJ})>\text{'} \\ \text{OBJ} & \begin{bmatrix} \text{INDEX} & [\,\text{GEND} \quad \text{NEUT}\,] \\ & \left\{ \begin{bmatrix} \text{PRED} & \text{'LIVER'} \\ \text{INDEX} & \begin{bmatrix} \text{PERS} & 3 \\ \text{NUM} & \text{SG} \\ \text{GEND} & \text{NEUT} \end{bmatrix} \\ \text{CASE} & \text{NOM} \end{bmatrix}, \begin{bmatrix} \text{PRED} & \text{'HEART'} \\ \text{INDEX} & \begin{bmatrix} \text{PERS} & 3 \\ \text{NUM} & \text{SG} \\ \text{GEND} & \text{NEUT} \end{bmatrix} \\ \text{CASE} & \text{NOM} \end{bmatrix} \right\} \end{bmatrix} \end{bmatrix}$$

In the EPC construction, we assume that the 'raised' possessor is a non-thematic object of the verb, and is identified with the possessor selected by the incorporated nominal (which is itself an OBL) (9). As shown in the associated (partial) f-structure (10), the POSS will distribute appropriately across all members of the coordinated set. We omit here for reasons of space the semantics NS08 assign to the incorporation, but provide full details of the semantic part of the analysis in the paper.

(9) *nga-ma-laga-wagiwaa*
$(\uparrow \text{PRED}) = \text{'break}< (\text{SUBJ})(\text{OBL}) > (\text{OBJ})\text{'}$
$(\uparrow \text{OBJ}) = (\uparrow \text{OBL POSS})$
$(\uparrow \text{OBL} \quad (\in)) = \downarrow$
$(\downarrow \text{PRED}) = \text{'leg}< (\text{POSS}) > \text{'}$
$(\downarrow \text{INDEX PERS}) = 3$
$(\downarrow \text{INDEX NUM}) = \text{SG}$
$(\downarrow \text{INDEX GEND}) = \text{NEUT}$
$(\uparrow \text{OBJ INDEX GEND}) = \text{VEG}$

(10)
$$\begin{bmatrix} \text{OBJ} & [1] \begin{bmatrix} \text{PRED} & \text{'TURKEY'} \\ \text{INDEX} & \begin{bmatrix} \text{NUM} & \text{SG} \\ \text{PERS} & 3 \\ \text{GEND} & \text{VEG} \end{bmatrix} \end{bmatrix} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{INDEX} & \begin{bmatrix} \text{NUM} & \text{SG} \\ \text{PERS} & 1 \end{bmatrix} \end{bmatrix} \\ \text{PRED} & \text{'BREAK}< (\text{SUBJ}) \ (\text{OBL})> (\text{OBJ})\text{'} \\ \text{OBL} & \begin{bmatrix} \text{INDEX} & [\,\text{GEND} \quad \text{NEUT}\,] \\ & \left\{ \begin{bmatrix} \text{PRED} & \text{'LEG}< (\text{POSS}) >\text{'} \\ \text{INDEX} & \begin{bmatrix} \text{PERS} & 3 \\ \text{NUM} & \text{SG} \\ \text{GEND} & \text{NEUT} \end{bmatrix} \\ \text{CASE} & \text{OBL} \\ \text{POSS} & [1] \end{bmatrix}, \begin{bmatrix} \text{PRED} & \text{'ARM}< (\text{POSS}) >\text{'} \\ \text{INDEX} & \begin{bmatrix} \text{PERS} & 3 \\ \text{NUM} & \text{SG} \\ \text{GEND} & \text{NEUT} \end{bmatrix} \\ \text{CASE} & \text{OBL} \\ \text{POSS} & [1] \end{bmatrix} \right\} \end{bmatrix} \end{bmatrix}$$

Thus, we provide a comprehensive analysis of body part incorporation in Wubuy and show how existing analyses of different aspects of the grammar – external possession, incorporation, agreement and coordination – interact to provide a single analysis of the complex empirical facts. This approach highlights the strength of LFG in accommodating typologically diverse languages, and will have important implications for the analysis of polysynthetic languages cross-linguistically.

[3] We use NOM to refer to the direct (unmarked) case that is found on subjects and objects in Wubuy.

References

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

Lødrup, Helge. 2009. Looking possessor raising in the mouth: Norwegian possessor raising with unergatives. In M. Butt and T. H. King, eds., *On-line Proceedings of the LFG2009 Conference*.

Nordlinger, Rachel and Louisa Sadler. 2008. From juxtaposition to incorporation: an approach to Generic-Specific constructions. In M. Butt and T. H. King, eds., *Proceedings of LFG08 the Conference*. Stanford, CA: CSLI Publications: http://www-csli.stanford.edu/publications.

Rosen, Sara. 1989. Two types of noun incorporation: a lexical analysis. *Language* 64:294–317.

Schrock, Terrill. 2007. Syntactic Exorcism: Toward an LFG analysis of Swahili external possession. Tech. rep.

Second-position and endoclitics in Pashto

Tina Bögel
University of Konstanz

The purpose of this paper is to continue the discussion on the prosody-syntax interface within LFG (Butt and King 1998, O'Connor 2004, MyCock 2006, Bögel et al. 2008, Bögel et al. 2009) by introducing Pashto (endo)clitics, which challenge the view of prosody as being derivative from the syntax (e.g. Selkirk 1984) and the principle of lexical integrity (Bresnan and Mchombo 1995).

Pashto is an Iranian language spoken in Afghanistan and parts of Pakistan. Clitics are quite common in this language and have been thoroughly described in Tegey (1977). This paper focuses on one group of clitics that have special properties, thereby challenging the common understanding of the interaction of morphology, syntax and phonology. They are subject to both prosodic and syntactic contraints and can be inserted INTO the verb (thus questioning the principle of lexical integrity). Numerous approaches have been suggested, involving for example Optimality Theory (van der Leeuw 1997, Roberts 1997, Anderson 2005), prosodic inversion (Halpern 1995, Dost 2005) and Transformational Grammar (Kaisse 1981). However, all of these attempts to account for the clitic phenomena in Pashto have resulted in unnecessarily complex algorithms.

As has been shown for Dutch clitics (Bögel et al. 2009), the tension between the syntactic and prosodic properties of clitics can be solved by assuming an LFG architecture that allows an interaction of prosody and syntax, but decouples the two components. Instead of regarding the prosodic projection as being based on the syntactic tree, prosody and syntax are viewed as equal partners trying to align with each other as much as possible. The prosodic and syntactic properties of Pashto clitics can similarly be accounted for in a simple and intuitive way by assuming prosody and syntax as interacting, but separate and parallel components. This paper shows that it is the prosodic component that is responsible for the placement of the clitics WITHIN words. This leads to the conclusion that in cases of 'disagreement', the prosodic component may take precedence over the syntactic component.

Pashto clitics seem to have the general properties of a second position clitic that attaches to the first element in a specific syntactic constituent (e.g. CP/ S) within which it originates. In (1a), the clitics can be found in the main and in the subordinate clause. In (1b), another syntactic constraint becomes obvious: the clitic(s) may not interrupt a syntactic constituent, the coordinated NP (clitics are underlined).

(1) a. tor me wəlidə [magar spin me wə nə lidə]
Tor I saw but Spin I PF not saw
'I saw Tor, but I didn't see Spin.' (Tegey 1977, 127)

b. [xušαl aw patang]$_{NP}$ ba ye dər ta rαwṛi
Koshal and Patang will it you to bring
'Koshal and Patang will bring it to you.' (Tegey 1977, 84)

However, Pashto clitics are not only subject to syntactic constraints, they are also bound by prosodic requirements. The element they attach to has to bear some sort of lexical stress. In (2), the clitic is placed after the first stress-bearing element, even though this causes the clitic to appear at the right edge of the phrase.

(2) rα ta te rα ẓolawál de
me for from_it here collect you
'You were collecting them for me from it (and bringing them) here.' (Tegey 1977, 119)

Like South-Asian languages in general, Pashto is a argument-dropping language (Butt 2007). Sentences can therefore consist of only a verb and a clitic. The interesting endoclitics appear in these short sentences in the context of an aspect-caused stress shifting; the perfective aspect of the verb is formed with the help of a verb-internal stress shift. The verbs fall roughly into three classes, depending on their word-internal structure. Class 1 verbs are monomorphemic. In the imperfective, they bear stress on the last foot; the clitic is placed after the verb ((3a)). In the perfective, class 1 verbs take on a perfective prefix *wə-* that bears the main stress. In this case, the clitic occurs after the prefix ((3b)):

(3) a. **imperfective**
təxnawála me
tickle I
'I was tickling (her).' (Tegey 1977: 86)

b. **perfective**
wá me təxnawəla (*wátəxnawəla me)
PREF I tickle
'I tickled (her).' (Tegey 1977: 92)

In contrast to class 1 verbs, class 2 and 3 form the perfective by means of a stress shift from the last to the first foot of the verb. The verbs of both clases are bimorphemic: class 2 verbs are formed by a derivational prefix and a root; class 3 verbs are complex predicates consisting of a adjective, adverb or a noun and a light verb. Their behavior with respect to clitics is the same, which is why only an example of a class 2 verb is given ((4)):

(4) a. **imperfective**

porewestó me

carry across I

'I was carrying it across.'

b. **perfective**

póre me westə (*pórewestə me)

PREF I carry across

'I carried it across.' (Tegey 1977: 92)

Apart from these three classes, there is another group of verbs that can have alternative stress in the imperfective, but form the perfective with the perfective prefix of class 1 (*wə-*), thus adopting properties of all other classes. Within this group, there are verbs that begin with consonants, which do not show any special behavior in the imperfective: even if the stress is on the front vowel, the clitic is placed after the verb.

However, there are nine verbs in this group with an initial vowel *a-*, which show a very distinct behavior:

(5) a. **imperfective — stress on second foot**

axistə́lə me

buy I

'I was buying them.'

b. **imperfective — stress on first foot**

á me xistələ

? I buy

'I was buying them.' (Tegey 1977, 89)

There is evidence supporting the fact that the clitic is inserted INTO the verb postlexically. In the perfective, the *a*-verbs take the perfective prefix *wə-* like all other class 1 verbs. In contrast to the consonant-initial verbs however, perfective *a*-verbs display vowel coalescence, a process that is assumed to take place in the lexicon. In (6a), the adjacency of the perfective prefix *wə-* and the *a-* causes a fusion: *wα-*. When a clitic is inserted after the perfective prefix, the vowel coalescence is still present ((6b)), providing evidence that the clitic has actually been inserted INTO the word postlexically.

(6) a. **Vowel coalescence — without clitic**

tə ye wαxla (*wə axla)

you it buy

'You buy it.' (Tegey 1997, 149)

b. **Vowel coalescence — with clitic**

wα ye xla

PREF it buy

'You buy it.' (Tegey 1977, 163)

The above data leads to an interesting conclusion about this particular group of Pashto clitics. If one assumes the syntactic and the prosodic component to be parallel, clitics can be viewed as subject to prosodic and syntactic constraints, respectively. This property is shared by clitics in other languages (e.g. Serbian/Croation/Bosnian; Franks and King 2000), which seems to be natural, given that clitics are prosodically deficient items. By granting prosody an independent and strong position, prosodic constraints can cause a morphological word to be interrupted by a clitic, thus violating the principle of lexical integrity (Bresnan and Mchombo 1995), **but on the basis of prosodic constraints only**. As the Pashto data shows, the clitics are equally subject to syntactic and prosodic constraints in sentences where the host is non-verbal ((1)). In the context of short, verb-based sentences as in (2)-(6) however, the prosodic component dominates the syntactic component — the clitics are placed after the foot that bears main stress.

I thus conclude that the requirement of the clitic to attach to a host is a strong requirement and that while syntax may not intervene in the word-internal structure after the morphological word is formed, prosody still has access to the internal structure of the prosodic word (e.g. the foot) and has the power of moving prosodically deficient items INTO the (prosodic) word.

Bibliography

Anderson, S. 2005. *Aspects of the Theory of Clitics*. Oxford University Press.

Bögel, T., M. Butt and S. Sulger. 2008. Urdu Ezafe and the Morphology-Syntax Interface. *Proceedings of LFG08*.

Bögel, T., M. Butt, R. Kaplan, T.H. King and J. Maxwell III. 2009. Prosodic Phonology in LFG: A New Proposal. *Proceedings of LFG09*.

Bresnan, J. and S. Mchombo 1995. The Lexical Integrity Principle: Evidence from Bantu. *Natural Language and Linguistic Theory* 13(2): 181-254.

Butt, M. 2007. The Role of Pronominal Suffixes in Punjabi. In Zaenen et al. (eds.), *Architectures, Rules and Preferences*. CSLI Publications: 341-368.

Butt, M. and T.H. King. 1998. Interfacing Phonology with LFG. *Proceedings of LFG98*.

Dost, A. 2005. A Domain-Based Approach to 2P Clitics in Pashto. *Proceedings of the Texas Linguistics Society IX Conference*.

Franks, S. and T.H. King. 2000. *A Handbook of Slavic Clitics*. Oxford University Press.

Halpern, A. 1995. *On the Placement and Morphology of Clitics*. CSLI Publications.

Kaisse, E. 1981. Separating Phonology from Syntax: A Reanalysis of Pashto Cliticization. *Journal of Linguistics* 17: 197-208.

Luís, A. and R. Otoguro. 2005. Morphological and Syntactic Well-formedness: The Case of European Portuguese Clitics. *Proceedings of LFG05*.

Mycock, L. 2006. *The Typology of Constituent Questions: A Lexical- Functional Grammar Analysis of wh-questions*. Ph.D.thesis, University of Manchester.

O'Connor, R. 2004. *Information Structure in Lexical-Functional Grammar: The Discourse-Prosody Correspondence in English and Serbo-Croation*. Ph.D.thesis, University of Manchester.

Roberts, T. 1997. The Optimal Second Position in Pashto. In G. Booij and J. van de Weijer (eds.), *Phonology in Progress: Progress in Phonology*. HIL Phonology Papers III. Holland Academic Graphics: 367-401.

Sadler, L. and A. Spencer. 2000. Workshop on Morphology and LFG. *Proceedings of LFG00*.

Selkirk, E. 1984. *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge: MIT Press.

Tegey, H. 1977. *The Grammar of Clitics: Evidence from Pashto and Other Languages*. Ph.D. thesis, University of Illinois.

van der Leeuw, F. 1997. *Clitics: Prosodic Studies*. Holland Academic Graphics.

Second position and the prosody-syntax interface

Tina Bögel University of Konstanz
Miriam Butt University of Konstanz
Ronald M. Kaplan Microsoft
Tracy Holloway King Microsoft
John T. Maxwell III PARC

In a recent paper Bögel et al. (2009) outlined a new architecture for modeling the interaction between prosody and syntax. Rather than the co-description approach first suggested by Butt and King (1998), Bögel et al. propose a pipeline arrangement of components in which prosodic information is developed in a module that operates independently of the syntax while still allowing for syntactic rules and preferences to be conditioned on prosodic boundaries and other features. This is made possible because the terminal string of the syntactic tree (the LFG c-structure) is a sequence of lexical formatives intermixed with elements inserted by the prosodic component. This architecture allows for misalignments between prosodic units and syntactic constituency because syntactic rules can simply ignore prosodic information that would otherwise disrupt syntactic patterns. But it also incorporates a Principle of Prosodic Preference that causes syntactic structures that do not coincide with prosodic boundaries to be dispreferred.

By way of illustration, Bögel et al. show how this new architecture provides for an insightful account of bracketing misalignments between the prosodic units of spoken English and syntactically motivated phrase structures. They also give an account of the bracketing misalignment exhibited by the Urdu clitic *ezafe*. The clitic *ezafe* attaches prosodically to a preceding host but it functions syntactically as an element of a following phrase (Bögel et al. 2008).

The present paper explores how the pipeline architecture can be applied to an additional class of syntax-prosody misalignments, the second position clitics that appear in many languages. Second position clitics have presented a challenge to many if not all theoretical frameworks, and there is a substantial literature on the subject (e.g. Halpern and Zwicky 1996, Franks and King 2000 and references therein). The crucial aspects of the problem, from an architectural point of view, are demonstrated by the following example from Serbian/Croation/Bosnian (SCB):

(1) [Taj joj ga je čovek] poklonio.
That her it AUX man presented

'That man presented her with it.' (Schütze 1994)

The clitic sequence *joj ga je* surfaces as an interruption between the demonstrative *Taj* and the head noun *čovek* of what would otherwise be an ordinary initial NP, and those clitics contribute feature and argument information not to the interpretation of the NP that they are contained within but to the clause enclosing that NP. Layered on top of these syntactic issues is the interaction with prosody: these clitics appear in the second position of a prosodic unit without regard to syntactic alignments. This is illustrated by the prosodic bracketing in (2):

(2) (((((Taj)$_w$ joj)$_{cl}$ ga)$_{cl}$ je)$_{cl}$ (čovek)$_w$)$_p$ (poklonio)$_p$
That her it AUX man presented

'That man presented her with it.'

In our account of this phenomenon the syntactic and prosodic components have a shared responsibility: the syntactic component deals with the clausal scope of functional information while the prosodic component provides for proper placement. In keeping with the Bögel et al. (2009) architecture, the components are coordinated by virtue of a limited amount of information visible on a commonly accessible string. In particular, the shared string carries a distinctive mark, notated as LB$_S$, by which the left-boundaries of syntactic clauses are made known to the prosody.

We start with the observation that clitics would naturally have clausal functional scope if they appeared as immediate daughters of the clause node in the syntactic c-structure. This can be achieved by

a simple extension of the c-structure rule that derives the normal patterns of clausal daughter sequences, as schematized in (3).

(3) $S \rightarrow LB_S \quad (\underset{\uparrow=\downarrow}{CCL}) \quad [...]$

Here, LB_S is the clausal left-boundary marker, the optional CCL covers the set of clitic sequences that can appear in second position, and [...] denotes the usual expansion of the clausal S category. The features of the clitics apply to the clausal f-structure by virtue of the ↑=↓ annotation.

We rely on the prosodic component to provide a mapping that correlates the clitics in the c-structure terminal string with their attested realization after the first prosodic word. The clitics are thus inverted in the prosodic representation so that they are realized in second position and can therefore attach to a prosodically acceptable host. Since the clitics are drawn from a given set of lexical/prosodic formatives and since they cluster according to a fixed set of patterns, we know that there are only a finite number of clitic sequences that are subject to the inversion mapping. This fact enables us to provide a characterization of the inversion mapping within the formal space of regular relations. Suppose CS in (4a) denotes the finite set of clitic sequences, the lexical/prosodic sequences that can be realizations of the CCL category (e.g. CS_1 for SCB might be the string *joj ga je*). Also let *W* stand for any prosodic word, presumably marked by distinctive prosodic-word brackets. Then the inversion mapping is the regular relation denoted by the expression (4b):

(4) a. $CS = \{CS_1, CS_2, \ldots, CS_n\}$

b. $[\ \overline{\Sigma^* LB_S\ CS\ \Sigma^*}\ (\ \bigcup_{CS_i}\ [\ LB_S\ \ CS_i{:}0\ \ W\ \ 0{:}CS_i\ ]\ )\ ]^*$

In this traditional notation (see Kaplan & Kay, 1994) the term Σ* stands for any number of prosodic items, and complementation is indicated by the overline. The overlined term thus describes the identity map on all strings that do not contain any of the clause-initial clusters. If a clause-initial cluster does appear, it must be treated by one of the expressions inside the optional union on the right. The term CS_i:0 indicates that there is nothing (0) on the prosodic side of the map corresponding to a particular cluster on the syntactic side. The following prosodic word *W* is unchanged in the mapping. After that word the term 0:CS_i indicates that that same i^{th} cluster appears on the prosodic side corresponding to nothing on the syntactic side. The effect is that strings with syntactically clause-initial clitic sequences are mapped to strings where those particular clusters appear on the other side of an adjacent word.

We thus provide a simple account of second position clitics as a minor variation within the general Bögel et al. (2009) architecture for the interface of prosody and syntax and without extending the formal power of LFG. The architecture allows syntactic and prosodic constraints to govern the distribution and interpretation of second-position clitics by applying in parallel to a limited amount of shared information.

Bibliography

Bögel, T., M. Butt and S. Sulger. 2008. Urdu Ezafe and the Morphology-Syntax Interface. *Proceedings of LFG08*.

Bögel, T., M. Butt, R. Kaplan, T.H. King and J. Maxwell III. 2009. Prosodic Phonology in LFG: A New Proposal. *Proceedings of LFG09*.

Butt, M. and T.H. King. 1998. Interfacing Phonology with LFG. *Proceedings of LFG98*.

Franks, S. and T.H. King. 2000. *A Handbook of Slavic Clitics*. Oxford University Press.

Halpern, A. and A. Zwicky 1996. *Approaching Second — Second Position Clitics and Related Phenomena*. CSLI Publications.

Kaplan, R. and M. Kay. 1994. Regular Models of Phonological Rule Systems. *Computational Linguistics* 20, 331-478.

Schütze, C. 1994. Serbo-Croatian Second Position Clitic Placement and the Phonology-Syntax Interface. In A. Carnie, H. Harley and T. Bures (eds.), *MIT Working Papers in Linguistics 21, Papers on phonology and morphology*: 373-473. (revised version).

Complements of adjectives: a diachronic approach

Kersti Börjars & Nigel Vincent
The University of Manchester

Can an adjective have an object? Traditional grammar says no (Huddleston & Pullum 2001: 527), and in similar vein Principles & Parameters Case Theory relies on the inability of nouns and adjectives to assign objective case to explain the distribution of English *of* (Chomsky 1981: 50-1). Compare too the theory of categories proposed by Jackendoff (1977), according to which adjectives are [–obj, –subj], thus contrasting with verbs: [+s, +o], nouns: [+s, –o] and prepositions:[–s, +o]. However, from the semantic point of view there is in fact good reason to expect that the range of complements available to adjectives is the same as for verbs, as evidenced by such near-synonymous pairs as *fear/be afraid of, love/be fond of, regret/be regretful of*. In practice grammarians have been happy to assign adjectives subcategorizations for COMP (*certain that S*), XCOMP (*keen to VP*), OBL (*similar to NP*). The question we ask in this paper therefore is whether an adjective can also take an OBJ.

Recent work in LFG (Mittendorf & Sadler 2008, Al Sharifi & Sadler 2009) argues that certain constructions in Welsh and Arabic involve an adjective taking an OBJ. These constructions have the general shape in (1), where A denotes a property of NP_1 with respect to NP_2, as in the archaic English *a girl fair of face*. Welsh does not show case on non-pronominal nouns, but in Arabic NP_2 occurs in the genitive, which might alternatively suggest that the GF in question is OBL, although this possibility is not explicitly considered.

(1) NP_1 [A NP_2]$_{AP}$.

From a different perspective, Maling (1983) and Platzack (1982a, b) note the existence of a category of so-called 'transitive adjectives' in the earlier stages of the Germanic languages. Thus in Old Swedish (examples from Platzack 1982b):

(2) a) Adjectives taking the dative case: *trygger* 'faithful', *hemul* 'familiar'
b) Adjectives taking the genitive case: *vis* 'sure', *forespar* 'farsighted', *vilder* 'stray'
c) Adjectives taking accusative case: *rätter* 'suitable', *godher* 'kind'

The situation in Old Swedish can be compared to that in Latin shown in (3), where partly similar patterns of case assignment are attested (examples and classification from Gildersleeve & Lodge 1895 §§359, 374, 390.3, 395):

(3) a) Adjectives of likeness, fitness, friendliness, nearness take the dative, e.g. *similis* 'like', *idoneus* 'suitable', *communis* 'common'
b) Adjectives of fullness, participation, power, knowledge, desire, etc take the genitive, e.g. *plenus* 'full', *compos* 'sound', *diligens* 'careful'
c) Adjectives of separation, origin and source take the ablative, e.g. *liber* 'free', i*mmunis* 'exempt', *natus* 'born'

On the basis of the cases assigned to the complements, we could argue that adjectives in Latin and earlier forms of Germanic are associated either with the function OBJ or with OBL. Pinkster (1990) shows that the case assigned to the adjectival complements in Latin can be considered the result of two conflicting principles, a Structural Principle, which assigns genitive case to complements of nouns and adjectives, and a Semantic Principle, namely that 'optional constituents with adjectives often have the same case that is used to express a comparable semantic relation on the sentence level' and that 'there is a certain regularity in the case marking of adjectives and semantically related verb'. Though the connection between the semantics of the predicate and the case of the complement is not so obvious in Old Swedish, it does play some role and we can assume that the semantic principles reflect a shared inheritance from the Indo-European case system. The second question then relates to historical change. Following Pinkster's distinction, we argue that those cases where the Structural Principle determine the case function as OBJ, whereas those where the Semantic Principle wins out represent OBLs. In languages which have case, both functions are represented by noun phrases. Given Maling's (1983:254) observation that 'there is something essentially correct about the idea that it is less natural for A and N to take NP complements than for V and P to do so', the issue is what happens to the case marked complements as case is lost in Romance and Germanic?

Some Germanic languages retain case and modern German for instance have adjectives with nominal complements in genitive and dative. For languages which lose case, there are essentially three options:

(i) change the syntactic status of the complement to PP as the general exponent of OBJ of nouns and adjectives and OBL in the language;
(ii) change the syntactic status of the head to P, a category that does accept noun phrase OBJ;
(iii) maintain the syntactic status of head and complement, but permit adjectives to take OBJ.

The modern Romance languages have taken the path described in (i). We argue that there are two reasons for this: on the one hand, there are no transitive adjectives because Latin had no adjectival accusatives, and on

the other the development was aided by the fact that there is a single preposition, *de/di* [< Lat *de* 'down, from, about'] which serves to mark dependents right across the nominal and adjectival domains. All previous genitives are therefore replaced by this one type of PP.

To a large extent, English has also followed the path in (i) and it is not likely to be a coincidence that it behaves like the Romance languages rather than like its sister Germanic languages; we attribute this to the influence of Norman French at the time when the Old English case system was being lost. However, as Maling (1983) argues convincingly, *worth* and *like* have been recategorised as prepositions and hence naturally take nominal complements. In short, they have followed the path in (ii). The only adjective in English that can take a nominal complement is *near* and this Maling reasonably describes as a historical remnant.

Swedish, Norwegian and Dutch, we argue, have taken the path in (iii). Exemplifying from Swedish, we show that the bold elements in (4) are truly adjectives; for instance in that they show agreement (4a, b) and the phrases they head distribute like adjective phrases, in that they can occur attributively and in the pre-verbal position (4c, d).

(4) a. Verkligheten blev oss **övermäktig**.
reality.COM.DEF become.PST us overpowering.COM
'Reality defeated us.'

b. Livet blev oss **övermäktigt**.
life.NT.DEF become.PST us overpowering.NT
'Life defeated us.'

c. den fienden **överlägsna** armén
the enemy.DEF superior army
'the army which was superior to the enemy'

d. Sitt samvete **kvitt** kunde han återgå till sitt brottsliga liv
POSS.REF conscious rid could he return to POSS.REFL criminal life
'Having got rid of his conscience, he could return to his criminal life.'

We argue that the OBL of the earlier stages of Swedish has developed into OBJ and that this change has been mediated by the existence of accusative OBJ in the earlier stages. The fact that Swedish has not developed one single preposition marking the complements of nouns and adjectives is also argued to have played a role in this development. It is also striking that Swedish developed novel adjective+OBJ combinations after case had been lost. Swedish (and other northern Germanic languages) then truly have nominal OBJ with adjectives.

In summary the present paper argues for the following conclusions:

a) adjectives may subcategorise for the full range of GFs, although OBJ is less widely attested, it comes about in Swedish because of conspiring historical facts;
b) adjectival OBJ can be realised as either the verbal structural case (accusative) or the nominal one (genitive);
c) when prepositional marking replaces morphological case, languages continue the preferences for structural vs semantic marking attested in the morphological stage, representing OBJ and OBL, respectively.

Selected references

Al Sharifi, Budour & Louisa Sadler (2009) The adjectival construct in Arabic. Miriam Butt & Tracy Holloway King (eds) *Proceedings of the LFG09 conference*. http://www-csli.stanford.edu/, pp. 26-43.

Gildersleeve, B.L. & Gonzalez Lodge (1895) *Gildersleeve's Latin Grammar*. 3rd ed. London: Macmillan.

Jackendoff, Ray S. (1977) *X-bar syntax*. Cambridge, Mass: MIT Press.

Maling, Joan. 1983. Transitive adjectives: a case of categorial reanalysis. In *Linguistic categories: auxiliaries and related puzzles*. *Vol 1*, eds. F. Heny & B. Richards 253–289.

Mittendorf, Ingo & Louisa Sadler (2008) NP would like to meet GF: a Welsh adjectival construction. Miriam Butt & Tracy Holloway King (eds) *Proceedings of the LFG08 Conference*. http://www-csli.stanford.edu/, pp. 373-393.

Pinkster, Harm (1990) *Latin syntax and semantics*. London: Routledge.

Platzack, Christer. 1982a. Transitive adjectives in Swedish: a phenomenon with implications for the theory of abstract Case. *Linguistic Review* 2:39-56.

Platzack, Christer. 1982b. Transitive adjectives in Old and Modern Swedish. In *Papers from the 5th International Conference on Historical Linguistics*, ed. Anders Ahlqvist, 273–282. Amsterdam: John Benjamins.

A movement paradox in Zapotec

George Aaron Broadwell
University at Albany, State University of New York

1. Introduction

Bresnan (2001) has drawn attention to *movement paradoxes* in syntactic theory. These are cases where the category of a constituent in a derived position differs from the category of the same constituent when it is *in situ*. The following pairs of English sentences show that a CP is disallowed as object of a preposition(1a, 2a), but grammatical as subject of the passive (1b, 2b) or when topicalized:

1a.) *This theory accounts for that languages are learnable
b.) That languages are learnable is accounted for by this theory.
2a.) *We talked about that he was sick for days.
b.) That he was sick, we talked about for days.

Theories which use movement to derive passive and topicalization must posit some category-change or other mechanism to account for the facts. In contrast, a theory with base-generated passive subjects and topics faces no such difficulties.

San Dionisio Ocotepec Zapotec (SDZ), an Otomanguean language of Mexico, also shows a movement paradox. For a number of verbs with semantics like 'cover/fill/be spread' which subcategorize for a Theme and a Location, the usual argument realization has two NPs after the V. But if the Theme argument is fronted (e.g as topic or interrogative focus), it optionally occurs with the preposition *cùn* 'with':

3a.) Rr-sè'w nìjs lòò yùù. 'Water covers the floor.'
hab-cover water on floor
b.) (Cùn) nìjs rr-sè'w lòò yùù 'Water [TOPIC] covers the floor.'
with water hab-cover on floor
c.)¿ Xhíí (cùn) rr-sè'w lòò yùù?' 'What covers the floor?'
what (with) hab-cover on floor

However, *cùn* 'with' may not appear if the Theme is *in situ*:

4a.) *Rr-sè'w cùn nìjs lòò yùù. (*intended:* Water covers the floor.)
hab-cover with water on floor
b.) *Rr-sè'w lòò yùù cùn nìjs.
hab-cover on floor with water

From the perspective of a movement-based theory, the initial PPs in (3b,c) are very difficult to explain, since the verbs of this class do not normally allow PP subjects.

2. A lexical rule of Zapotec

In order to account for these facts, I will argue that SDZ has a lexical rule which relates two forms of verbs in the 'cover/fill/be spread' class. In the usual realization, the Theme is realized as an NP subject, but the lexical rule produces an alternate argument realization where the Theme appears as PP subject.

The Zapotec alternation is not found with all Theme arguments, but only with those in the semantic class where the Theme is in complete contact with a location. I will refer to such Themes as Cover-Themes Using a slighly modified version of the formalism of Jackendoff (1990:160ff), the semantic representation of such verbs contains the following:

5.) $[_{\text{Event}}\ \text{INCH}\ [_{\text{State}}\ \text{BE}\ ([_{\text{Thing}}\ ]_i, [_{\text{Place}}\ \ \text{IN}_d/\text{ON}_d\ [_{\text{Thing}}\ ]_j\ ])\ ]\ ]$

IN_d and ON_d are distributive versions of the IN and ON locational predicates.

Studies of similar alternations in English (Jackendoff 1990, inter alia), Korean (Kim, Landau, and Phillips 1999), Hungarian (Ackerman 1992), and Modern Greek (Kordoni 2003). The generalization seems to be that Cover-Themes frequently alternate between a.) the morphosyntax characteristic of Themes and b.) the morphosyntax characteristic of Instruments, whether this is an adposition (Zapotec, English) or a case marker (Korean, Hungarian, Greek).

Thus for SDZ, we want the combination of lexical rule and Lexical Mapping Theory to produce two entries for verbs like *rr-sè'w* 'cover'. (6a) shows the Cover-Theme encoded as Theme; (6b) shows the Cover-Theme encoded as Instrument:

6a.) $[_{Event}$ INCH $[_{State}$ BE $([_{Thing}\ ]_i, [_{Place}$ IN_d/ON_d $[_{Thing}\ ]_j\ ])\ ]\ ]$
SUBJ OBL_{Loc}

b.) $[_{Event}$ INCH $[_{State}$ BE $([_{Place}$ WITH $[_{Thing}\ ]_i\ ], [_{Place}$ IN_d/ON_d $[_{Thing}\ ]_j\ ])\ ]\ ]$
SUBJ OBL_{Loc}

Regular rules of correspondence between semantic and syntactic categories result in Things being realized as NPs and Places being realized as PPs in SDZ.

3. SYNTACTIC REALIZATION

Assume that the phrase structure rules of SDZ include the following:

7.)

CP –>	(COMP)	(XP)	IP		
	↑=↓	(↑INTERROG)=↓ (↑GF)=↓	↑=↓		
IP –>	(Infl)	(XP)	S		
	↑=↓	(↑TOPIC)=↓ (↑GF)=↓	↑=↓		
S –>	V	(NP)	(NP)	(NP)	PP*
	↑=↓	(↑SUBJ)=↓	(↑OBJ)=↓	(↑OBJ_0)=↓	(↑OBL)=↓

The TOPIC and INTERROG functions may be assigned to any XP, but the SUBJ, OBJ, and OBJ-θ positions are restricted to NPs. Crucially, a postverbal PP cannot be a SUBJ.

We then predict that when a verb like *rrsè'w* 'cover' has a lexical entry like (6b), the PP subject will only be possible when it appears in position like [Spec, CP] or [Spec, IP] which allows the SUBJ function to be assigned to any XP.

4. ADDITIONAL IMPLICATIONS

A lexical rule that allows for alternation between NP and PP realizations of Cover-Themes is key to this analysis, but the correct formulation of such a rule must be sensitive to details of Lexical-Conceptual Representation. LFG analyses have occasionally used detailed lexical semantic representations of this kind (Butt 1995, Broadwell 1998, inter alia), but most work uses only labels like *Agent* and *Theme.* Simple labels for Thematic Roles are not sufficient to correctly identify the class of verbs that show this alternation in SDZ, since PP subjects are only available with Cover-Themes, and not for other Themes in the language. Thus an additional implication of the alternation seen in SDZ is that a successful account of the full range of lexical rules in a language requires more articulated lexical semantic representations of the sort found in Jackendoff (1983, 1990).

Two Approaches to Automatic Matching of Atomic Grammatical Features in LFG

Anton Bryl
CNGL, Dublin City University
abryl@computing.dcu.ie

Josef van Genabith
CNGL, Dublin City University
josef@computing.dcu.ie

The recent increase in attention to Lexical-Functional Grammars in syntax-based statistical machine translation (SMT) [1, 2] poses new problems for processing richly annotated data. Alignment is a core issue in machine translation. In our work we focus on deepening the automatic cross-language structure alignment by adding the possibility to effectively align not only words, but also atomic f-structure features. Though sets of atomic features (such as case, number, etc.) differ for different languages, they are far from being disjoint. A number of features, such as number or case are shared between many languages. It is common practice to hardcode these similarities in the grammars; that is, to give the same names to the same or similar linguistic properties in the grammars for different languages. However, when one wants to make use of correspondences between such features in the framework of a language-agnostic syntax-based SMT system, such "feature name alignment" between source and target grammars cannot simply be taken for granted. Moreover, the degree of correspondence may differ from feature to feature across grammars. This motivates the need for an automatic way to judge correspondences between atomic features in f-structure representations for arbitrary language pairs.

We show that, provided we have parsers for two languages and a parallel corpus for these languages, it is possible to automatically identify at least part of the correspondences between atomic-valued grammatical features. Once identified, these feature pairs can further be used to improve the coverage of transfer-based machine translation. For example, if the algorithm identifies that NUM in English and NUM in German generally co-vary (we presume again, that we do not use any prior knowledge about this correspondence), then we can safely induce transfer rules (from aligned parsed bitext corpora) which abstract over the number feature, providing an effective back-off to more specific transfer rules.

Another application in which automatic grammatical feature matching is potentially useful is parallel grammar design. More specifically, feature matching is able to provide empirical evidence of the similarity between features of different languages, thus helping to determine whether they are to be treated as the same or different features.

We define and evaluate two methods of grammatical feature matching.

Method 1. The idea behind the first method is that if a feature **A** in one language corresponds to feature **B** in another language, then a change in the value of **A** in the source language (SL) frequently corresponds to a change

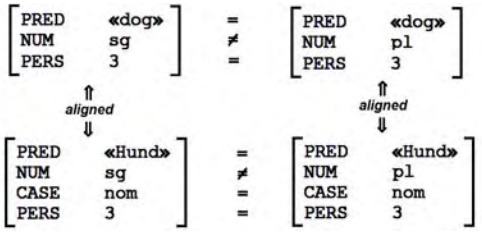


Figure 1: Simultaneous change of the values of NUM in parallel data. Finding such a situation, the algorithm increases the probability counter for the (Eng.NUM ⇒ Ger.NUM) correspondence.

German-to-English			English-to-German		
Pair	N_{it}	$\frac{C_{A,B}}{C_A}$	*Pair*	N_{it}	$\frac{C_{A,B}}{C_A}$
(NUM ⇒ NUM)	1	0.88	(NUM ⇒ NUM)	1	0.67
(TNS-ASP ⇒ TNS-ASP)	1	0.70	(TNS-ASP ⇒ TNS-ASP)	1	0.81
(CLAUSE-TYPE ⇒ CLAUSE-TYPE)	1	0.62	(CASE ⇒ CASE)	1	0.86
(CASE ⇒ CASE)	1	0.51	(DEGREE ⇒ DEGREE)	1	0.98
(ATYPE ⇒ ATYPE)	1	0.79	(PASSIVE ⇒ PASSIVE)	2	0.55
(COMP-FORM ⇒ COMP-FORM)	2	0.92			
(PASSIVE ⇒ PASSIVE)	2	0.64			

Table 1: Experimental results for Method 1. N_{it} is the number of iteration on which the pair emerged. $\frac{C_{A,B}}{C_A}$ is the normalized score (see algorithm).

Pair	Inc. of prediction accuracy	Pair	Inc. of prediction accuracy
NUM ⇒ NTYPE	0.17	NUM ⇒ PERS	0.25
NUM ⇒ NUM	0.35	NUM ⇒ CASE	0.15

Table 2: A part of experimental results for Method 2 (German-to-English), showing the increase of prediction accuracy over the pick-most-frequent baseline. Apart from the NUM, which is the correct match, CASE, PERS, NTYPE also gained high scores. This is due to their frequent co-presence with NUM in f-structures, which results in accurate prediction of the *feature absent* special value. However, the correct match clearly outscores them. All the other features got 0 score when matched with NUM, and are not included in the table for the sake of space.

in the value of **B** in the aligned translation. More precisely, the method identifies pairs of sub-f-structures in the SL data, which differ only in one atomic feature value, and then checks the difference in the aligned target language sub-f-structures (see Figure 1). The accumulated data from all such pairs is then used to find the best matches between the features. Complex PRED-less features like TNS-ASP are treated as atomic and considered equal if all their child attributes are equal, and unequal otherwise.

Method 2. The second method makes use of the mutual predictability of features of the two languages. For each possible pair (**Lang1.A**, **Lang2.B**) we calculate the best possible accuracy of prediction of the value of **Lang2.B** in the TL structure by the value of **Lang1.A** in the aligned SL f-structure. The accuracy of pick-most-frequent baseline is then subtracted from this value. The resulting value, that is the increase in prediction accuracy over the baseline, is used as a measure of similarity between the features. The absence of a certain feature in a structure is considered a special *feature absent* value of this feature.

Both methods are evaluated experimentally on 219,667 sentences of parced Europarl [3] German-English data and show promising results. The results for two iterations of Method 1 are presented in Table 1. Table 2 contains some example numbers for Method 2.

References

[1] E. Avramidis and J. Kuhn. Exploiting XLE's finite state interface in LFG-based statistical machine translation. In *Proceedings of LFG'09*, 2009.

[2] Y. Graham and J. van Genabith. An open source rule induction tool for transfer-based SMT. *The Prague Bulletin of Mathematical Linguistics*, Special Issue: Open Source Tools for Machine Translation, 91:37–46, 2009.

[3] P. Koehn. Europarl: A parallel corpus for statistical machine translation. In *MT Summit X*, pages 79–86, 2005.

Case Attraction in Modern Greek Free Relative Clauses

Kakia Chatsiou

Case matching effects in relative clauses occurs when the case of the relative pronoun introducing relative clauses matches the case requirements of the verb of the matrix clause and not those of the relative clause verb. Nominal Modern Greek Relative Clauses (henceforth FRCs), such as *ópjos*-FRCs in (1), display matching effects, as the free relative pronoun usually takes matrix rather than subordinate case:

(1) *Voithises* *ópjos / ópjon irthe.
helped.$_{2SG}$ whoever.$_{MSG\cdot NOM}$ whoever.$_{MSG\cdot ACC}$ came.$_{3SG}$
'You helped whoever came.'

When FRCs are fronted, however, case matching is not required and the free relative pronoun can receive either matrix or subordinate case as in (2a), an observation referred in the classical literature as *forward attraction of case* (Tzartzanos, 1996: 169). The presence of a doubling clitic is necessary, demonstrated by the unavailability of the nominative case in (2b):

(2) a. Ópjos / Ópjon irthe, ***ton*** *voithises*
whoever.$_{MSG\cdot NOM}$ whoever.$_{MSG\cdot ACC}$ came.$_{3SG}$ him.$_{MSG\cdot ACC}$ helped.$_{2SG}$
'Whoever came, you helped him.'

b. *Ópjos / Ópjon irthe, ___ *voithises*
whoever.$_{MSG\cdot NOM}$ whoever.$_{MSG\cdot ACC}$ came.$_{3SG}$ helped.$_{2SG}$
'Whoever came, you helped him.'

Case attraction seems quite robust and independent from the thematic role of the free relative pronoun in the matrix and the FRC, as illustrated in (3):

(3) a. *Voithises* ópjon / *ópju i Maria tu edose ena doro
helped.$_{2SG}$ whoever.$_{MSG\cdot ACC}$ whoever.$_{MSG\cdot GEN}$ the Mary him.$_{MSG\cdot GEN}$ gave.$_{3SG}$ a gift
'You helped whoever Mary gave a gift to'

b. ópjon / ópju i Maria tu edose ena doro,
whoever.$_{MSG\cdot ACC}$ whoever.$_{MSG\cdot GEN}$ the Mary him.$_{MSG\cdot GEN}$ gave.$_{3SG}$ a gift
ton *voithises.*
him.$_{MSG\cdot ACC}$ helped.$_{2SG}$
'Whoever Mary gave a gift to, you helped him '

In fronted FRCs, the free relative pronoun alternatively fulfils the case requirements of the matrix clause or the FRC. This could pose a challenge for unification-based frameworks like LFG, since in certain environments, the value of a feature of a single f-structure (the CASE feature of the free relative pronoun f-structure) can alternatively realise the CASE of the FRC or the matrix clause grammatical function.

Previous LFG analyses will be discussed and it will be shown that the Modern Greek data cannot be accommodated using proposals previously put forward for case mismatching phenomena in other languages, such as indeterminacy (Dalrymple & Kaplan, 2000), underspecification (Dalrymple, King & Sadler, 2009) or lexical sharing (Wescoat, 2005).

I propose an LFG analysis which treats the Free relative pronoun as the head of the FRC's f-structure and the rest of the relative clause as an adjunct to the free relative pronoun, a treatment similar to that of restrictive and non-restrictive relative clauses. Building on Echevarría & Ralli's (2000) observations on the role of the doubling clitic in facilitating case alternation in clitic left dislocating constructions, I propose an alternative solution that uses anaphoric binding and relies on the use of an additional feature on the f-structures of the doubling clitic, the free relative pronoun

and the within FRC thematic role. This feature is used to constraint case alternation on the relative pronoun introducing a fronted FRC and to ensure either matrix or FRC case is allowed.

Examples of the f-stuctures of a non-topicalised (1) and a topicalised FRC (2a) are shown in (4) and (5):

(4)
$$\begin{bmatrix} \text{PRED} & \text{`helped}\langle \text{SUBJ, OBJ} \rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`pro'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`pro'} \\ \text{CASE} & \text{acc} \\ \text{INDEX} & i \\ \text{PRONFORM} & \text{opjos} \\ \text{ADJ} & \left\{ \begin{bmatrix} \text{PRED} & \text{'came}\langle \text{SUBJ} \rangle\text{'} \\ \text{SUBJ} & \boxed{1}\begin{bmatrix} \text{PRED} & \text{'pro'} \\ \text{INDEX} & i \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{TOPIC} & \boxed{1} \\ \text{RELPRO} & \boxed{1} \end{bmatrix} \right\} \end{bmatrix} \end{bmatrix}$$
f-structure for example (1)

(5)
$$\begin{bmatrix} \text{PRED} & \text{`helped}\langle \text{SUBJ, OBJ} \rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`pro'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`pro'} \\ \text{CASE} & \text{acc} \\ \text{INDEX} & i \\ \text{PRONTYPE} & \text{clitic} \end{bmatrix} \\ \text{TOPIC} & \begin{bmatrix} \text{PRED} & \text{'pro'} \\ \text{CASE} & \text{acc} \\ \text{INDEX} & i \\ \text{PRONFORM} & \text{opjos} \\ \text{ADJ} & \left\{ \begin{bmatrix} \text{PRED} & \text{'came}\langle \text{SUBJ} \rangle\text{'} \\ \text{SUBJ} & \boxed{2}\begin{bmatrix} \text{PRED} & \text{'pro'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{TOPIC} & \boxed{2} \\ \text{RELPRO} & \boxed{2} \end{bmatrix} \right\} \end{bmatrix} \end{bmatrix}$$
f-structure for example (2a)

References

Dalrymple, D. & Kaplan, R. (2000). "A set based approach to feature resolution". In Butt, M and King, T. *Proceedings of the LFG97 Conference*. CSLI publications, 1-15.

Dalrymple, M., King, T. & Sadler, L. (2009). "Indeterminacy by Underspecification." *Journal of Linguistics* **45**/1.

Echevarría, M & Ralli,A. (2000). "Case Mismathces in Greek: Evidence for the autonomy of morphology". In *Acta Linguistica Hungarica*, **47** (1-4), 179-203.

Tzartzanos, A. (1996). *Neoelliniki Syntaxis (Modern Greek Syntax).* Thessaloniki: Kiriakides.

Wescoat, M. (2005). "English nonsyllabic auxiliary contractions: An analysis in LFG with lexical sharing". In M. Butt and T. H. King, eds., *Proceedings of the LFG05 Conference*, 468Ű486. Stanford: CSLI.

Less-beaten paths from pronoun to agreement: The case of Uralic objective conjugations

Elizabeth Coppock and Stephen Wechsler, The University of Texas at Austin

The transition from pronoun to agreement marker is standardly characterized as a loss of the referential property of the pronoun, with a retention of person, number, and gender (ϕ) features (Bresnan and Mchombo, 1987). In LFG, this can be modelled as the loss of a PRED 'pro' equation in the lexical specification of the affix. But the path from pronoun to agreement marker does not always follow this simple scheme. For example, as Bresnan (2001, 146) notes, "finer transition states" are possible, in which affixes retain sensitivity to properties like definiteness and animacy. We argue that the Hungarian objective conjugation is an agreement marker historically derived from a pronoun through an even more complex and varied set of transitions, which nevertheless can be modelled naturally in LFG as the loss of lexical specifications on the affix.

Hungarian has two subject-verb agreement paradigms, the *subjective* and *objective conjugations,* whose distribution depends on the presence of a 'definite' object (they are glossed as INDEF and DEF, respectively):

(1) a. Vár-ok
wait-1SG.INDEF
'I'm waiting'

b. Lát-ok egy madar-at
see-1.SG.INDEF a bird-ACC
'I see a bird'

c. Lát-om a madar-at
see-1.SG.DEF the bird-ACC
'I see the bird'

The objective conjugation is an agreement marker rather than an incorporated object pronoun (Coppock and Wechsler in prep.). But it is unlike normal cases of pronoun-derived agreement in that it cross-references the definiteness, rather than the ϕ-features, of the object. We argue that, seen from the perspective of its historical provenance, the objective conjugation may nevertheless be understood as a variation on the more familiar cases of ϕ-feature agreement. Sensitivity to specificity or definiteness can be lost before ϕ-features in the transition from pronoun to affix; Bininj Gun-Wok is an example of a language where ϕ-features remain but specificity requirements do not (Evans, 1999). In Hungarian, we suggest that feature loss occurred in the opposite order: ϕ-features were (almost completely) lost, but sensitivity to specificity, definiteness, or topicality was retained, and this property was reanalyzed as formal definiteness.

Following several (but not all) other Hungarian linguists, we propose that the objective conjugation endings derive historically from a third person singular object marker (OM) agglutinated to a subject marker (SM). Support for this view comes from the fact that phonologically, many of the objective conjugation endings consist of a glide or similar sound followed by an element that is similar to the corresponding subjective conjugation ending, as shown in (2) (the glide element is indicated with a box).

(2)

	INDEF	DEF
1SG	-ok/-ek/-ök	-om/-em/-öm
2SG	-(a)sz/-(e)sz or -ol/-el/-öl	-od/-ed/-öd
3SG	∅	-ja/-i
1PL	-unk/-ünk	-juk/-jük
2PL	-(o)tok/-(e)tek/-(ö)tök	-já tok/-i tek
3PL	-(a)nak/-(e)nek	-ják/-ik

Further evidence for this view comes from (i) the fact that the verbs in other Uralic languages, including Ostyak and Mordva, follow a V+OM+SM template and (ii) the special *-lak/-lek* suffix, used for first person singular subjects and second person objects, which can be analysed as second person *l* + 1SG *k*.

Furthermore, the Hungarian objective conjugation is not totally insensitive to ϕ features; first and second person (non-reflexive) objects trigger the subjective conjugation, unlike third person objects:

(3) a. Lát-ják őt/őket
see-3.PL.DEF it/them
'They see it/them'

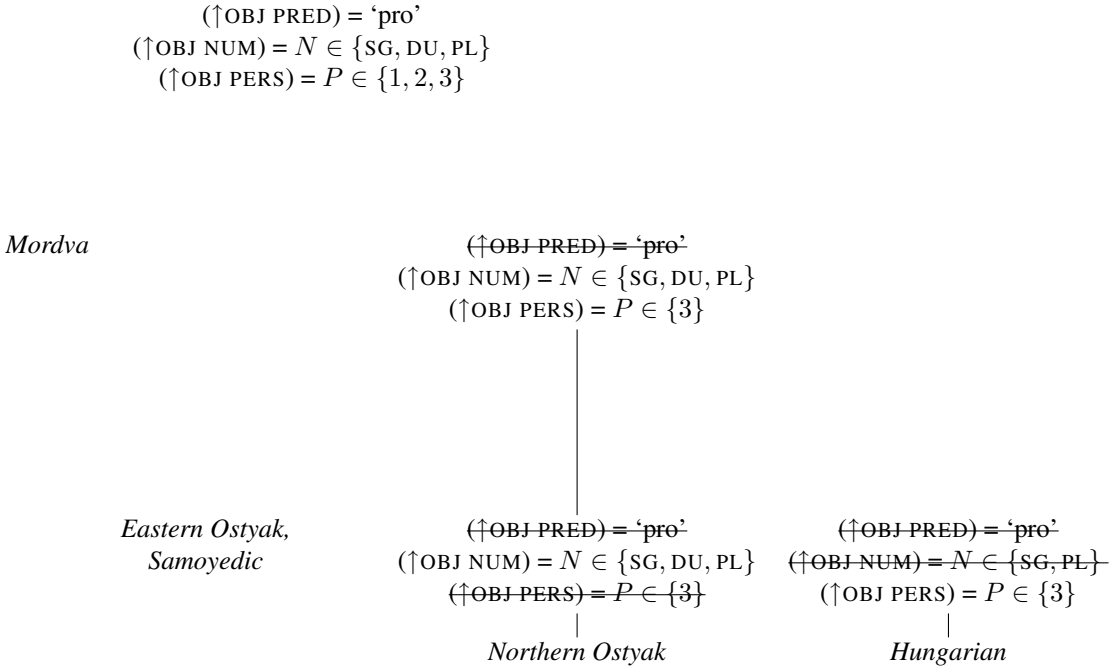
b. Lát-nak engem/téged/minket
see-3PL.INDEF me/you/us
'They see me/you/us'

How Hungarian verbal morphology managed to retain sensitivity to object definiteness while losing (most of) its object ϕ-features, we argue, is as follows: First, due to factors related to topicality, object pronouns (which display both person and number) were incorporated only in third person, so person *distinctions* were lost, and only number distinctions remained, as in the Eastern Ostyak (Gulya, 1966) and Samoyedic (Honti, 1984), both Uralic. Then number distinctions were lost within third person. This yielded a language with an agreement system expressing exactly *one* ϕ-feature: third person. This feature in a sense "carried" the sensitivity of the phenomenon to topicality, a sensitivity

that is shared by Northern Ostyak (Nikolaeva, 1999). The absence of ϕ-distinctions, along with the sensitivity to topicality, led the marker to be reanalyzed as a 'definiteness' marker.

The proposed historical development is given in (4). We assume that in all of the languages in question, the verb has the template V+OM+SM, and the object marker contributes features to the OBJ *f*-structure. The annotations contributed by the object marker are given in the nodes of the historical tree. Strike-throughs indicate that the annotation has been lost; we hypothesize that the historical development proceeds in part by removal of these constraints. The inventory of object markers is indicated by the sets of features; loss of inventory is another proposed historical mechanism.

(4)



To account for the absence of the object marker in first and second person singular (cf. (2)), we argue that the objective conjugation endings in those cells were replaced by possessive markers, due to the identity in form between the third person singular objective conjugation ending and the third person singular possessive marker (along with the presence of possessive markers in other arenas of verbal inflection). Unlike competing theories of the Hungarian objective conjugation (reviewed and contributed to by Havas (2004)), this theory accounts for all of the quirks of its distribution and morphology and sheds light on other Uralic languages, using simple and independently grounded historical mechanisms (feature loss, analogy).

We conclude that the provenance of this phenomenon bespeaks a richer array of historical possibilities for the feature loss that leads from pronoun to agreement: agreement markers can end up deficient in ϕ-features through feature loss akin to the process by which the referential property of pronouns is lost, and sensitivity to properties like definiteness and animacy can "survive" on a single ϕ-feature.

References

Bresnan, J. (2001). *Lexical-Functional Syntax*. Blackwell, Malden, Mass.

Bresnan, J. and Mchombo, S. (1987). Topic, pronoun, and agreement in Chicheŵa. *Language*, 63:741–782.

Evans, N. (1999). Why argument affixes in polysynthetic languages are not pronouns: Evidence from Bininj Gun-Wok. *Srachtypologie and Universalienforschung*, 52:255–281.

Gulya, J. (1966). *Eastern Ostyak Chrestomathy*. University of Indiana Press, Bloomington.

Havas, F. (2004). Objective conjugation and medialisation. *Acta Linguistica Hungarica*, 51:95–141.

Honti, L. (1984). *Chrestomathia Ostiacica*. Tankönyvkiadó, Budapest.

Nikolaeva, I. (1999). Object agreement, grammatical relations, and information structure. *Studies in Language*, 23:331–376.

Agreement patterns and coordination in Lexical Functional Grammar

Mary Dalrymple and Bozhil Hristov
Centre for Linguistics and Philology
University of Oxford

Coordination and its interactions with agreement have been a focus of research in LFG over the past decade (Dalrymple & Kaplan, 2000; King & Dalrymple, 2004; Dalrymple et al., 2006; Sadler, 1999, 2003, 2006; Arnold et al., 2007), though an account that captures the full range of agreement patterns in an elegant manner has proved elusive. Many previous proposals account for patterns of feature resolution but do not extend to single-conjunct agreement (Dalrymple & Kaplan, 2000; King & Dalrymple, 2004; Dalrymple et al., 2006; Sadler, 2006). Other proposals address single-conjunct agreement, but provide an account of standard resolution patterns that is less than satisfying. We provide a means of stating a typology of agreement patterns that handles resolution and single-conjunct agreement as well as agreement requirements that apply in an across-the-board fashion to all of the conjuncts of a coordinate phrase. We rely on the standard distinction between CONCORD and INDEX features, treating them both as syntactic features represented at f-structure (Wechsler & Zlatić, 2003). We follow King & Dalrymple (2004) in treating INDEX in coordinate structures as a nondistributive (resolving) feature and CONCORD as distributive. Previous literature has not been explicit about what features are active when a target agrees with only one conjunct.

A number of Serbian/Croatian nouns have mismatched CONCORD and INDEX features: for example, *deca* 'children' and *unučad* 'grandchildren' have FemSg CONCORD but NeutPl INDEX. We can use Serbian/Croatian nouns with mismatched CONCORD and INDEX features to demonstrate the existence of at least the following agreement patterns.

• Agreement with resolved INDEX features:

(1) Deca i unučad koja/koji su došl-a/došl-i su gladn-a/gladn-i
children and grandchildren who.NeutPl/who.MascPl AUX.3PL come-NeutPl/come-MascPl be.3PL hungry-NeutPl/hungry-MascPl
'The children and grandchildren who came are hungry.' [web/informant]

The MascPl agreement option on the relative pronoun, the verb and the adjective must be resolved agreement over the NeutPl INDEX features of the conjuncts, since resolved agreement involving FemSg conjuncts would give FemPl, and everything else, including neuter, resolves to MascPl. NeutPl agreement is closest-conjunct agreement, also illustrated in (2).

• Agreement with INDEX features of the closest conjunct:

(2) Tinejdžeri i deca koja preglasno i prečesto slušaju muziku ...
Teenagers.MascPl and children who.NeutPl too.loudly and too.often listen.to music ...
'Teenagers and children who listen to music too loudly and too often...' [web]

'Teenagers' has MascPl CONCORD and INDEX. The relative pronoun shows closest-conjunct agreement with the NeutPl INDEX features of the closest conjunct. Closest-conjunct CONCORD agreement would be FemSg, and resolved agreement would be MascPl.

• Agreement with the concord features of each conjunct (distributive concord agreement):

(3) porodicu i decu koju imate u Ausraliji
family.FemSgAcc and children.FemSgAcc who.FemSgAcc you.have in Australia
'family and children whom you have in Australia' [web]

'Family' has FemSg CONCORD and INDEX. The relative pronoun shows FemSg agreement with both conjuncts; resolved agreement would be FemPl for the CONCORD feature, and MascPl for INDEX. All examples of this type which we have collected involve uniform CONCORD features of the conjuncts.

• Agreement with the concord features of the closest conjunct:

(4) sve njegove molbe i uveravanja ni-su pomagali ništa
all.FemPl his.FemPl prayers.FemPl and assurances.NeutPl Neg-PL helped.MascPl nothing
'All his prayers and assurances did not help at all.'
(Corbett 1979, 206; Corbett 1991, 283)

Although there are no concord/index mismatches in this example, we argue on the basis of agreement patterns with mismatched nouns that attributive agreement ('all', 'his') is with the concord features of the initial conjunct. Following Kuhn & Sadler (2007), we propose to handle these agreement patterns by defining functional metavariables to allow reference to peripheral conjuncts in a coordinate phrase. We adopt Kuhn & Sadler's notation f_L and f_R, but define them differently:

(5) $f_{(L)} \equiv f \quad \in^* \quad \neg[(\leftarrow \in) <_f \rightarrow]$

$f_{(L)}$ is an f-structure possibly embedded within f as a conjunct in a coordinate set. If $f_{(L)}$ is embedded as a member of f, it must be the leftmost member: this is accomplished by the off-path constraint, which states that there may not be any (other) members of the coordinate structure that f-precede $f_{(L)}$.

The definition of $f_{(R)}$ is similar except for reversed f-precedence requirement on the other conjuncts. The definitions of f_L and f_R add the requirement that the f-structure that is the target of agreement must not itself be a coordinate structure:

(6) $f_L \equiv f \quad \in^* \quad : \neg(f_L \ \in) \quad \neg[(\leftarrow \in) <_f \rightarrow]$

Like Kuhn & Sadler, we encode agreement requirements lexically. (7) gives the lexical entry for the Serbian/Croatian possessive determiner *njegove* 'his', which shows obligatory closest-conjunct CONCORD agreement, agreeing either with a noncoordinated noun or with the left conjunct of a coordinate structure:

(7) njegove 'his': $(\uparrow_L$ CONCORD GEND) = F
$(\uparrow_L$ CONCORD NUM) = Pl

Our approach contrasts with Kuhn & Sadler (2007), who require features to be assigned to exactly one classification, and to behave uniformly as that classification requires. The main difficulty with this proposal is the existence of optional single-conjunct agreement. A classification of features entails that a feature will always behave in a certain way: always requiring resolved agreement, for example, or always requiring single-conjunct agreement. However, either closest-conjunct or resolved INDEX agreement is possible for example (1), showing that the INDEX feature can participate in both single-conjunct agreement and resolved agreement in the same construction, which is unexpected on Kuhn & Sadler's view.

Arnold, Doug, Louisa Sadler, & Aline Villavicencio. 2007. Portuguese: Corpora, coordination and agreement. In Sam Featherston & Wolfgang Sternefeld (editors), *Roots: Linguistics in Search of its Evidential Base*, Mouton.

Corbett, Greville G. 1979. The agreement hierarchy. *Journal of Linguistics* 15, pp. 203–224.

Corbett, Greville G. 1991. *Gender*. Cambridge, UK: Cambridge University Press.

Dalrymple, Mary, Helge Dyvik, & Louisa Sadler. 2006. Gender resolution in coordination and disjunction. Abstract in *Proceedings of LFG2006*.

Dalrymple, Mary & Ronald M. Kaplan. 2000. Feature indeterminacy and feature resolution. *Language* 76(4).

King, Tracy Holloway & Mary Dalrymple. 2004. Determiner agreement and noun conjunction. *Journal of Linguistics* 40(1), pp. 69–104.

Kuhn, Jonas & Louisa Sadler. 2007. Single conjunct agreement and the formal treatment of coordination in LFG. In *Proceedings of LFG2007*.

Sadler, Louisa. 1999. Non-distributive features and coordination in Welsh. In *Proceedings of LFG99*.

Sadler, Louisa. 2003. Coordination and asymmetric agreement in Welsh. In Miriam Butt & Tracy Holloway King (editors), *Nominals: Inside and Out*, CSLI.

Sadler, Louisa. 2006. Gender resolution in Rumanian. In Miriam Butt, Mary Dalrymple, & Tracy Holloway King (editors), *Intelligent Linguistic Architectures: Variations on themes by Ronald M. Kaplan*, CSLI.

Wechsler, Stephen & Larisa Zlatić. 2003. *The Many Faces of Agreement*. Stanford, CA: CSLI Publications.

SYNTACTIC STRUCTURE OF K'ICHEE' MAYAN

Lachlan Duncan

The majority of Mayan language research, including K'ichee'an, was effected from the sixties to the late nineteen eighties. The research was almost exclusively descriptive in nature, with the literature concentrating on phonology, historical linguistics, or epigraphy. Pedagogical grammars typically covered a broad spectrum of grammatical description, concentrating mainly on phonology and morphology, some pragmatics, and elementary morphosyntax and clause structure. At best, analysis was speculative and pre-theoretical. Since the early 1990s, however, more contemporary analytical research based on formal theories of syntax have surfaced (Aissen 1987, 1992, 1996, 1999a,b, 2000; Broadwell 2000, 2001, 2005; Woolford 1991, 1997). My dissertation adds to this list, using theoretical analyses based on the formal architecture of OT-LFG. Previous proposals on the syntactic structures of the sister K'ichee'an languages of Tz'utujiil (Aissen 1992; King 1995) and Kaqchikel (Broadwell 2000), and K'ichee' itself (Larsen 1988) are reviewed and their proposals analyzed for comparative purposes. The abundant interlinear-glossed data include cited material drawn from a variety of published sources. Nevertheless, the data on which the analyses are based are taken overwhelmingly from the author's fieldwork, elicited from first language K'ichee' Mayan speakers. Hence a substantial resource of never-before-seen data of an endangered language is now made available. In addition, the official Mayan language orthography is used exclusively in the dissertation.This has facilitated tracking Mayan language because much of the cited Mayan data in the dissertation was published previous to standardization, and employed a variety of inconsistent and confusing orthographies.

In brief, K'ichee' is an ergative-absolutive, pro-drop, head-marking language that marks agreement on the finite verb with ergative and absolutive agreement markers. Possessed nouns agree in person and number with their possessors. Complex prepositions agree in person and number with their object complements. The dissertation begins with the nominals, examining, for example, the bi-determiner DP, which I argue, is a type of demonstrative, with no pragmatics involved, as is usually claimed. The nominals use three distinct forms of pluralization, one morphological, the other two free morphemes, and are analyzed accordingly. After a literature review, I consider in detail the predicate-initial clause, in effect expanding on Aissen (1992). But contra Aissen's VP proposal, I argue for a predicate-initial, non-endocentric S(ENTENCE), with canonical word order as [$_{\mathrm{S}}$ V^0 XP*]. Incontrovertible evidence is presented using finite predicates that conclusively proves that the VP is not universal, as the derivational generativists assume from first principles. Argument word order is determined by lexical properties like animacy, definiteness, and phrasal weight. I include an OT-LFG analysis for predicate-initiality in K'ichee'. OT-LFG remains indispensable on this account because phrase-structure rules or generalized linear precedence rules are insufficiently fine-grained to capture the natural variation of argument distribution in the predicate-initial clause.

After reviewing sentential topics, I argue contra Aissen (1992) that the so-called external topic adjoins to CP, while the internal topic is located at the left edge of Spec,IP. I contend that the i-topic position is always blocked in non-verbal predicates. If any one of the focus positions in Spec,IP is occupied, the i-topic position, excluding relative pronouns, is blocked in the finite predicates as well.

Two types of predicates occur in K'ichee', finite predicates and non-verbal predicates, the latter of which include the perfect aspect as a special case. Aspect, not tense, is morphologically-marked on the verb, and is the foremost identifier of non-perfect finite predicates. Hitherto little-known structural correspondences are identified between non-verbal predicates and other linguistic constructions, in particular the non-finite

perfect aspect and the various mix of actor focus constructions. K'ichee' evidences five types of non-verbal predicates. The nominal and adjectival predicates, which are clearly zero-copula, and the existential, possessive, and locational predicates, which require the predicating non-verbal copula *k'oolik* 'exist.' I contest the single-tier analysis as the default for verbless clauses (Nordlinger and Sadler 2007), preferring instead the double-tier analysis using 'null be' (Dalrymple et al. 2004) for zero-copula and VCop for *k'oolik*. Following Butt et al. (1999), I reject the generalizing principle that adjectives and nominals can also function as clausal heads, which select for subjects, thus requiring additional equations in their lexical entries. Following Attia (2008) and Rosén (1996), I argue that agreement should be specified in phrase-structure rules, not in lexical entries. I depart, however, from the above approaches in rejecting Butt et al.'s (1999) closed grammatical function (GF) PREDLINK. Because K'ichee's non-verbal predicates are morphologically-marked with non-bound intransitive absolutive agreement markers, the non-verbals are thus intransitive requiring SUBJ-only f-structure semantic forms. The PREDLINK argument is thus infelicitous.

I propose instead an intermediate argument-non-argument category called function thematic (FN_Θ), a GF that is thematically-selected for but is *not* syntactically-selected for. In a binary feature array, FN_Θ fills an obvious gap in a two-feature, four way division: arguments are [+syntactic, +thematic], non-arguments are [–syntactic, –thematic], expletive subjects/objects of raising verbs are [+syntactic, –thematic], and FN_Θ is [–syntactic, +thematic]. Hence FN_Θ is not part of the f-structure's semantic form but is listed as a thematic role in a-structure. F-structure's completeness requirement is thereby satisfied, although accounting for coherency is somewhat more involved. Additional candidates for FN_Θ include head-adjoined incorporated nouns of detransitivized periphrastic noun incorporation constructions, nominal complements in copula inversion constructions in the Romance languages (cf. Alsina 2007), and even Rákosi's (2006) thematic adjunct (ADJ_Θ), which describes circumstantials in Hungarian. Because the binary argument-non-argument distinction (Bresnan 1982) is axiomatic in the strategic design of LFG, expanding the inventory of GFs will undoubtedly raise some objections. Notwithstanding this, I maintain that FN_Θ is well-founded and empirically motivated.

Contrastives, interrogatives, and negatives are also considered, and in each of these—except clausal negation—the argument is always focused. Focus is located in Spec,IP, and ordered such that INTFOC $\prec$ CONFOC $\prec$ NEGFOC. Crucially in all cases focused arguments represent non-verbal predicates. In that sense, K'ichee' clauses with focused arguments resemble English clefts without relative pronouns. The clause's primary predicate is determined on the focused argument's grammatical category. If an object, the transitive predicate remains unaltered. But if a subject, the actor focus construction is required. Actor focus is an intransitive predicate with an obligatorily preverbal focused actor and with agreement determined according to argument salience on the participant hierarchy. Although a morphological intransitive, the actor focus verb obligatorily retains both semantic roles of the transitive verb. Nevertheless the syntactic-thematic mismatch, I argue, can be accounted for by positing FN_Θ as the non-actor argument.

What triggers actor focus? Again all focused arguments in Spec,IP manifest as non-verbal predicates. As such, focused arguments form sentences with two predicates—the non-verbal predicate and the finite predicate. I maintain that when the subject of the non-verbal predicate co-indexes the subject of the primary predicate, the actor focus is triggered. The actor focus is not triggered when the subject of the non-verbal predicate does not co-index the subject of the primary predicate. I conclude that all clauses with focused arguments form complex predicates. The actor focus presents, therefore, as a subject-sharing complex predicate.

An Unmediated Analysis of Relative Clauses

Yehuda N. Falk, The Hebrew University of Jerusalem

The standard view of (restrictive) relative clause constructions is that they consist of three parts: the head, the relative pronoun, and the clause.

(1) the word processor which Bill prefers *e*
head relative pronoun clause

This view, standard both in P&P and in LFG, holds the relative pronoun to be the central component of this construction, as it serves to link the other two elements of the construction. The relation between the head and in-clause function is indirect, mediated anaphorically by the relative pronoun. I will refer to this as the mediated analysis of relative clauses. From the perspective of this analysis, it is very odd that there is an alternative form for relative clauses in English, one in which there is no relative pronoun:

(2) the word processor (that)Bill prefers *e*
head clause

The existence of this kind of relative clause suggests a direct relation between the head and the in-clause position, an unmediated analysis. An unmediated analysis of relative clauses has appeared from time to time in the transformational literature under the name "raising analysis" (e.g. Schachter 1973, Vergnaud 1974, Kayne 1994). This paper argues for an unmediated analysis in the context of LFG.

Under the LFG conception, long-distance dependency (or *wh*) constructions are constructions in which one element has (at least) two grammatical functions in potentially distant clauses. Given that the element in question has multiple grammatical functions, it could potentially be realized in the position of either function. The choice between the two options for realization give rise to the distinction between "ex-situ *wh*" and "in-situ *wh*". In the realm of relative clauses, the choice is between an externally-headed relative clause construction (EHRC), the equivalent of the ex-situ construction (as in English), and an internally-headed relative clause construction (IHRC), the equivalent of the in-situ construction (illustrated in (3)).

(3) a. Mooré (Culy 1990: 76)

[Yãmb	sẽn	yã	dao	ninga	zamẽ	wã]	bee	ka.
2PL	AUX	saw	man	INDEF	yesterday	DEF	be	there

'The man that you saw yesterday is here.'

b. Imbabura Quechua (Cole 1982: 49)

[Wambra	wagra-	ta	randi-	shka]	ali	wagra-	mi.
boy	cow-	ACC	buy-	NMNL	good	cow-	FOC

'The cow which the boy bought is a good cow.'

In an IHRC, the relativized element occupies the canonical position of its in-clause function, and the external "head" position is merely a determiner or nominal inflection appended to the relative clause. Comparing the EHRC and IHRC constructions, we see that the relativized element functions both as the head of the construction and as the in-clause element. There is no evidence of a relative pronoun in the IHRC construction. There are no in-situ relative clause constructions in which the relativized element occupies the external head position and a *wh* element (relative pronoun) occupies the in-clause position. The IHRC construction thus provides evidence for an unmediated analysis.

Another typological argument for the unmediated analysis is the cross-linguistic distribution of EHRC constructions: those with relative pronouns and those without. The mediated analysis suggests that constructions with relative pronouns should be common and those without (missing, as they do, the linchpin of the entire construction) should be relatively rare. The facts do not back this up: Maxwell (1979) shows, based on the 49 languages in the database of Keenan and Comrie (1979), that languages with pronoun-less relative clause constructions are quite common. In some languages, such as Toba Batak and Japanese, these are the only kind of relative clause, while in others, such as Spanish and Czech (and English), relative-pronoun relatives also exist. The widespread distribution of relative-pronoun-less relative clause constructions argues for an unmediated analysis.

An argument originally due to Schachter (1973) relates to idiom chunks (examples from Hulsey and Sauerland 2006):

(4) a. Mary praised the headway that John made.
b. I was shocked by the advantage that she took of her mother.

The idiom chunks *headway* and *advantage* are licensed by being arguments of *make* and *take* respectively. However, in these examples, they are only arguments of the correct verbs under the unmediated analysis. Under the mediated analysis, the relative pronoun, functionally an element distinct from *headway*/*advantage*, is an argument of the idiomatic verb, while the idiom chunks are arguments of the main verbs. The mediated analysis thus predicts that these should be ungrammatical; it is only under the unmediated analysis that we have an account of their grammaticality. On the other hand, a

transformational implementation of the unmediated analysis (a.k.a. the raising analysis) also fails, because it is also possible for the licensing verb to be in the main clause:

(5) Mary never made the headway that had been expected of her.

An LFG implementation of the unmediated analysis can account for both kinds of idiom-chunk examples, and is thus superior both to a mediated analysis (which can only account for (5)) and to a "raising"-type implementation of the unmediated analysis (which can only account for (4)).

The paper then proceeds to work out the details of an unmediated analysis in LFG, drawing on both EHRCs and IHRCs. It emerges from a careful consideration that the relativized element is not feature-identical in the two functions: the two funcions differ in the features CASE and DEF. The following illustrates this for DEF: the head (or rather the larger NP, the OBJ of 'buy') is definite, while the in-clause function (OBJ of 'make') is indefinite.

(6) Lakhota (Williamson 1987: 171)
Mary owíža wą kağe ki he ophewathų.
Mary quilt a make the DEM I.buy
'I bought the quilt that Mary made.'

The equations licensing the relative clause constructions thus need to use the restriction operator (Kaplan and Wedekind 1993). For English, approximately:

(7) NP → NP CP
↑=↓ ↓∈(↑ ADJ)
(↓ OPER) / DEF/ CASE =↑ / DEF/ CASE

An "operator" function OPER is hypothesized here as a formal device (providing a specific function in the relative clause that the larger NP can be related to), but it is simply part of the "chain" of linked functions, not a distinct functional element. It transpires that this operator plays a role in allowing relative clauses with relative pronouns.

Under an unmediated analysis, the existence of relative clauses with relative pronouns initially looks as mysterious as the relative clause without relative pronouns looks under the mediated analysis. The evidence shows that even relative clauses with relative pronouns involve an unmediated analysis. For example, idiom chunks show the same behavior.

(8) Mary praised the headway which John made.

An examination of the distribution of relative pronouns in a variety of language, including in infinitival relative clauses in English, reveals that the primary purpose of such relative clauses is to allow pied-piping: constructions in which the operator is only part of the topic of the relative clause.

(9) a. a word processor [to mangle the text] /*[which to mangle the text]
b. a word processor [to hate with a passion] /*[which to hate with a passion]
c. a word processor [to crash the computer with] /*[which to crash the computer with]
d. a word processor [with which to crash the computer]

The use of a relative pronoun allows the relative clause to have an element with the function TOPIC (in specifier of CP position), an element which contains the relativized element (the relative operator). It is an extension of the functional equations already in place for *wh* questions (see, e.g., Falk 2001), under which an element in the specifier position of CP bears a grammaticized discourse function and contains (or is) an OPER. Here, the existence of those functional equations is exploited, along with the OPER in relative clauses, to allow greater flexibility.

References

Cole, Peter (1982) *Imbabura Quechua*. Amsterdam: North Holland.

Culy, Christopher D. (1990) *The Syntax and Semantics of Internally Headed Relative Clauses*. Ph. D. dissertation, Department of Linguistics, Stanford University.

Falk, Yehuda N. (2001) *Lexical-Functional Grammar: An Introduction to Parallel Constraint-Based Syntax*. Stanford, Calif.: CSLI Publications.

Hulsey, Sarah and Uli Sauerland (2006) "Sorting Out Relative Clauses." *Natural Language Semantics* 14: 111–37.

Kaplan, Ronald M. and Jürgen Wedekind (1993) "Restriction and Correspondence-Based Translation." *Proceedings of the 6th Conference of the Association for Computational Linguistics European Chapter*. Utrecht University 193–202.

Kayne, Richard S. (1994) *Antisymmetry in Syntax*. Cambridge, Mass.: MIT Press.

Keenan, Edward and Bernard Comrie (1979) "Data on the Noun Phrase Accessibility Hierarchy." *Language* 55: 333–51.

Maxwell, Daniel N. (1979) "Strategies of Relativization and NP Accessibility." *Language* 55: 352–71.

Schachter, Paul (1973) "Focus and Relativization." *Language* 49: 19–46.

Vergnaud, Jean-Roger (1974) *French Relative Clauses*. Ph. D. dissertation, MIT.

Williamson, Janis S. (1987) "An Indefiniteness Restriction for Relative Clauses in Lakhota." in Eric J. Reuland and Alice G. B. ter Meulen, eds., *The Representation of (In)definiteness*. Cambridge, Mass: MIT Press. 168–90.

Particle Verbs in Computational LFGs: Issues from English, German, and Hungarian

Martin Forst, Tracy H. King, and Tibor Laczkó

In a number of languages, especially Germanic and Finno-Ugric, there are classes of verbs commonly called "particle verbs" (Ackerman 1983, Piñón 1992, Lüdeling 2001, Toivonen 2001). Particle verbs are verbs whose meaning and argument structure depends on the combination of a (base) verb and a particle. Often the meaning and argument structure of a particle verb is not compositional, i.e. it is not predictable from the combination of its components, but it must be listed in the lexicon. An example of a meaning expressed by a particle verb in English, German, and Hungarian is *He **gave up** the fight.* = *Er **gab** den Kampf **auf**.* = *Ő **fel#adta** a küzdelmet.* However, particle verbs can also be compositional (*push them up/in/out*; *push up/in/out the boxes*) and highly productive, which is a challenge for the coverage of computational grammars (Villavicencio 2003).

In this paper, we present the, so far undocumented, ways in which particle verbs are implemented in two relatively mature computational grammars, the English and the German *ParGram* LFGs (Butt et al. 2002), and we will address the issues that arise with respect to particle verbs in the development of a computational LFG for Hungarian. We will see that considerations concerning the *ParGram* LFG implementation of productive Hungarian particle + verb combinations raise questions as to the current treatment in the other two grammars. In addition, a set of Hungarian particles exhibit inflectional properties as well; we will also outline an LFG analysis of this phenomenon.

Particle verbs — syntactic or morphological objects? English particle verbs are typically analyzed in such a way that the two components are separately inserted in their respective syntactic positions, which is not surprising given that particles are always written as separate words and short NPs can intervene between base verbs and particles. In German and Hungarian, however, particle + verb combinations are generally spelled as a single word when the particle immediately precedes the verb (although a certain variation with respect to the spelling as one or two words can be observed with semantically compositional particle + verb combinations), and this order is in a way the default order, since only clearly definable conditions (V1 and V2 in German; negation, imperatives, etc. in Hungarian) cause particles to appear in positions other than the immediately preverbal one. In addition, there are verbs in German and Hungarian that do not exist on their own, but only in combination with particles; examples of such verbs are *aus#flippen* 'to flip/freak out' (German) and *be#fejez* 'to finish' (Hungarian). As a result, there is substantial controversy in the linguistic literature concerning the status of particle + verb combinations as syntactic or morphological objects. We will argue for a uniformly syntactic treatment of particles across the three LFG implementations and offer analyses that nevertheless capture the lexical properties of particle verbs in a principled manner.

Current Implementations in the English and German *ParGram* LFGs As verb particles are always spelled as separate words in English, particle verbs receive a syntactic analysis in the English *ParGram* LFG. The lexical entries of verb particles contribute a feature called PRT-FORM, which simply records the form of the respective particle, and the lexical entries of base verbs introduce the semantic form of the particle verb with its argument structure. Finally, the lemma of the base verb and the form of the particle are concatenated via an implementational device (CONCAT) so that the combination of the two, rather than just the lemma of the base verb, is the PRED of the respective f-structure. All particle verbs are listed with their argument structures in the verb lexicon of the grammar, and they appear under the corresponding base verb, but restricted to co-occuring with the appropriate particle. Below are the lexical entries involved in the analysis of our English example sentence to illustrate this treatment, as well as the f-structure associated with it. This analysis captures the syntactico-semantic facts in that the PRED reflects the potentially idiosyncratic particle verb meaning and the corresponding argument structure. However, it does not allow the system to construct productive particle verbs on the fly nor does it differentiate between compositional and non-compositional particle verbs.

```
give  V (^ PRED) = '%NewPred<(^ SUBJ)(^ OBJ)>'
        (^ PRT-FORM) =c up
        @(CONCAT %stem # (^ PRT-FORM) %NewPred).

up    PART  (^ PRT-FORM) = up.
```

```
[PRED  'give#up<[1:he], [99:fight]>'
SUBJ 1[PRED 'he']
      [PRED 'fight'                ]
OBJ   [SPEC [DET [PRED 'the']]     ]
   99 [                            ]
```

In German V1 and V2 clauses, particle verbs are spelled as separate words. In these contexts, the German *ParGram* LFG thus treats them in the same way as its English counterpart. In verb-final clauses and in headed VPs, however, particle verbs are usually spelled as single words. Compare, e.g., *Er **lud** seine Kusine **ein**.* 'He invited his cousin.' and *Er wird seine Kusine **einladen**.* 'He will invite his cousin.'

The finite-state morphology currently used by the German *ParGram* LFG outputs analyses like the following for forms of particle verbs:

```
einlud
ein#laden +V .13 .Sg .Past .Ind
```

The hash mark indicates the boundary between the particle and the base verb and thus potentially disambiguates analyses involving a separable verb particle from analyses involving homophonous non-separable verb prefixes; however the entire lemma is still a single unit. As a result, the grammar must analyze spelled-together particle verbs as morphological objects, and the lexical information for the particle verb *ein#laden* must be listed both under the base verb lemma (as in the English *ParGram* LFG), e.g. *laden*, and under the particle verb lemma, e.g. *ein#laden*. In order to allow for a uniformly syntactic analysis of particle verbs like the one in the English grammar, the analysis produced by the finite-state morphology would have to separate the particle from the verb, as done, e.g., by SMOR, a morphology developed at the IMS of Stuttgart University:

```
ein <VPART> laden <+V> <13> <Sg> <Past> <Ind>
```

As in the English grammar, all particle verbs must currently be listed with their argument structures in the German verb lexicon, so that the system exhibits the same limitations with respect to productively formed combinations. The `CONCAT` template makes it possible to project analogous f-structures regardless of whether a given particle verb is spelled together or as separate words.

Compositional and Productively Formed Particle Verbs As pointed out already, the implemented analyses do not differentiate compositional particle + verb combinations from idiomatic particle verbs. This is a problem for the coverage of computational grammars because new combinations inevitably show up in corpus texts and because the regular character of these combinations is not captured. E.g., the particles *hinterher* (German) and *rá* (Hungarian) can basically combine with any motion verb and (optionally) introduce an $\textsc{Obj}_\theta$/$\textsc{Obl}_\theta$, as the following sentences exemplify:

Lauf	dem	Glück	nicht	länger	hinterher!	Mari	rá-lépett	a	doboz-ok-ra.
run.IMP.2SG	the-DAT	happiness	not	longer	after	Mari.NOM	onto step.PAST.3SG	the	box-PL-onto
'Don't run after happiness any longer!'						'Mari stepped onto the boxes.'			

This behavior can be analyzed by means of a lexical entry for the particle where, rather than a PRT-FORM feature, it contributes a PRED that subcategorizes for the argument it introduces, and a predicate composition rule involving restriction similar to the one proposed for Urdu causatives by Butt et al. (2003). Other productively used particles fill argument slots of the base verb or simply contribute aspectual information. We will provide fully worked-out and implemented analyses at the conference.

Hungarian Inflected Preverbs In addition to the uninflected particles found in Germanic, Hungarian has inflected preverbs: when certain particle verbs take a pronominal argument, their preverbs are inflected for the person and number features of this argument. The pronominal argument is not overtly realized in that case.

Mari	rá-juk	lépett.
Mari	onto-3PL	step.PAST.3SG
'Mari stepped onto them.'		

We propose that these inflected particles can be straightforwardly treated in LFG by dint of an analysis whereby the inflected particle provides the PRED 'pro', as well as number and person information, for the particle verb's argument (e.g. *rájuk* (↑ OBL PRED)='pro', (↑ OBL PERS)=3, (↑ OBL NUM)=pl).

Conclusions We propose that the implemented LFG analysis of particle verbs for English and German is appropriate and feasible for non-compositional particle constructions in Hungarian. In addition, LFG, and its implementation via XLE, allows for a straight-forward analysis of Hungarian inflected preverbs. A closer examination of the productive, compositional Hungarian particle constructions has resulted in a re-examination of the analysis of such constructions in English and especially German. An orthogonal issue to those addressed here is that of how particle verbs participate in derivational morphology and how best to implement this (e.g. English *bystanders*, German *Einladung* 'invitation'); we leave this area for future work.

References

Ackerman, F. (1983) *Miscreant morphemes: phrasal predicates in Ugric.* PhD thesis, UC Berkeley. **Booij, G. (1990)** The boundary between morphology and syntax: separable complex verbs in Dutch. In *Yearbook of Morphology 3.* **Butt et al. (2002)** The Parallel Grammar Project. In *Proceedings of the COLING 2002 Workshop on Grammar Engineering and Evaluation.* **Butt et al. (2003)** Complex Predication via Restriction. In *Proceedings of the LFG 2003 Conference.* **Dehé, N. et al. (2002)** *Verb-Particle Explorations.* Mouton de Gruyter. **Lüdeling, A. (2001)** *On Particle Verbs and Similar Constructions in German.* CSLI Publications. **Piñón, C. J. (1992)** The preverb problem in German and Hungarian. In *Proceedings of BLS, 395–408.* **Toivonen, I. (2001)** *The Phrase Structure of Non-Projecting Words.* PhD thesis, Stanford University. **Villavicencio, A. (2003)** Verb-Particle Constructions and Lexical Resources. In *Proceedings of the ACL 2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment.*

I-structure, s-structure and multiple questions in French and Hungarian

Anna Gazdik
Université Paris 7
ELTE Budapest

In this paper we discuss multiple questions (questions which contain more than one information gap within the sam clause *e.g. Who said what?*) in two typologically different languages, French and Hungarian.

In both languages, different syntactic structures of multiple questions can be identified, with different interpretations (multiple questions have either a pair-list reading *e.g. Who left when? -Mary left in the morning, John in the afternoon...*, or a single-pair reading *Who left and when? Mary left in the morning.*). In Hungarian (and in other languages, for Romanian see Comorovski (1996), and for Czech Skrabalova 2007), it is possible to extract both (all) question words. Such questions license a pair-list answer. This structure is agrammatical in French:

(1) *Ki kivel ment moziba?/ *Qui avec quoi est venu à la fête ?*
who who.with went cinema.to/ who with what is come to the party
Who went to the cinema with whom?/ (Who brought what to the party?)

In the next type, one *wh*-word is extracted, the other is *in situ*:

(2) *Kinek mutattál be kit?/ Qui as-tu présenté à qui ?*
whom.to introduced VM whom/ whom have you introduced to whom
Whom did you introduce to whom?

This structure exists both in Hungarian and in French. In the former, only a single-pair reading is available and there is a strong preference for cases in which the question words belong to the same lexeme (*ki (who) and its declined forms, or mi (what) and its declined forms*), in other words, they refer to the same set of entities. In French, this question type is ambiguous between the single-pair and the pair-list reading. In informal French, all question words can appear *in situ*, with the same interpretation conditions, which, in turn, is not possible in Hungarian (in which at least one question word is obligatorily extracted):

(3) *Tu as donné quoi à qui?/ *Te adtál mit kinek?*
you have given what to whom/ you gave what to whom
What did you give to whom?

In the last case, the *wh*-words are coordinated in a sentence-initial position:

(4) *[Mikor és mit] adott János Marinak a múzeumban?/ [Quand et pourquoi] est-il parti ?*
when and what gave John Mary.to the museum.in/ when and why is he left
What did John give to Mary in the museum and when?/ When and why did he leave?

Although this structure is present in both languages, there are important differences between the two. First, in Hungarian, it licenses only a single-pair reading, whereas in French both interpretations are possible. Second, in Hungarian any two question words can be coordinated, while in French, coordination is more restricted: the conjuncts have to share all their functions, thus (5) is agrammatical:

(5) **[Quand et qui] est parti ?*
when and who is left
(Who left and when?)

This structure is problematic in that in Hungarian it is possible to coordinate question words that have different grammatical functions. The phenomenon, at first sight, is difficult to handle in a set-based feature-resolution analysis distinguishing between + and - distributive features in coordinated structures(Dalrymple & Kaplan 2000), or in Peterson (2004)'s framework, according to which only grammatical features distribute, lexical features do not (especially if we assume that discourse functions are represented at a separate i-structure and not at f-structure, thus *focus* cannot be the common grammatical function). In the present analysis, we concentrate on 3 types of problems mentioned above and show how the LFG architecture can account for them. We build on [Mycock:2006]'s analysis in that the focus status of *wh*-questions in the information structure can come from different sources (syntax, prosody, context, etc.) and on [Dalrymple:2010]'s proposal based on [DalrympleNikolaeva:toappear], concerning the relationship between information and semantic structure (categorization of meaning constructors (semantic information) according to their information structure role in a complex semantic structure. First of all, it is well-known that in pair-list questions the *wh*-words do not have the same status. One of them has to denote a contextually determined set, all the elements of which are to be paired up, in the answer, with one element of the set denoted by the other question word. [Comorovski:1996] refers to this phenomenon as the *D-linkedness* of question words. In Hungarian (1), this difference is indicated syntactically: D-linked question words precede non-D-linked ones. In French, on the other hand, the syntax is not revelatory in this respect. In a structure like (2), any of the question words can be D-linked depending on the context. We propose, therefore, that D-linkedness in multiple questions is related to information structure phenomena. We assume an information structure architecture based on that of Halliday (1967) and Steedman (2000) (theme/background and rheme partitions, both divided in a focus (prominent, highlited) and a background (non-prominent) part), which is compatible with Butt & King's (1996) approach as well. Contrary to [Mycock:2006], who places all question words to the *focus* set, we associate D-linked question words with the highlighted part of the *theme*. To support this view, we refer to the following facts: only one preverbal focus is permitted in Hungarian (in declaratives, the second focus is obligatorily sentence-final); in the answer, (contrastive) topics and not foci correspond to D-linked question words (expressed in the prosody as well); D-linked question words, like topics, refer to entities that are salient and (often) that have already been introduced into the discourse; finally, being the *sorting key* (Kuno & Takami 1993), D-linked question words thematize the answer to the question.
Secondly, we propose a restriction at the semantic structure to the problem of Hungarian *wh*-questions in (2). According to our approach, when the two question words are in the same clause and they are not coordinated, their feature *animate+/-* must agree, *i.e.* all the elements of the *focus* set must share the same feature, otherwise the structure is ill-formed. This approach is an elegant way of accounting for the fact that the question words can only differ in their cases, but they have to belong to the same lexeme (*who* ot *what*).
Finally, we account for the differences in coordination between the two languages in f-structure constraints. We propose that, whereas in Hungarian it is enough if the conjuncts share at least one of their grammatical functions, in French they have to share all of them. At f-structure in Hungarian, the shared grammatical function of *wh*-words is that of *extracted* (Q), *i.e.* it is enough if both conjuncts are extracted. This predicts that the same coordination *in situ* would lead to the degradation of the acceptability of the sentence, which prediction is borne out:

(6) *??* *Ki* *adott* *Marinak* *[mikor* *és* *mit]* *a* *múzeumban?*
who gave Mary.to when and what the museum.in
Who gave what to Mary in the museum?

French						Hungarian				
XP	→	(XP)*	Conj	XP		XP	→	(XP)*	Conj	XP
		$\downarrow \in \uparrow$		$\downarrow \in \uparrow$				$\downarrow \in \uparrow$		$\downarrow \in \uparrow$
		$\downarrow \in$Q		$\downarrow \in$Q				$\downarrow \in$Q		$\downarrow \in$Q
		$\uparrow$ GFα $\downarrow$		$\uparrow$ GFα $\downarrow$						

Selected References

[Comorovski:1996] Comorovski, Ileana (1996). *Interrogative Phrases and the Syntax-Semantics Interface.* Kluwer Academic Publishers: Dordrecht, Boston, London.

[Dalrymple:2010] Dalrymple, Mary (2010). *Glue and information structure.* Paper presented at the LFG Meeting, SOAS, 6 February 2010.

[DalrympleNikolaeva:toappear] Dalrymple, Mary & Nikolaeva, Irina (to appear). *Objects and information structure.* CUP.

[Mycock:2006] Mycock, Louise (2006). *The Typology of Constituent Questions: A Lexical-Functional Grammar analysis of 'wh'-questions.* Ph.D. thesis, University of Manchester.

Closing the Gap Between Stochastic and Hand-crafted LFG Grammars

Annette Hautli
Universität Konstanz
annette.hautli@uni-konstanz.de

Özlem Çetinoğlu Josef van Genabith
Dublin City University
{ocetinoglu,josef}@computing.dcu.ie

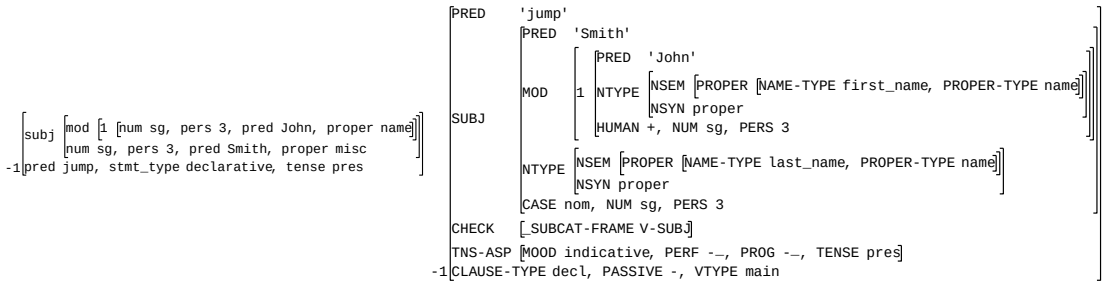
Developing large-scale deep grammars in a constraint-based framework such as Lexical Functional Grammar (LFG) is time-consuming and requires significant linguistic insight. This paper presents an approach to extend the stochastic DCU LFG annotation algorithm with more detailed f-structure information. It thereby reaches the feature detailedness of state-of-the-art hand-crafted grammars such as the English XLE grammar, while profiting from the robustness and the good coverage of stochastic grammars.

The DCU annotation algorithm (DCU grammar) (Cahill, 2004) comprises of the following parts: a stochastic parser (Charniak, 2000, Bikel, 2002) that creates trees in the Penn-II Treebank style (Marcus et al., 1993), a module that automatically annotates these trees with f-structure equations, a constraint solver that unifies the equations and produces f-structures, and a module that resolves long-distance dependencies. The system has been successfully evaluated on gold standards such as the PARC700 (King et al., 2003), outperforming the hand-crafted English XLE grammar by 2% (Cahill et al., 2008).

This provides an excellent basis for a further development of the DCU grammar. However, as the system was solely tuned to produce presentations restricted in detailedness with only some syntactic and semantic features, it lacks the detailedness of the hand-crafted English XLE grammar. In order to close the gap between the stochastic and the hand-crafted grammar, we need to extend the restricted feature space of the DCU grammar to get f-structures as detailed as XLE f-structures. Table 1 gives the original DCU features (34 in total) with the newly added f-structure features in bold (29 added). Extending the feature

F-structure feature space of the DCU grammar
adegree, adjunct, **adjunct-type**, **adv-type**, aquant, **atype**, **case**, **clause-type**, **common**, comp, conj, **coord**, coord-form, **deg-dim**, **degree**, **deixis**, **det**, det-form, **focus**, focus-int, **gend-sem**, **human**, **inf-type**, mod, **mood**, **name-type**, **nsem**, **nsyn**, **ntype**, num, number, number-type, obj, obj-th, obl, obl-ag, obl-compar, **part**, passive, pcase, perf, poss, prog, pron-form, pron-int, pron-rel, proper, **proper-type**, prt-form, **psem**, **ptype**, quant, **spec**, stmt-type, subj, subord-form, tense, **time**, **tns-asp**, topic-rel, **vtype**, xcomp, **xcomp-pred**

space includes renaming the already existing features and restructuring their representation in the f-structure. The following f-structures for *John Smith jumps.* exemplify how the tense/aspect paradigm is represented in the restricted and the extended DCU grammar. The existing TENSE feature is embedded into the TNS-ASP f-structure which now also contains the features MOOD, PERF and PROG. The extended f-structures also contain features that



distinguish first names from last names (as in *John Smith*). To add this information, the

relation of the tree nodes was taken into consideration. If two (or more) proper nouns are sisters and the lemma of the leftmost node can be found in a list of first names, we annotate it with the information that it is a first name. This is successively done until one of the sisters to the right is not a first name, annotating this node with the feature for last names. Therefore, names with a middle name are also detected (e.g. *John Adam Smith*).

A recent approach (Hautli and King, 2009) attempted to overcome the differences between the hand-crafted XLE grammar and the stochastic DCU grammar by using a set of ordered rewrite rules that added missing f-structure information, accepting the DCU system as a black box. Both architectures (extended DCU grammar vs. rewrite rules) were tested on a testsuite of 720 sentences, used to test the semantics of the English XLE grammar. The results show that the features can be reconstructed successfully in both scenarios, however extending the DCU grammar is more effective than using a set of rewrite rules, because we can allow for operations that are unavailable to the rewrite approach, such as taking into account the relation of nodes in the tree, as exemplified in annotating first and last names. For external validation, we took the PARC700 gold standard with a core feature

	Extended DCU grammar			**Set of rewrite rules**		
	precision	recall	f-score	precision	recall	f-score
sem_test	82.14	76.23	79.08	70.31	67.69	68.98

structure and evaluated the DCU grammar against it. This allowed us to check whether the extended system also performs well on other data than the testsuite, as in the case of PARC700, newspaper text of the Wall Street Journal. Performance with the development set of PARC700 (140 sentences), has the following results:

	precision	recall	f-score
development set PARC700	83.93	74.5	78.93

These initial experiments show that the gap between stochastic and hand-crafted grammars can be closed, which means that there is a possibility of generating deep LFG grammars on the basis of treebanks for other languages as well, benefiting from the aspect that the trees are automatically created with a robust parser and have a very good coverage for unknown text. As opposed to the earlier approach of employing the DCU grammar as a black box and using rewrite rules in addition, this new approach results in a single DCU grammar, which is more efficient and also has a higher accuracy due to the extra information that is available within the DCU grammar. By being able to take into account tree information which we previously could not, we allow for more LFG linguistic insight that can be captured in the final f-structure representation.

References

D. M. Bikel. Design of a multi-lingual, parallel-processing statistical parsing engine. In *The Proceedings of HLT 2002*, 2002.

A. Cahill. *Parsing with Automatically Acquired, Wide-Coverage, Robust, Probabilistic LFG Approximations*. PhD thesis, School of Computing, Dublin City University, 2004.

A. Cahill, M. Burke, R. O'Donovan, S. Riezler, J. van Genabith, and A. Way. Wide-coverage deep statistical parsing using automatic dependency structure annotation. *Computational Linguistics*, 34(1):81–124, 2008.

E. Charniak. A maximum entropy-inspired parser. In *Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics*, pages 132–139, 2000.

A. Hautli and T. H. King. Adapting stochastic lfg input for semantics. In Miriam Butt and Tracy Holloway King, editors, *Proceedings of the LFG '09 Conference*. CSLI Publications, Stanford, 2009.

T. H. King, R. Crouch, S. Riezler, M. Dalrymple, and R. Kaplan. The PARC700 dependency bank. In *Proceedings of the EACL03: 4th International Workshop on Linguistically Interpreted Corpora (LINC-03)*, 2003.

M. P. Marcus, B. Santorini, and M. A. Marcinkiewicz. Building a large annotated corpus of English: The Penn treebank. *Computational Linguistics*, 19(2):313–330, June 1993.

The Syntax of Discontinuous Reciprocal Constructions

Peter Hurst – University of Melbourne

Many verbally marked reciprocal constructions allow what has become known as a discontinuous alternation formed from the combination of a subject (which may be singular) and a comitative phrase - and it is the entities contained within these phrases which symmetrically participate in the situation described by the predicate (for example, see (1c) below). This additional phrase has some surprising properties: despite being marked as a comitative phrase (and this is usually the case cross-linguistically), it is obligatory and both semantically and syntactically distinct from the subject NP. The exact status of this construction and its relation to its more basic counterparts has proved challenging for theoretical analysis (see Rákosi 2008, Siloni 2008, Dimitriadis 2004, 2008 for discussion). In this paper I present an LFG-based analysis of these constructions and show how these properties are predicted by my explanation for how the discontinuous reciprocal construction is related to both its intransitive reciprocal counterpart and ultimately the transitive base from which they are derived. My explanation draws upon the work of two researchers in the area of LFG; György Rákosi (2008) who introduces the idea of an argument unspecified for a thematic role (a "partner" argument), and James Webb (2008) who proposes a two-tiered extension to argument structure as a means of understanding the distribution of instruments in English. By analysing a partner argument in a manner similar to that of Webb's analysis for instruments, I show how the resulting predicate's richer a-structure explains the relationship between transitive verbs and their two non-transitive reciprocal counterparts – allowing for an explanation of discontinuous reciprocal constructions which can be incorporated into a general theory of verbally marked reciprocals.

Bantu languages in particular form these constructions productively as exemplified by the Swahili data below. I follow the terminology used by Rákosi (2008) and call the discontinuous reciprocal construction in (1c) a *dyadic* reciprocal construction and the construction in (1b) *monadic:*

(1a)	*Juma a-na-m-penda Halima*	*transitive:*
	Juma he-prs-her-loves Halima	SUBJ Verb OBJ
	"Juma loves Halima"	
(1b)	*Juma na Halima* w*a-na-pend-an-a*	*monadic reciprocal construction:*
	Juma and Halima they-prs-love-rec	$SUBJ_{pl}$ Verb-rec
	"Juma and Halima love each other"	
(1c)	*Juma a-na-pend-an-a na Halima*	*dyadic reciprocal construction*
	Juma he-prs-love-rec with Halima	SUBJ Verb-rec OBL
	lit: Juma loves each other with Halima	
	"Juma and Halima love each other"	Vitale (1981:145,146)

In previous work, Rákosi (2008) goes part way to answering how these constructions might be related by examining the monadic and dyadic alternations of symmetric verbs in Hungarian. He limits his analysis to lexicalised verbs which do not productively form these reciprocal constructions from transitive counterparts. This allows him to consider the discontinuous reciprocal construction as being basic with the monadic reciprocal construction being formed from it by a process similar to argument binding (see Alsina 1996). The key insight he makes is that these verbs have two arguments, one being a proto-agent (which maps to SUBJ), and the other being a partner (which maps to OBL). The partner argument (represented by [ ] in his examples below) is under-specified for a thematic role and as such, standard feature assignment requires that it receive a [-o] feature (see Rákosi 2008:444-446):

(2a) Hungarian: dyadic reciprocal construction		$quarrel\_{dyadic}$<[P-A] [ ]>
A katoná-k vesz-eked-t-ek az őrmester-rel	intrinsic	-o -o
the soldier-pl quarrel-rec-pst-3pl the sergeant-with	default	-r +r
"The soldiers quarrelled with the sergeant"		SUBJ OBL

The monadic construction is formed from the dyadic construction by grouping two arguments in a-structure and treating them as a whole with respect to argument mapping. Notationally this is indicated by the extra pair of square brackets around the two arguments:

(2b) Hungarian: monadic reciprocal construction		$quarrel\_{monadic}$<[[P-A] [ ]]>
A katoná-k vesz-eked-t-ek	intrinsic	-o
the soldiers quarrel-rec-pst-pl	default	-r
"The soldiers quarrelled"		SUBJ

Despite providing a synchronic analysis of the monadic/dyadic alternation in Hungarian, Rákosi's analysis cannot address how these constructions came to be formed from a transitive verb in the first place – and why dyadic reciprocal constructions should contain reciprocal morphology. As such, this analysis cannot be used to explain the three-way alternation seen in Bantu languages (exemplified by the Swahili above). This is because the basic lexical item in their analysis must be the transitive verb: and given that its argument structure typically selects a proto-agent and proto-patient, there is no obvious way to account for the oblique argument in the corresponding dyadic reciprocal construction.

The approach I take is to treat the accompaniment phrase in (1c) as a type of argument-adjunct along the lines of Webb's analysis for instrument phrases for English. Grimshaw (1990:108) defines an a-adjunct as one which cannot be assigned a theta-

role but which nevertheless is licensed by a-structure. As such, a-adjuncts have some sort of intermediate status between an argument and an adjunct. Accompaniment phrases are suitable candidates for analysis as a-adjuncts: like arguments, they participate in the event described by the predicate and are usually analysed as forming part of their predicate's conceptual structure (see Jackendoff 1990); like adjuncts they are optional and can be productively added to any semantically suitable predicate. Under this analysis, there are two tiers of a-structure, the first tier specifies canonical arguments: those which are uniquely selected by the predicate and which are obligatory. The second tier specifies the a-adjuncts and, if present, their mapping takes place after the first-tier arguments. Note that I will not follow Webb's analysis of assigning a thematic role to the a-adjunct, and instead leave its thematic description as underspecified (as per Rákosi's analysis and in line with Grimshaw (1990)). In my analysis below, the monadic reciprocal construction is first formed through a process of argument binding (3b), and the discontinuous reciprocal construction is subsequently formed from it with the addition of an a-adjunct (3c):

(3a) *Juma a-na-m-penda Halima*		love<[P-A] [P-P]>	
Juma he-prs-her-loves Halima	intrinsic	-o -r	
"Juma loves Halima"	default	-r	
		SUBJ OBJ	
(3b) *Juma na Halima* w*a-na-pend-an-a*		$love\_rec_{monadic}$<[[P-A] [P-P]]>	
Juma and Halima they-prs-love-rec	intrinsic	-o	
"Juma and Halima love each other"	default	-r	
		SUBJ	
(3c) *Juma a-na-pend-an-a na Halima*		$love\_rec_{dyadic}$<[[P-A] [P-P]]> , tier 2: <[ ]>	
Juma he-prs-love-rec with Halima	intrinsic	-o	-o
lit: Juma loves each other with Halima	default	-r	+r
"Juma and Halima love each other"		SUBJ	OBL

This analysis not only accounts for the syntax of reciprocal constructions in Bantu languages, but also has the virtue of being able to provide a natural account for the grammaticalisation process so common to verbally marked reciprocal constructions more generally. The relatively complex argument structure of the discontinuous reciprocal construction maps to just two grammatical functions and so is highly susceptible to grammaticalisation. In this process, the two bound arguments mapped to the subject NP are treated as a single argument and the partner a-adjunct becomes a first-tier partner argument:

(4) $verb\_rec_{dyadic}$<[[P-A][P-P]] >, <[ ]> → $verb_{sym}$<[P-A][ ]>
RECIP({entities},$\lambda x.\lambda y.verb_{basic}(x,y)$) $\lambda x.\lambda y.verb_{lexicalised}(x,y)$

This newly formed verb is now *inherently* symmetric - i.e., the symmetry of the event is no longer associated with the reciprocal morpheme (see Dalrymple et al. (1998) for discussion), but is now implied as part of the meaning of the new verb. As such, it is possible for the symmetry of the event to be cancelled - unlike that of the equivalent monadic reciprocal construction in the same situation. This is in fact seen in Hungarian (and other languages); for example, "quarrel" - *veszeked* in Hungarian when used in a dyadic construction does not have to be symmetric as evidenced by (5) below:

(5) *Én num veszeked-t-em János-sal ő veszeked-ett vel-em*
I not quarrel-pst-1sg John-with he quarrel-pst with-1sg
"I was not quarrelling with John, he was quarrelling with me" Rákosi (2008:423)

Conclusion

By building on recent work in LFG (Rákosi 2008, Webb 2008), I have provided a unified account of verbally marked reciprocal constructions whereby a dyadic reciprocal construction is analysed as being formed from a monadic construction in conjunction with a partner a-adjunct. This analysis is not only sensitive to these constructions' diachronic development from a basic transitive verb, but also provides some insight into why the dyadic reciprocal construction is so prone to grammaticalisation – and why it has subtly different semantics with respect to symmetry when compared to its monadic counterpart.

References

Alsina, A. 1996. *The role of argument structure in grammar: Evidence from Romance.* Stanford: CSLI Publications.
Dalrymple, M., Kanazawa, M., Kim, Y., Mchombo, S., Peters, S. (1998). Reciprocal Expressions and the Concept of Reciprocity in *Linguistics and Philosophy* 21:2, 159-210.
Dimitriadis, A. 2004. *Discontinuous Reciprocals*. Manuscript: Utrecht institute of Linguistics OTS.
Dimitriadis, A. 2008. *Irreducible symmetry in reciprocal constructions.* In König & Gast 2008.
Grimshaw, J. 1990. *Argument structure.* Cambridge, MA: The MIT Press.
Jackendoff, R. 1990. *Semantic structures.* Cambridge, MA: The MIT Press.
König, E. & Gast, V. 2008. *Reciprocals and reflexives: theoretical and typological explorations.* Berlin, Germany.
Rákosi, G. 2008. *The argument structure of inherently reflexive and reciprocal predicates.* In König & Gast 2008.
Siloni, T. 2008. *The syntax of reciprocal verbs: an overview.* In König & Gast 2008.
Vitale, A. 1981. *Swahili syntax*. Dordrecht – Cinnaminson: Foris Publications.
Webb, J. 2008. *Instruments in LFG's argument structure.* Manuscript: Master's thesis, University of Oxford.
Available at: http://ora.ouls.ox.ac.uk/objects/uuid%3Ae887fcc8-c2d3-4de6-a439-05e8461efdf3

RUSSIAN VERBAL PREFIXES AND THE PROJECTION ARCHITECTURE

Kateryna Kent & Jean-Philippe Marcotte, University of Minnesota

Russian perfectivity paradigms raise a complex network of formal issues for the projection architecture of LFG, including the structure of morphological representation and its relationship to the c-, f- and a-structures, with some consequences that appear to favor description-by-analysis over codescription for semantic interpretation. This paper presents the data and navigates its formal implications, suggesting in the end that a Paradigm Function Morphology (Stump 2001; Sadler & Nordlinger 2004, 2006) approach to the m-structure allows a clear description of the Russian facts that is equally compatible with both codescription and description-by-analysis.

Russian verb roots have inherent perfectivity specifications (1a), and stems can be successively perfectivized (1b,d) and imperfectivized (1c) by affixation:

(1)
a. *stroi- -yet*
build.[IMPF] 3SG.SUBJ
'she/he builds [houses]'

b. *na- stroi- -yet*
SP.CUMUL.[PERF] build.///// 3SG.SUBJ
'she/he builds a lot [of houses]'

c. *na- stra- -yva -yet*
SP.CUMUL.///// build.///// [IMPF] 3SG.SUBJ
'she/he builds a lot [of houses]'

d. *po- na- stra- -yva -yet*
SP.DISTR.[PERF] SP.CUMUL.///// build.///// ///// 3SG.SUBJ
'she/he builds a lot [of houses] everywhere'

The root *stroi-* in (1a) is inherently IMPF, but the prefixed cumulative verb in (1b) is PERF; it has a further IMPF form (1c), which the distributive prefix in (1d) perfectivizes. The words in (1b-d) can also be glossed and translated as in (1b′-d′):

(1)

b′. *na- stroi- -yet*
LP.[PERF] build.///// 3SG.SUBJ
'she/he tunes [a guitar]'

c′. *na- stra- -yva -yet*
LP.///// build.///// [IMPF] 3SG.SUBJ
'she/he tunes [a guitar]'

d′. *po- na- stra- -yva -yet*
SP.DISTR.[PERF] LP.///// build.///// ///// 3SG.SUBJ
'she/he tunes [all her/his guitars]'

The foregoing verbs exemplify the relevant perfectivity affixes of Russian, which are categorized in the literature (Ramchand 2008; Smith & Rappaport 1997; Svenonius 2004) as: the secondary imperfective suffix (2Impf), i.e. *-yva* in (1c,d,c′,d′), which here triggers a stem alternation; lexical prefixes (LPs), as *na-* in (1b′-d′); and superlexical prefixes (SPs), as *na-* in (1b-d), and *po-* in (1d,d′).

Secondary imperfective can only be suffixed to perfective stems—compare the derived perfective in (1b) to (1c)—or any inherently perfective root. Attaching 2Impf directly to an inherently imperfective root, e.g. ∗*stra–yva–yet*, is ungrammatical.

The two prefix types are drawn from the same set of forms and are uniformly perfectivizing, but they differ in their other properties. LP-derived stems have idiosyncratic meanings (compare (1a) 'build' with (1b′) 'tune', can have additional arguments compared with underived stems, always have a 2Impf form, but cannot take additional LPs. SP-inflected stems mostly disallow 2Impf ((1c) is an exception), though SPs can be attached to 2Impf stems (compare (1c) and (1d)), and can stack with each other (1d) and with LPs (1d′). SPs do not license argument structure but add information about the progress of the event, such as cumulativity and distributivity.

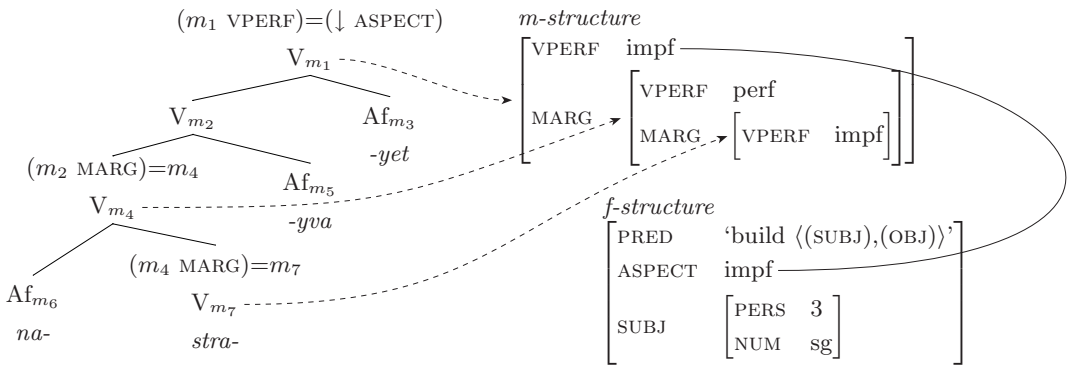
An analysis of Russian perfectivity paradigms must account for the argument-structure and idiosyncratic effects of LPs, the co-occurrence restrictions between affixes, and the alternating perfectivity values of verbs. We account for these various facts by exploiting the projection architecture of LFG.

First, we propose that LP-stems are memorized. This immediately accounts for their idiosyncratic meanings: *na-stroi-* 'tune' may be historically derived from *stroi-* 'build', but they bear no synchronic relationship. Memorization of LP-stems also explains how their argument structures can differ from those of the bare roots from which they are historically derived: different lexemes project different a-structures. The status of LP-stems in our account is somewhat analogous to that of particle verbs in English.

Second, we show that an m-structure account of co-occurrence restrictions is desirable. Consider (1b): the f-structure of a sentence headed by this verb must get its PRED from the verb root, and its [ASPECT perf] feature from the SP *na-*, which is accounted for directly if these two morphemes share an f-structure. But functional uniqueness prevents a pure f-structure account, since the verb root *stroi-* has its own [ASPECT impf] feature, as (1a) shows.

One solution is to use an m-structure, projected from the c-structure and codescribed in sublexical rules via the correspondence funtion μ. In this view perfectivity affixes function as m-structure heads with their own VPERF specfications, and with stems as their morphological arguments (MARG). Thus the structure of (1c) is as in (2), where all f-structure annotations on the daughter nodes are $\uparrow{=}\downarrow$, and all unmarked m-structure annotations are $\mu(\mathcal{M}(*)){=}\mu(*)$ (in the notation of Kaplan (1987)):

(2)



Co-occurrence restrictions are captured as contraints placed by affixes on the VPERF value of their MARG; the 2Impf suffix is thus lexically specified as $(\mu(\mathcal{M}(*))$ MARG VPERF$)=_c$perf. The c-structure rule for VP ensures that the ASPECT attribute of the f-structure of its V daughter has the value of outermost m-structure defined by the sublexical rules. The alternating perfectivity values of verbs are, on this account, a mere matter of morphosyntactic accounting.

This, however, poses problems for the glue semantics codescription approach to the syntax–semantics interface (Dalrymple 1999). Perfectivity is semantically interpreted. Hence morphemes codescribing perfectivity must carry a meaning constructor, projecting via the f-structure to the s-structure through the correspondence function composition $\sigma \circ \phi$, and composing with other such constructors to yield appropriate interpretations (we follow Ramchand (2008) in treating perfectivity as temporal definiteness). But both perfectivity affixes in (2), plus the root, must then have such a meaning constructor. All these meaning constructors would then enter the derivation, an undesirable consequence since this word clearly is semantically imperfective (and cumulative), not a perfectivized–re-imperfectivized imperfective, if a coherent (as opposed to merely successful) glue proof could even be achieved for such a thing.

This indicates that codescription is inadequate for modeling the composition of perfectivity in Russian. In contrast, a description-by-analysis syntax–semantics interface (Halvorsen & Kaplan 1988) would read the single perfectivity value in the f-structure and translate it into an appropriate logical expression for semantic composition.

Such an approach to the semantic contribution of atomic f-structure features may be compatible with a codescription approach to predicate saturation and scope relations. But we argue that it is preferable to sidestep this question altogether by acknowledging, with Stump (2001) and Sadler & Nordlinger (2006), that lexical-incremental approaches to morphology—of which the analysis of Russian verbs we sketch above is an example—are descriptively inferior to inferential-realizational approaches. In a Paradigm Function Morphology approach to Russian perfectivity, a word like *na-stra-yva-yet* (1c), rather than having the structure in (2), occupies a cell in a paradigm, along with a complete well-formed property set {CUMUL:+,ASP:impf,AGR:{PERS:3,NUM:sg}}. The form of the word is determined by realization rules applying in blocks based on the word's category and morphosyntactic properties. We assume that these property sets are translated to f-descriptions by an appropriate morphology–syntax interface (Sadler & Nordlinger 2004; Andrews 2005). The question of a word having multiple ASPECT specifications does not arise and, since the cell can also be associated directly with its semantic contribution, the issue of multiple aspectual meaning constructors does not arise if a codescription approach to the syntax–semantics interface is taken.

The properties of Russian perfectivity paradigms help shed light on LFG's projection architecture: a lexical-incremental m-structure is not compatible with a uniquely codescriptive syntax–semantics interface, but an inferential-realizational m-structure is neutral between codescription and description-by-analysis. We envision that further analysis of SP–stem co-occurrence restrictions—see for example Tolskaya (2007) on SPs compatible with verbs of motion—will have consequences for the relationship of morphology to semantics or lexical semantics.

Abstract: Walman *and*-verbs and the nature of Walman serialization

Lewis C. Lawyer
University of California, Davis

Walman is a language in the Torricelli family spoken on the northern coast of Papua New Guinea. In a recent article, Brown and Dryer (2008) discuss the phenomenon of Walman *and*-verbs, words which are morphologically transitive verbs but syntactically serve as coordinators coordinating noun phrases.

(1) *Kum* *m-etere-y* [*John* *n-aro-Ø* *Mary*].
1sg 1sg.S-see-3pl.O 3sg.m.S-and-3sg.f.O
"I saw [John and Mary]."
Brown and Dryer (2008, 539)

In the present paper I wish to demonstrate that the functional nature of these *and*-verbs is partially parasitic on the functional nature of a particular type of serial verb construction (SVC) found in Walman. This type of SVC is called inclusory serialization (following Crowley 2002, 41), and is characterized by the subject of one verb being the aggregate of the subject and object of the previous verb (see 2, below). I will argue that the functional structure required for the maximal projections of transitive verbs in this kind of serialization is a structure shared by *and*-verbs, and thus conducive to the innovation of *and*-verbs in a language.

The grammatical framework I will use in this paper is lexical-functional grammar (LFG). I hope this paper will also demonstrate the utility of an LFG-like model to descriptive linguistics. The relationship between inclusory serialization and *and*-verbs is difficult to discern in a traditional Chomskyan theoretical framework in which constituent structure is the sole determinant of functional roles and functional structure. In an LFG framework, where constituency and functional structure are modeled separately, the functional similarity between *and*-verbs and inclusive serial constructions is readily apparent.

(2) *Kum* *m-rachere-Ø* *pelen* ***k*-esi** *nakol.*
1sg 1sg.S-chase-3sg.f.O dog **1pl.S**-go.out house
"I chased the dog out of the house."
Brown and Dryer (2008, 551)

The agreement morphology on the second verb in (2) indicates agreement with the subject and object of the preceding verb, taken together. We may thus ask, where is the entity with which the subject agreement morphology on the second verb agrees? The simplest answer is that the first verb has created an indexable entity in f-structure which is the conjunction of its subject and object

arguments. It is this entity with which the subject agreement morphology on the second verb agrees, and it is this entity which functions as the subject of *kesi* at unification. Since any transitive verb in Walman may occur in an inclusive SVC, we must assume that every transitive verb in Walman has the capacity to construct an indexable conjunction of its subject and object. Turning again to *and*-verbs, their functional peculiarity is that they are apparently verbs yet they are used as NP coordinators. Yet from the preceding discussion it seems that all transitive verbs in Walman must have the capability to construct indexable conjunctions of their subject and object arguments. Thus in f-structure, *and*-verbs do not look very different from ordinary transitive verbs.

This similarity is obscured in a traditional constituency-based grammatical description. The simple coordinative function of *and*-verbs becomes buried underneath their perplexing distribution. This distribution is indeed unusual. The constituent formed by the *and*-verb together with its subject and object arguments is distributed as if it were an NP. It can occur as a possessor NP, an argument of a verb, the object of an adposition, or even a coordinand in a coordinated NP. (Brown & Dryer 2008, 538-545) However, the *and*-verb itself is in some ways like a verb. It takes the same set of subject and object pronominals as a normal transitive verb, so that its first and second coordinand appear to be its subject and object, respectively. Furthermore clause-level particles such as the negative particle and the perfective particle, which occur immediately before the first verb in a clause, may occur immediately before the *and*-verb if it is the first "verb" in the clause. (Brown & Dryer 2008, 546-7) The categorial nature of the *and*-verb is thus very unusual -- it appears to be a verb whose maximal projection is a noun phrase! This improbable mixed categorial affiliation, together with the unlikelihood of a language developing a verb with no semantic content but only a coordinative function, would account for the rarity of *and*-verbs cross-linguistically. Walman is the first language in which *and*-verbs have been observed.

In spite of the perplexing nature of the distribution and categorial affiliation of the Walman *and*-verb, in an LFG analysis the functional nature of the *and*-verb is clear, and not unusual. Like other NP coordinators crosslinguistically, the *and*-verb constructs an indexable conjunction of its subject and object arguments. In addition to allowing us this simple observation, the LFG analysis of Walman SVCs further reveals that all transitive verbs in Walman must have the capacity to perform this coordinative function. Thus we see that the linguistic ecology of Walman is well suited for the development of the unusual phenomenon of the *and*-verb.

Brown, Lea, and Matthew S. Dryer (2008) "The verbs for 'and' in Walman, a Torricelli language of Papua New Guinea." *Language* 84 (3), 528-565.

Crowley, Terry (2002) *Serial Verbs in Oceanic: A descriptive typology.* Oxford University Press.

Helge Lødrup (U of Oslo): Are Norwegian 'type anaphora' really surface anaphora?

Introduction The syntax of the Norwegian pronoun *det* 'it, that' presents challenges. This personal pronoun is basically non-human third person neuter singular. Under certain conditions, however, *det* can have noun phrase antecedents that are not neuter singular. Cf. (1)-(2), with masculine antecedents.

(1) John kjøpte en laks. **Det** kjøpte Marit også
John bought a salmon that bought Marit too ('Marit bought one too')
(2) John kjøpte en laks. Marit liker **det** ikke
John bought a salmon Marit likes that not ('Marit does not like that (i.e. salmon)')

det used as in (1)-(2) is called a type anaphor in Borthen (semanticsarchive.net 2003, DAARC 2004). This anaphor does not refer to the same object as its antecedent; in Borthen's view it refers to the same type of thing as as its antecedent.
This paper claims that type anaphora must be divided into two groups, with very different properties. One group realizes what was called surface anaphora in Hankamer and Sag (LI 1976), Sag and Hankamer (LP 1984). The other group will be called generic/eventive *det*.

Surface anaphora A deep anaphor refers to something in its linguistic or non-linguistic context. It has no internal structure in syntax. A surface anaphor, on the other hand, must have a linguistic antecedent, it cannot be deictic. One example is the zero proform in VP ellipsis (as in *Everything has changed, and he has Ø, too*). A surface anaphor must have internal structure in syntax; its antecedent must be available to check standard conditions on well-formedness. In Norwegian (and Scandinavian), the regular surface anaphor is the pronoun *det*, both with VP pronominalization, as in (3), and complement pronominalization with copular verbs (see Houser et al. WECOL 2007 on Danish).

(3) Har du spist? Nei, jeg har ikke **det**
have you eaten no I have not that ('Have you eaten? No, I haven't.')

The architecture of LFG makes it possible to to utilize the distinction between c-structure and f-structure to account for surface anaphora. The surface anaphor must be present in c-structure, while the corresponding position in f-structure contains the f-structure of the antecedent (see Levin MRGR 1982 on sluicing).
Some type anaphora share properties of surface anaphora (from Hankamer and Sag LI 1976). The focus will be on those *det* that are complements of verbs that include a 'have'-relation in their semantics, such as *kjøpe* 'buy', *gi bort* 'give away', etc. Other groups of verbs that seem to take *det* with the same properties include verbs of creation and consumption.
->Surface anaphora cannot be deictic, they must always have an antecedent in the linguistic context. This is also true of the relevant type anaphora (Borthen DAARC 2004).
->Surface anaphora show the 'missing antecedent' phenomenon. This is also true of the relevant type anaphora. In (4), the antecedent of the pronoun *den* 'it' cannot be the type anaphor itself, because *den* 'it' is masculine. The antecedent of *den* 'it' must be the f-structure representation of the type anaphor, based upon the antecedent *hund* 'dog' (masculine).

(4) John fikk ikke hund, men **det** fikk Marit. Den er veldig stor.
John got not dog, but that got Marit It is very big
(John did not get a dog, but Marit did. It is very big.)

->Other facts also show that there must be access to the antecedent of *det* in f-structure.
- The interpretation potential of the relevant type anaphor is always the same as that of the antecedent. The antecedent can be non-specific or specific (but normally not definite). An example is (5).

(5) John vil kjøpe hund / en hund / hunder / noen hunder. **Det** vil Marit også kjøpe
John will buy dog / a dog / dogs / some dogs that will Marit too buy

The antecedent determines the interpretation of *det*. With the bare singular *hund* 'dog', it can only be non-specific; with *en hund* 'a dog', it can be specific or non-specific, etc. (Note that the interpretation is not always identical in the two sentences.) This analysis is different from the one in Borthen, who insists that the type anaphor is always type referring and nonreferential (Borthen 2003:286-94).

-The antecedent must satisfy the selectional requirements of the predicate that takes the anaphor. For example, type anaphora can take antecedents that are non-referring (Asudeh and Mikkelsen, HPSG-book, CSLI 2000), including idiom chunks, as in (7).

(6) John røyker pipe. **Det** røyker Marit også
John smokes pipe that smokes Marit too ('So does Marit.')
(7) John får hetta / fnatt. **Det** får Marit også
John gets 'the-hood' / 'scabies' that gets Marit too ('John freaks out. So does Marit')

->The relevant type anaphora can occur in the object position of a presentational sentence, see (8). They are not affected by the indefiniteness restriction when the antecedent is indefinite.
->The relevant type anaphora are normally accented, and often topicalized, just like surface anaphora with VP pronominalization. They are reluctant to be object shifted, see (9), and compare (3) above.

(8) (Alkohol er tillatt) når det blir gitt tillatelse, og **det** ble det gitt av NN (google)
alcohol is allowed when there is given permission, and that was there given by NN
('Alcohol is allowed when permission is given, and it was given by NN.')
(9) Har du bil? Nei, jeg har ikke **det** / ??**det** ikke
have you car no I have not that that not ('Do you have a car? No, I don't.')

Surface anaphora The suggestion here is that type anaphora with some groups of verbs are surface anaphora. A surface anaphor must be allowed lexically by the verbs in question as a possible c-structure realization of a complement function. Verbs differ concerning the option of taking the surface anaphor *det*.

Generic/eventive *det* Example (2) above differs from (1). The antecedent of *det* is specific, but (2) cannot mean that Marit does not like some specific salmon. It can only mean that she does not like salmon in general. This is in itself reason enough that *det* cannot be a surface anaphor here. It is therefore necessary to assume another *det* that can take antecedents that are not neuter singular. Examples are (2) and (10)-(11) (with masculine antecedents)

(10) Sykkel er kult, og **det** er et nyttig framkomstmiddel. (Borthen 2003:41)
bike is cool, and it is a useful conveyance
(11) Jeg foreslår en ferie i Florida. **Det** vil gi familien mange minner.
I suggest a vacation in Florida that will give the-family many memories

The non-agreeing *det* in (2) and (10)-(11) raises many difficult questions. Tentatively, it could be split in two, generic *det* (examples (2) and (10)), and eventive *det* (example (11)).
Generic/eventive *det* does not share properties with surface anaphora. The generic *det* can be deictic. However, it can only be used generically; it does not establish a discourse referent (but possibly a 'short term referent' in a modal context). Referring to to a 'missing antecedent', as in (12), does not work

(12) John liker ikke en stygg jente, men **det** liker Per. *Hun er rik
John likes not an ugly girl, but that likes Per she is rich

The fact that generic/eventive *det* must always be non-specific makes it very different from the surface anaphor *det*. Generic/eventive *det* can take a referring noun phrase as an antecedent without being referring itself, as in (2) above. It can even take a definite antecedent; this is normally impossible with surface anaphora.
Generic *det* is also different from surface anaphora prosodically; it does not have to be accented, and it can object shift (as in (2) above).
An important difference between surface *det* and generic/eventive *det* is that generic/eventive *det* seems to have the distribution of a regular nominal phrase (semantic restrictions aside), there is no special selection by its predicate.
The view of generic/eventive *det* sketched above ties in with another issue in Norwegian (and Scandinavian) syntax. An adjectival XCOMP normally agrees with its subject, but any noun phrase that can be referred to by generic/eventive *det* can take an adjectival XCOMP with singular neuter morphology. An example is (10) above (*kult* 'cool' is neuter). This phenomenon is a classical problem in Scandinavian grammar. The last contribution, Josefsson (NJL 2009), proposes that there are two groups of disagreeing subjects, making a distinction which roughly parallels the one made here between eventive and generic *det*.

Using Templates to Account for English Resultatives

Elizabeth MacGregor

February 14, 2010

The English resultative construction is a secondary predication relationship in which the main verbal predicate is affected by the existence of a resultant state. This resultant state describes the outcome of the action which was performed on the patient of the main predicate. In this study I will attempt to describe a manner in which the resultative can be formalized through the use of LFG templates and Glue semantics.

For the sake of expedience during this treatment of the resultative, we will focus on two types of resultatives, (1) transitive resultatives, and (2) unaccusative resultatives (demonstrated below), although several other types can be identified.

(1) I dyed my skirt red.

(2) The river froze solid.

LFG templates, as presented in [ADT08], act as a bridge between syntactic form and semantic meaning. This is because the the lexical entries for verbs are associated with a given argument relationship, as represented by a statement in Glue Semantics. The lexical entry calls a structural form by default through a template call. This structural and semantic arrangement can be overwritten by a template which only calls to the original argument name, and then changes its syntactic form and Glue equation. So, given the structural rules for transitive (3) and intransitive (4) below, one can utilize the verbal predicates given in a lexical entry to generate any transitive or intransitive sentence, without the lexical entry having to specify the argument structure, as demonstrated in 5. A template for a more complex structure would then call for the function name which can be paired with a certain structure (TRANSITIVE or INTRANSITIVE). This template call will specify new values for both the argument structure and semantic relationships.

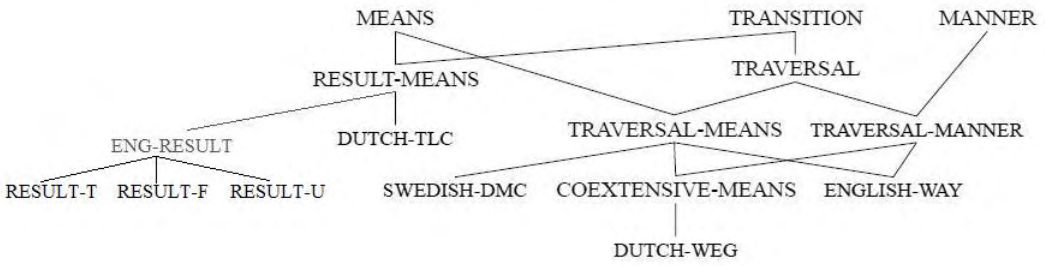
(3) TRANSITIVE(FN) = $(\uparrow PRED) = `FN < (\uparrow SUBJ)(\uparrow OBJ) >'$

(4) INTRANSITIVE(FN) = $(\uparrow PRED) = `FN < (\uparrow SUBJ) >'$

(5) hammer V $\left(\begin{array}{l} \lambda e.hammer(e) : (\uparrow_\sigma REL) \\ \text{@TRANSITIVE(hammer)} \\ \lambda P \lambda x \lambda y \lambda e.P(e) \wedge agent(e) = x \wedge patient(e) = y: \\ (\uparrow_\sigma REL) \multimap (\uparrow SUBJ)_\sigma \multimap (\uparrow OBJ)_\sigma \multimap \uparrow_\sigma \end{array} \right)$

With these basic rules and templates in place, we can look at the more complex structure of the resultative. Both of the kinds of resultative we are looking at take the meaning of the function, and apply it to a new verbal frame and Glue equation. Using a template hierarchy extended from [ADT08] (shown as 6), the English Resultative becomes an extension of the use of the RESULT-MEANS template, and inherits its properties and meanings. The two kinds of resultative that we are concerned with then inherit and further specify the ENG-RESULT template (7), forming the final templates of RESULT-T (transitive) and RESULT-U (unaccusative), shown below as 8 and 9, respectively.

(6) Extended Template Hierarchy



(7)
ENG-RESULT = @RESULT-MEANS
$\lambda R \lambda s.R(s)$:
$[(\uparrow_\sigma \text{EVENT2}) \multimap \uparrow_\sigma] \multimap [(\uparrow_\sigma \text{EVENT2}) \multimap \uparrow_\sigma]$

(8)
RESULT-T(FN) = @ENG-RESULT
@TRANSITIVE-X(FN)
$\lambda R \lambda P \lambda e \lambda s \lambda y.P(e) \wedge R(s) \wedge agent(e) = \neg y \wedge patient(e) = y$
$\wedge\ experiencer(s) = y$:
$(\uparrow \text{SUBJ})_\sigma \multimap (\uparrow \text{OBJ})_\sigma \multimap (\uparrow \text{XCOMP})_\sigma \multimap (\uparrow_\sigma \text{EVENT1}) \multimap (\uparrow_\sigma \text{EVENT2}) \multimap \uparrow_\sigma$

(9)
RESULT-U(FN) = @ENG-RESULT
@INTRANSITIVE-X(FN)
$\lambda R \lambda P \lambda e \lambda s \lambda x.P(e) \wedge R(s) \wedge agent(e) = x \wedge patient(e) = x$
$\wedge\ experiencer(s) = x$:
$(\uparrow \text{SUBJ})_\sigma \multimap (\uparrow \text{XCOMP})_\sigma \multimap (\uparrow_\sigma \text{EVENT1}) \multimap (\uparrow_\sigma \text{EVENT2}) \multimap \uparrow_\sigma$

Finally, the resultative is combined into the syntax by the c-structure rules in 10 and 11. These rules can only be used if the resultative interpretation is available to the verb, and create the extra XCOMP argument which takes the secondary predication.

(10)

V'	→	V	NP	$\{NP\|AP\|PP\}$
		↑=↓	$(\uparrow OBJ) = \downarrow$	$(\uparrow XCOMP) = \downarrow$
				$(\downarrow SUBJ) = (\uparrow OBJ)$
				$@RESULT-T((\uparrow PRED\ FN))$

(11)

V'	→	V	$\{NP\|AP\|PP\}$
		↑=↓	$(\uparrow XCOMP) = \downarrow$
			$(\downarrow SUBJ) = (\uparrow SUBJ)$
			$@RESULT-U((\uparrow PRED\ FN))$

When everything is combined, the resultative is created because the lexical entry permits the optional use of the resultative template, the resultative template calls in the extra verbal argument, and the c-structure rule provides the extra constituent. From the semantic side, the Glue equation combines the meaning of the subject with the meaning of the XCOMP and the meaning of the two events to create the overall meaning of the sentence.

References

[ADT08] Ash Asudeh, Mary Dalrymple, and Ida Toivonen. Constructions with lexical integrity: Templates as the lexicon-syntax interface. In Miriam Butt and Tracy Holloway-King, editors, *Proceedings of the LFG08 Conference*. CSLI Publications, 2008.

Incorporation and Complex Predication in Persian

Fatemeh Nemati
University of Tehran, University of Konstanz

Persian is a predominantly SOV language which displays a variety of multiword verbal expressions. This paper focuses on one such verbal expression, namely the N+V construction. Following Dabir-Moghaddam (1997) and Megerdoomian (2009), I propose a distinction between what I label *incorporation* and cases of so-called *complex predication*. Noun incorporation is here taken to refer to all cases of N+V compounds where the noun has an argument relationship to the verb, as shown in [1a] (Mithun 1984, Baker 1988). The term complex predicates, in contrast, is applied to constructions where more than one grammatical element contributes to the overall meaning of the complex and the argument structure of the complex results from the interaction of the contributing predicative elements, as illustrated in [1b] (Alsina et al., 1997; Butt, 1995, Mohanan 1994).

1a) āryā [Gazā]$_N$ [xord-Ø]$_V$
Arya food eat-Past.3.sg
'(lit.) Arya ate food.'

1b) āryā [kotak]$_N$ [xord- Ø]$_V$
Arya beating hit-Past.3.sg
'Arya was beaten.'

The contribution of this paper is threefold. First I survey the state-of-the-art for incorporation and complex predicates. After refining the definition of incorporation and adopting an argument merging analysis for complex predicates (Butt, 1995), I introduce a distinction between these two classes of syntactic multiword verbal expressions. Finally, I provide a c-structure and a-structure analysis of incorporation and complex predicates in Persian, drawing on the tools provided by the LFG framework.

The investigation of N-V constructions in Persian has taken either a lexical or a syntactic approach. Some researchers claim that all multiword verbal expressions are lexical and that they are the result of the morphological processes of incorporation and combination (Dabir-Moghaddam 1997, Vahedi-Langrudi 1996). However, a lexicalist approach falls short of explaining the syntactic behavior of both constructions. Other researchers discuss multiword verbal expressions as syntactic constructions, but they fail to observe the distinction between the two types of N+V sequences and take incorporation into account. (Karimi 1992, Folli et al. 2005). Megerdoomian (2009), in contrast, deals with the syntactic and semantic differences between these two N+V sequences, treating the nominal part under the term bare nominal as opposed to preverbal nominal (in my analysis incorporated noun and nominal part of the complex predicate, respectively). She does not deal with scrambling, topicalization, and relativization of preverbal noun, and these processes pose a challenge to a derivational analysis. The incorporation analysis I put forward, not only accounts for the bare nominals in Megerdoomian's analysis, but it can also be extended to include other multiword verbal expressions in Persian, such as PP+V and most of the Adv+V constructions, which are treated in the literature as complex predicates (Foli et al, 2005; Megerdoomian, 2009; Pantcheva, 2010).

What follows is a reanalysis of N+V sequences in the light of the LFG theory. The subcategorized noun in N+V sequences shows some of the properties that are usually associated with incorporation: a) With respect to stress pattern, the whole complex behaves like one unit and the nominal part receives the main VP stress; b) The whole complex can be nominalized (example 2); c) the noun is not available to bind a nominal, but the whole complex can bind a pronominal, as illustrated in [3a] and [3b] (Dabir-Moghaddam, 1997).

2. Gazā xordan-aS
food eating-his
(Lit.) 'His food-eating'

3a) *man Gazā$_i$ xordam va kami az ān$_i$ rā be gorbe dādam.
I food eat-past.1.sg. and some from it OM to the cat give-past.1.sg.
'I ate the food and gave some of it to the cat.' (examples adapted from Dabir-Moghaddam, 1997)

3b) man [Gazā xord-am]$_i$ va bad az ān$_i$ xābid-am
I food eat-past.1.sg. and after from it sleep-Past.1.sg.
(Lit.) 'I ate food and after that I slept.'

Inflectional morphology, however, rules out a lexical analysis of this N+V sequence: if they were the results of morphological compounding, the incorporated element would not be separated from the verb by the Negative marker (*ne-*), the Imperfective prefix (*mi-*), the Subjunctive prefix (*be-*) and the future auxiliary(xāstan). Therefore, I posit that Incorporation in Persian is syntactic, involving Semantic Incorporation and packaging of the event as a conceptual whole.

The claim is that incorporation encompasses all cases where a verb and its arguments are combined to make a conceptual whole. I follow an argument merger analysis as a diagnostic for the relationship between non-verbal and verbal elements. As a result, it can be shown that other Persian multiword verbal expressions, which have been treated unambiguously as cases of Complex Predicate by most scholars (Dabir-Moghaddam, 1997; Foli et al, 2004; Megerdoomian, 2009; Pantcheva, 2010), are in fact cases of incorporation which have undergone metaphorical extension. For instance, Folli et al (2004) treat [[*be bād*]$_{PP}$ [*dādan*]$_V$] (to wind give 'to lose') and [[*bālā*]$_{Adv}$ [ā*vardan*]$_V$](up bring 'to vomit') as cases of Complex

Predicates, while PP and Adv are internal argument and optional argument of the verb, respectively, and as such are cases of the incorporation of prepositional phrase and the adverbial with the verb.

To account for the differences between Incorporation and Complex Predicates, I avail myself of the multi-representational nature of the LFG analyses. Based on their syntactic behavior, I conclude that both constructions are syntactic, but they have different c-structure representations. Following King and Butt (1996) and in line with the literature on Persian complex predicates, I posit that the incorporated noun appears adjacent to the verb as the daughter of V' and sister of V with a non-specific interpretation [-restricted] functioning as OBJ, in contrast to the specific NP in Spec, VP, functioning as OBJ_{θ}. As there is no evidence to distinguish cases of Noun Incorporation from bare Noun and Verb juxtaposition, this analysis can equally account for both cases without introducing a new grammatical function (Asudeh and Ball ,2005; Duncan,2007). For the syntactic representation of Complex Predicates, I follow Mohanan (1996) by positing a verb complex that includes not only a light verb and an auxiliary, but also the nonverbal element of a Complex Predicate. That the noun in Complex Predicates cannot be generated in the same position as non-specific objects follows from the fact that Complex Predicates can also incorporate nouns (Megerdoomian, 2009).

mā tamām-e ruz otāG jāru mi-kard-im
we complete-EZ day room sweep IMP-do-Past-1PL]
'We swept rooms all day.'

The other difference between the two N+V sequences lies in the argument structure. Incorporation involves a flat argument structure where the verb and its argument merge for a joint predication, illustrated in representation [4]. Complex Predicates are assumed to have a complex argument structure as a result of argument merging, where light verb has a variable(%PRED) in its a-structure that is to be filled by a predicative element, as illustrated in [5] (Butt, 1995).

4) Gazā xordan 'food-eat < __ >'

5) kotak xordan 'beating-hit' <exp %Pred>
kotak<th>

Argument merging analysis of complex predicate formation proposes a new look at Persian N+V constructions, and shows that they can be divided up on the basis of the relationship between the nominal and the verbal element in the two categories of Incorporation and Complex predicates. Assuming that some of the multiword verbal expressions are cases of incorporation, some of the main verbs like "āmadan" (to come) will not be listed among the light verbs. Moreover, the researchers have studied the PP+V and Adv+V constructions as argument bearing predicative parts of complex predicates that contribute to the event structure of the whole complex (Foli et al, 2004; Megerdoomian, 2009; Pantcheva, 2010). I claim that these categories are not predicative but rather arguments of the verbal element.

References:

Alsina, J. Bresnan, and P. Sells (eds).1997. *Complex Predicates*. Stanford: CSLI Publications

Asudeh, Ash and Douglas Ball. 2005." Niuean incorporated nominals as non-projecting nouns". In *2005 LSA Annual Meeting*, Oakland, Ca.

Baker, Mark C. 1988. *Incorporation. A Theory of Grammatical Function Changing*. The University of Chicago Press, Chicago and London.

Butt, Miriam. 1995. *The Structure of Complex Predicates in Urdu*. Stanford: CSLI Publications.

Butt, Miriam and Tracy H. King. 1996. "Structural Topic and Focus without Movement". *Proceedings of the First LFG Conference*, XRCE, Grenoble

Duncan, Lachlan. 2007. "Analytic Noun Incorporation in Chuj and K'ichee' Mayan. In M. Butt and T.H. King, eds., *Proceedings of LFG07 Conference*, pp. 163-183. Stanford: CSLI Publications.

Dabir-Moghaddam, Mohammad. 1997. "Compound Verbs in Persian". *Studies in LInguistic Sciences*, 27, 25-59.

Folli, Raffaella, Heidi Harley, and Simin Karimi. 2005. Determinants of event type in Persian complex predicates. *Lingua* 115(10):1365–1401

Megerdoomian, Karine. 2002a. "Aspect in Complex Predicates". Talk presented at the Workshop on Complex Predicates, Particles and Subevents, Konstanz, October 2002.

Megerdoomian, Karine. 2006. "The Status of the Nominal in Persian Complex Predicates". available at http://www.zoorna.org/papers/preverbs.pdf

Mithun, Marianne. 1984. "The evolution of noun incorporation", *Language* 60: 847–894.

Mohanan, Tara. 1996. "Multidimensionality of Representation: NV Complex Predicates in Hindi". In *Complex Predicates*. A. Alsina, J. Bresnan and P. Sells, eds. Stanford: CSLI Publications.

Pantcheva, Marina . 2010. "First Phase Syntax of Persian Complex Predicates: Argument Structure and Telicity" in *Journal of South Asian Linguistics*, V. 2, pp. .

Vahedi-Langrudi, Mohammad-Mehdi. 1996. The Syntax, Semantics and Argument Structure of Complex Predicates in Modern Farsi. Ph.D. thesis, University of Ottawa.

On snakes, locative binding and complex predicates

György Rákosi
University of Debrecen, Hungary http://ieas.unideb.hu/rakosi

1. Background Coreference marking in English locative PPs can be achieved either via the use of a pronominal or a reflexive form (1). In the standard LFG analysis, the non-complementarity of the pronoun and the anaphor is captured by assuming that there is an asymmetry in the binding theoretically relevant domain of the two elements (Bresnan et al. 1985, Bresnan 2001, Dalrymple 1993, 2001; cf. also Büring for a similar approach outside of LFG). Reflexives are +NUCLEAR in the sense that they must find an antecedent within the *minimal complete nucleus*, i.e., the smallest f-structure that contains the f-structure of the anaphor and a SUBJ. Pronouns are –NUCLEAR in the sense that they are constrained to be disjoint from their coarguments. The coargument domain is defined by the PRED feature, and this domain need not include a SUBJ. Therefore locative PPs, being predicative, will constitute a negative binding domain for pronouns, but, lacking a SUBJ, they will not constitute a (positive) binding domain for anaphors.

This analysis rests, among others, on three important assumptions. First, binding constraints are lexically associated with anaphoric/pronominal elements (in the general LFG-spirit). Second, the locative PP in question can in principle be an argument or an adjunct (see, for example, Dalrymple 2001: 280, and Lødrup 2007 for a specific discussion of this issue). Third, what matters for Binding Theory is whether the P-element itself is predicative or not. A P-element is arguably not predicative if it is directly selected by the verb (possibly as part of a larger idiomatic unit) and lacks independent semantic content. In such cases, only the anaphor can encode coreference (cf. 2).

2. The problem It is also quite well-known, however, that the cross-linguistic facts are more complex. In particular, languages differ wrt whether they only *allow* or *require* a reflexive strategy to apply in locative PPs (see Faltz 1985 for an overview). For example, in the German equivalent of (1), only the reflexive element is acceptable and the pronoun is ungrammatical (3). In the above described approach, the German facts can be explained by either of the following two ways. It could be assumed that the German pronoun (*ihm*) is associated with binding constraints in the lexicon that differ from those of the English pronoun *him*. For example, *ihm* can be taken to be –NUCLEAR in the sense that it cannot be bound in the *minimal complete nucleus* (i.e., there is no domain asymmetry in German between *ihm* and *sich*). Second, it can be argued that the constructions in the two languages are in fact not equivalent.

Reuland (2001, 2006) makes use of this second strategy to explain the difference between the French (4a) and the Dutch (4b): a corefering pronominal is licensed only in the former case (note that his particular examples do not represent locative contexts). Dutch licenses preposition stranding, which Reuland interprets as the sign of the covert reanalysis of the P-element with verb. The result is a covert complex predicate V-P, and only one binding domain for (4b). French does not allow preposition stranding, hence there is no covert complex predicate formation. The French pronoun survives in (4a) because the preposition does not incorporate into the verb and no complex predicate is formed.

This account does not readily explain the difference between the English (2a) and the French (4a), for in both cases we have what looks like a semantically empty P. This is a contrast that I will not explain here, and I also remain agnostic about the general feasibility of the covert complex predication formation analysis of Dutch. My aim in this paper is to scrutinize the delicate pattern of coreference marking in Hungarian PPs, and to show that the data can be explained within standard LFG-theoretic assumptions under recognition of the binding theoretic relevance of *overt* P-V complex predicate formation.

3. The Hungarian facts

Hungarian has two different types of postpositions (one set takes case-marked complements, the other takes caseless complements), plus a handful of locative case suffixes. All these P-elements used to be possessive structures historically, which origin has become obscure to different degrees. As a synchronic reflex of this etymology, the pronominal form of case suffixes and that of postpositions taking caseless complements is formally identical to possessive structures (with the possessor being *pro*-dropped). See (5).

At first sight, Hungarian seems to pattern up neatly with German, and not with English, inasmuch as it normally does not allow 3SG coreferential pronouns in locative PPs (6a). However, as (6b) testifies, pronominal coreference becomes an option in first and second person (there is some variation across speakers, but everyone finds a clear contrast between (6a) and (6b)). Notice that in (6b) the inflected PP is in the postverbal domain.

Directional postpositions/case suffixes by default occupy an immediately pre-verbal position. When this happens, pronominal coreference marking becomes very marginal or unacceptable (7a). If however, the (first or second person) pronominal P is a postverbal associate of an incorporated adverbial, as in (7b), then coreference becomes grammatical. I will show that this pattern is pervasive: pronominal marking of coreference in Hungarian PPs is best if the PP itself does not occupy the immediately preverbal position.

4. An explanation of the Hungarian facts

These data raise two immediate questions. First, why do we have the binding theoretically relevant difference between third and non-third person pronominal PPs in Hungarian? Second, what is the actual relevance of the immediately preverbal position (known as a host of *verbal modifiers* in Hungarian grammars) for Binding Theory?

I will argue that the answers to these two questions are interrelated. When a PP licenses pronominal coreference, its structure is actually reanalyzed as a possessive construction: (5b) or (5c) may reactive the underlying (historical) possessive structure, and they synchronically become more similar to a real possessive (5a). Informally, the PP *alattam*

'under me' is reanalyzed as 'under my place', and (6b) is in fact interpreted as 'I saw a snake under my place [i.e. the place associated with me].' Coreference is then between the possessor (represented by agreement morphology) and the subject antecedent. See (8) as an illustration for a simplified f-structure of (7b).

I will show that sometimes there is overt morphological evidence of this possessive reanalysis. I will also argue that it is blocked in 3SG because there is independent evidence that third person possessors do not agree with the possessum, and this lack of agreement precludes the possibility of possessive reanalysis in the PP cases.

Finally, possessive constructions are known not to be able to occupy the preverbal position in neutral sentences in Hungarian. This explains the ungrammaticality of (7b): the pronominal PP could be coreferent with the (*pro*-dropped) subject only as a reanalyzed possessive structure, but as such, it cannot occupy the preverbal position. It follows that only free pronominal PPs (which need not be reanalyzed as possessives) may occur preverbally, which is in fact the case (not shown). In fact, preverbal occurrence in neutral sentences will be analyzed as an instance of P-V complex predicate formation (and concomitant predicate composition in semantic structure), and whenever this happens, the incorporated PP cannot have a clause-mate antecedent (as suggested in Reuland 2006 for the Dutch (4b)).

All in all, Hungarian *is* like German: regular pronouns cannot code clause-internal coreference in PPs, *except* when they are reanalyzed as possessive structures. Such reanalysis is not available in German. What this account does not explain is why English differs from German (and from Hungarian) in allowing coreferent *simple* pronouns in PPs, which is a problem that needs an independent explanation.

(1) a. *John*$_i$ *saw a snake beside him*$_{i/k}$ / *himself*$_i$.

(2) a. *John*$_i$ *believes in him*$_{*i/k}$ / *himself*$_i$.
b. *John was beside *him/ himself with rage*.

(3) a. *Hans*$_i$ *sah eine Schlange neben ihm*$_{*i/k}$ / *sich*$_i$. German
Hans saw a snake beside him self

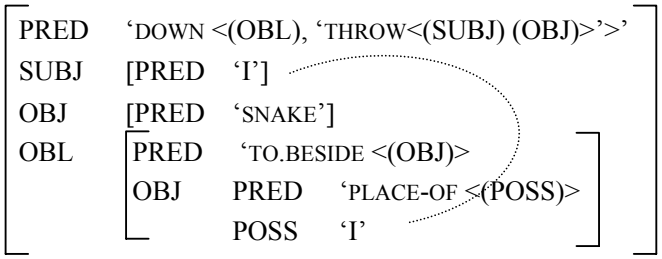
(4) a. *Jean*$_i$ *parle de lui*$_{i/k}$ / *lui-même*$_i$. French (Reuland 2006:65)
'Jean talks of him/himself.'
b. *Jan*$_i$ *praat over *zich*$_i$ / *hem*$_{*i/k}$ / *zichzelf*$_i$. Dutch
'Jan talks of himself/him.'

(5) a. *ház-am* [possessive] b. *nál-am* [case suffix] c. *alatt-am* [postposition]
house-1SG at-1SG under-1SG
'my house' 'at me' 'under me'

(6) a. *János*$_i$ *látott egy kígyó-t maga mellett*$_i$ / **mellett-e*$_i$
John saw a snake-ACC himself beside beside-3SG
'John saw a snake beside himself.'
b. *Látt-am*$_i$ *egy kígyó-t mellett-em*$_i$.
saw-1SG a snake-ACC beside-1SG
'I saw a snake beside me.'

(7) a. */$^{??}$*Mellé-m-dobt-am a kígyó-t*.
to.beside-1SG-threw-1SG the snake-ACC
'I threw the snake beside me.'
b. *Le-dobt-am a kígyó-t mellé-m*.
down-threw-1SG the snake-ACC to.beside-1SG
'I threw the snake down beside me.'

(8) (→7b)



BIBLIOGRAPHY

Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Oxford: Blackwell. • **Bresnan, Joan; Kanerva, Jonni M & Maling, Joan**. 1985. Logophoricity and bound anaphors. Ms. Stanford University. • **Büring, Daniel**. 2005. *Binding Theory*. Cambridge: CUP. • **Dalrymple, Mary**. 1993. *The Syntax of Anaphoric Binding*. Stanford: Center for the Studies of Language and Information. • **Dalrymple, Mary**. 2001. *Lexical Functional Grammar*. San Diego: Academic Press. • **Faltz, Leonard**. 1985. *Reflexivisation: A study in universal syntax*. New York: Garland. • **Lødrup, Helge**. 2007. A new account of simple and complex reflexives in Norwegian. *The Journal of Comparative Germanic Linguistics* 10 (3). 183-201. • **Reuland, Eric. 2001.** Primitives of Binding. *Linguistic Inquiry* 32 (2). 439-492. • **Reuland, Eric. 2006.** Gradedness: interpretive dependencies and beyond. In G. Fanselow, C. Féry, M. Schlesewsky & R. Vogel eds. *Gradience in Grammar. Generative Perspectives*. Oxford: OUP. pp. 45-69. •

Right Node Raising in Parsing and Generation

Christian Rohrer Jonas Kuhn Sina Zarrieß
University of Stuttgart, IMS, Germany

We present an implementation of Right Node Raising (RNR) in German in the framework of LFG. Like other types of non-constituent coordination, RNR is often considered notoriously difficult and ignored in grammar writing – although instances of it do occur in real text. Our analysis is inspired by the "rule splitting" technique of (Maxwell and Manning, 1996), but it does not leave the formal framework of standard LFG and at the same time requires few special-purpose rules, as it builds on the regular mechanisms implementing German clausal syntax.

Informally speaking, the term 'Right Node Raising'[1] refers to two coordinated clauses which lack some elements in the first conjunct. The missing parts in the first conjunct have obligatory phonetically overt counterparts in the same structural position in the second conjunct (Féry and Hartmann, 2005). Our implementation of RNR accounts for the following raised constituents: DPs, PPs and ADVPs (example 1), infinitival constructions (example 2), V-final clauses (example 3).

(1) Hans kauft und Maria verkauft Aktien.
Hans buys and Maria sells shares.

(2) Hans versuchte gestern und Emil probiert heute, das
Hans tried yesterday and Emil attempts today, the
Problem zu lösen.
problem to solve.

(3) Hans bedauert und Emil begrüßt, daß Maria geht.
Hans regrets and Emil welcomes, that Maria leaves.

The raised constituent may be "extracted" from different levels of embedding (example 4) and several constituents may be raised (example 5).

(4) Wulff sagte heute und Merkel soll morgen bestätigen,
Wulff said today and Merkel should tomorrow confirm,
daß die Steuern nicht gesenkt werden.
that the taxes not lowered be.

(5) Hans versprach seiner und Eva versprach ihrer Mutter, zu
Hans promised his and Eva promised her mother, to
kommen.
come.

In our analysis, the raised material is adjoined to the coordination of two sentences.

```
(a) ROOT --> CProot: ! $ ^;  and
             CProot: ! $ ^;
             RNR.
(b) RNR --> {DP*|PP*|VP|...}.
```

RNR specifies the possible raised constituents. The raised constituent must be optional in the rule which introduces it. The coordination rule is the familiar LFG-rule for same-constituent coordination. The functional information of the raised contituent(s) is distributed over the set of elements. For distribution to be possible, RNR has to be adjoined to CProot and cannot be part of CProot. The raised constituents are annotated roughly with the function they have in the position from which they are 'extracted'; with some exceptions: For instance, subjects cannot be RNR-ed. The annotation of the RNR-ed constituent follows the same principles as the annotation of topicalized constituents like the relative pronoun in example (9).

(6) Ich kenne den Roman, den Maria schreibt und Erich
I know the novel, which Maria writes and Erich
liest.
reads.

Problems with overgeneration: The implementation sketched so far overgenerates. In generation-mode we get, for instance, the following surface realisation for example (3).

(7) Gestern versuchte Hans und Emil probiert heute, das
Yesterday tried Hans and Emil attempts today, the
Problem zu lösen.
problem to solve.

In this surface realization, the syntactic parallelism of the original sentence (3) is destroyed. The analysis presented so far does not capture two basic features of RNR (Féry and Hartmann, 2005): (a) the raised material is extracted from the right periphery of the first conjunct, (b) the two conjuncts must exhibit a parallel syntactic and focus structure. In order to account for conditions (a) and (b) we introduce a new non-terminal category 'RightPeriphery' into the Middle Field (MF) and use linear precedence rules to enforce its final position in the MF. Furthermore we have to make sure that the constituents in the right periphery have the same function in both conjuncts.[2] Therefore, we introduce a discourse function RNR-FOCUS.

[1] We use the terms 'Right Node Raising' and 'raised constituent' despite the fact that in LFG there is no "extraction" rule.

[2] The grammar also parses sentences like (i) where the DP is 'extracted' from inside a VP.

(i) Der Mitarbeiter wird [DP ...] verfassen und der Chef wird
The colleague will [DP ...] write and the boss will
[DP ...] unterschreiben den Bericht.
[DP ...] sign the report.

```
RightPeriph -->
          e:(^RNR-FOCUS)= ! ;
          { ADVP[std]:! $ (^ADJUNCT)
          | DP[std]: ! = (^ {SUBJ|OBJ|OBJ-TH})
          |...}.
```

We use the term RNR-FOCUS because the constituent dominated by 'RightPeriphery' is Focus-marked. Phonetically speaking, it receives a pitch accent. To capture the notion of functional parallelism, we add a schematic constraint to rule (a) such that the RNR-FOCUS originates from the same grammatical function in both conjuncts. [3]

Syntactic Parallelism: One could argue that sentence (7) is not ungrammatical. We believe that the best way to capture parallelism is by a soft constraint. XLE's log-linear disambiguation component already provides a c-structure feature template 'cs_conj_nonpar' which counts non-parallel conjuncts within depth levels. A similar template could be added for f-structure parallelism.

(Féry and Hartmann, 2005) observed that RNR-construction exhibit a parallel focus structure. If one makes this focus structure accessible to syntax along the lines proposed by (Bögel et al., 2009), then the Principle of Prosodic Preference would prefer RNR-constructions where syntactic constituent boundaries coincide with prosodic boundaries.

Discussion: Maxwell and Manning (1996) propose a treatment of non-constituent coordination requiring a modification of c-structure parsing: coordination may pertain to an incomplete subspan of a c-structure rule's right-hand side. In principle, the LFG parsing algorithm can be modified with a stack-driven mechanism that will allow for coordination of partial constitutents, otherwise following standard LFG assumptions about coordination. The stack is used to keep track of the position in the rule's right-hand side up to which c-structure material was included/excluded in the coordination. To our knowledge, the mechanism has never been implemented, apart from (Zarrieß and Seeker, 2008), who propose a finite-state based rule compilation mechanism that can be combined with XLE.

In spirit however, Maxwell and Manning's analysis can be captured well with moderate extensions applied to a standard broad-coverage grammar of German. This is because for the typical cases of non-constituent coordination, detailed c-structure-level book-keeping over the position up to which material has been included/excluded in the coordination is not required: To capture argument and adjunct placement in the German *Mittelfeld*, no hard grammatical constraints are assumed. Simplifying somewhat, a binary right-branching rule, or, as Forst and Rohrer (2009) propose to capture coordination facts, a flat rule VP → XP* VC can be assumed (where VC is the verbal complex). In addition, Forst and Rohrer (2009) assume an "artificial" category VPargs (on the left edge of VC), which can span two or more of the XP arguments/adjuncts, excluding the verbal complex. By allowing for coordination of this VPargs category, typical conjunction reduction cases (like *John gave Mary apples and Sue bananas*) can be captured. The unconstrained span of XP* in the VP on one hand, and VPargs on the other generates the necessary options for the conjunction reduction phenomenon and is at the same time constrained at the level of f-structure. RNR is not captured by this analysis since it involves argument/adjunct material outside the coordination. But in the present paper, we propose the addition of a dual "artificial" category for this non-coordinated material, the RNR category, which again nicely combines with a flexible-span coordination.[4]

In the RNR analysis, additional f-structural constraints are needed to capture the parallelism constraint on the conjuncts. Note that this would be required in Maxwell and Manning's formalism too, and as we pointed out above, there is clear information structural evidence motivating this.

[3] It is parallelism of f-structure not c-structure parallelism which is required here. Consider example (i) where 'heute' and 'am Montag' belong to different syntactic categories but have the same funtion:

(i) Er kaufte heute und sie kaufte am Montag ein Auto.
He bought today and she bought on Monday a car.

[4] We note that the RNR category can be viewed as a generalized *Nachfeld* category, attached at a high structural level, such that it can distribute over a complete coordinated sentence. Normally, one would not allow for arbitrary functions for a *Nachfeld* category (since extraposed arguments are extremely rare and an extreme amount of local ambiguity is added), but in its interaction with coordination, the construction is sufficiently constrained.

References

Bögel, Tina, Miriam Butt, Ronald M. Kaplan, Tracy Holloway King, and John T. Maxwell III. 2009. Prosodic Phonology in LFG: A New Proposal. In *Proceedings of the LFG09 Conference*.

Forst, Martin, and Christian Rohrer. 2009. Problems of German VP Coordination. In *To appear in the Proceedings of the LFG '09 Conference*, Cambridge, England.

Féry, Caroline, and Katharina Hartmann. 2005. Focus and prosodic structures of German Gapping and Right Node Raising. *The Linguistic Review* 22:67–114.

Maxwell, John, and Christopher Manning. 1996. A Theory of Non-constituent Coordination based on Finite State Rules. In *Proceedings of the First LFG Conference*, Grenoble, France.

Zarrieß, Sina, and Wolfgang Seeker. 2008. Finite-State Rule Deduction for Parsing Non-Constituent Coordination. In *FSMNLP 6th International Workshop,*, Potsdam.

Applicativizing complex predicates: a case study from Murrinh-Patha

Melanie Seiss
University of Konstanz

Rachel Nordlinger
University of Melbourne

Murrinh-Patha, a polysynthetic language from the Northern Territory of Australia, is like many northern Australian languages in having a bipartite verbal system, in which one of a limited set of classifier stems combines with a lexical stem to form a complex predicate (see e.g Wilson 1999, Schultze-Berndt 2000, McGregor 2002, Bowern 2004 for discussion of related phenomena in other Australian languages). In this paper we present an analysis of these complex predicates within LFG that both accounts for the patterns of combination found in the data, and the interaction of these complex predicates with derivational processes such as applicativization and reflexive/reciprocal constructions. While there is a significant body of work investigating the analysis of complex predicates in LFG (see for example, Mohanan 1994, Butt 1995, Alsina 1996, Alsina et al 1997, Andrews and Manning 1999, Wilson 1999), very little of this work has dealt with polysynthetic languages like Murrinh-Patha and the interaction of complex predicates with other valency-changing morphological processes. This research thus brings new data into the discussion of complex predicates in LFG, and extends this discussion into a new typological domain.

Examples of Murrinh-Patha complex predicates are given in (1), in which we see the classifier stem (traditionally glossed with a number) followed by the lexical stem.[1] Each of these stem types can independently form other combinations – (1a) and (1b) show the same classifier stem (BE(4)) co-occurring with two different lexical stems, and (1c) and (1d) show the same lexical stem combining with two different classifier stems. A minority of classifier stems can function alone as a clausal predicate, all other classifier stems must always combine with a lexical stems. Lexical stems can never occur as the sole clausal predicate, and are only ever found in combination with a classifier stem.

(1a) *kanam-kaykay*
3sS.BE(4).nFut-call_out
'he continually calls out'

(1b) *nganam-kut*
1sS.BE(4).nFut-collect
'I collected (the money).'

(1c) *mam-kurrk*
1sS.HANDS(8).nFut-scratch
'I scratched something.'

(1d) *mem-kurrk*
1sS.HANDS:RR(10).nFut-scratch
'I scratched myself.'

Such data is interesting theoretically since the argument structure (and semantics) of the complex predicate as a whole is derived from the composition of its component parts, and accurately capturing how this composition works systematically is not always straightforward. In Murrinh-Patha, the problem becomes more interesting since these complex predicates also combine with various other types of verbal morphology that also affect argument structure. The examples in (2) illustrate the applicative *-ma-* which promotes a source to the function of direct object (as evidenced by the object agreement on the verb):

(2a) *nganam-**nhi-ma**-kut*
1sS.BE(4).nFut-**2sO-APPL**-collect
'I collected (the money) from you.'

(2b) *mangan-**nhi-ma**-art* *kura*
1sS.SNATCH(9).nFut-**2sO-APPL**-get water
'I got (some) water from you.'

Since applicativization is a valency-changing process, such examples have significant implications for an analysis of Murrinh-Patha complex predicates: namely, over what part of the complex predicate does the applicativization process apply?

On our analysis, both classifier stems and lexical stems contribute (partial) argument structures that combine to produce the argument structure for the complex predicate as a whole. Classifier stems are either intransitive (1a), transitive (1c) or reflexive/reciprocal (1d). Their argument structures specify argument slots, but are underspecified for thematic roles. Lexical stems, on the other hand, contribute the thematic roles and are either intransitive (1a) or transitive (1b, c, d).

A sketch of the analysis of some simple cases (for examples (1)) is presented in (3):

(3a) BE(4) <x>
kaykay, 'call out'<agent>
=> BE-*kaykay* <agent>

(3b) BE(4) <x>
kut, 'collect' <agent, theme>
=>BE-*kut* <agent, theme>

(3c) HANDS(8) <x, y>
kurrk, 'scratch' <agent, theme>
=> HANDS(8)-*kurrk* <agent, theme>

(3d) HANDS:RR(10) <x_1, y_1>
kurrk, 'scratch' <agent, theme>
=> HANDS:RR(10)-*kurrk* <$agent_1$, $theme_1$>

While we find transitivity matching in most cases, intransitive classifier stems can also combine with transitive lexical stems in which the intransitive classifier stem provides information about the posture of the agent and/or tense and aspect.

[1] Unless otherwise specified, examples are taken from author(s) unpublished fieldnotes.

We assume the following basic argument structure for the applicative in which the applicative adds a source argument to the already existing argument structure of the verb. In contrast to other applicative constructions, e.g. the Indonesian applicative *-i* (Arka et al 2009), no evidence for a co-indexing of the arguments is present in Murrinh-Patha.

$$\text{V - Appl} < \text{V} < \theta_{1, \ldots} \theta_{n} > \quad \text{Appl} < \text{source} >>$$

This basic argument structure accounts straightforwardly for the examples in (2) by assuming that the applicative operates over the complex predicate as a whole. It is clear that the applicative cannot operate on the classifier stem alone since, in a case like (2a), the intransitive classifier would first be made transitive, as in (4). However, as we have seen above, the combination of a transitive classifier with a transitive lexical stem results in a 2-place predicate (as in 1c), not the 3-place predicate we find in the applicative construction:

(4) BE(4) – Appl <x, y(source)>
kut, 'collect' <agent, theme>

Our analysis also extends naturally to account for examples such as (5), in which the applicative combines with a reflexive/reciprocal classifier:

(5a) *ngennham*
nhem-nham
1sS.POKE:RR.nFut-fear
'I'm afraid.'

(5b) *ngennhimanham*
nhem-**nhi-ma**-nham
1sS.POKE:RR.nFut-**2sO-Appl**-fear
'I'm afraid of you.'

As pointed out above, RR classifier stems specify a co-indexing of two arguments provided by the lexical stem. In (5b) the theme argument is co-indexed with the experiencer while the source remains unindexed. In order to ensure the correct co-indexing, the applicative must operate on the combined complex predicate (6).

(6) POKE:RR-*nham*<experiencer$_1$, theme$_1$>
=> POKE:RR-*nham*-Appl<experiencer$_1$, theme$_1$, source>

This applicativized complex predicate can also undergo a second reflexive/reciprocal process, marked by the RR marker *-nu-*, as shown in (7):

(7a) *them-**nu-ma**-nham*
1incS.POKE:**RR**(21).nFut-**RR-APPL**-fear.
'We're (inclusive) frightened of each other.'

(7b) POKE:RR(21)-*nham*<experiencer$_1$, theme$_1$>
=> POKE:RR(21)-*nham*-Appl<experiencer$_1$, theme$_1$, source>
=> POKE:RR(21)-*nham*-Appl-RR<experiencer$_1$, theme$_1$, source$_1$>

Thus, the argument structure approach presented can explain the applicativising process in Murrinh-Patha and its interaction with complex predicate formation and reflexivisation/reciprocalisation. Note that, in contrast to other applicative constructions like e.g. the Chichewa applicative (Lam 2007), the linear order does not reflect the order in which the valency-changing processes apply.

Building on Alsina & Mchombo's (1993) account of applicatives in Chichewa, and following Butt (2006) in applying [+o] to non-agent theta roles, we show how our approach provides a unified analysis of the full range of complex predicates in Murrinh-Patha, and especially their interactions with other valency-changing processes. Thus we extend the research on complex predicates into the domain of polysynthetic languages, and showcase the strengths of the LFG framework in analysing typological diversity.

References

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

Alsina, Alex and Sam Mchombo. 1993. Object asymmetries and the Chichewa applicative construction. In Mchombo, Sam (ed). *Theoretical Aspects of Bantu Grammar.* Stanford: CSLI Publications, 1-44.

Arka, I Wayan, Mary Dalrymple, Meladel Mistica, Suriel Mofu, Avery Andrews and Jane Simpson. 2009. A linguistic and computational morphosyntactic analysis for the applicative *-i* in Indonesian. In M. Butt and T.H. King, eds., *LFG Online-Proceedings of the 2009 Lexical-Functional Grammar Conference, Cambridge.*

Alsina, Alex, Joan Bresnan and Peter Sells (eds). 1997. *Complex Predicates*. Stanford: CSLI Publications.

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

Bowern, Claire. 2004. *Bardi Verbal Morphology in Historical Perspective*. Ph.D. Thesis, Harvard University.

Butt, Miriam. 2006. *Theories of Case*. Cambridge: Cambridge University Press.

Lam, Olivia. 2007. The applicative affix and morpheme ordering in Chichewa. In M. Butt and T.H. King, eds., *LFG Online-Proceedings of the 2007 Lexical-Functional Grammar Conference, Stanford.*

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

Schultze-Berndt, Eva. 2000. *Simple and Complex Predicates in Jaminjung*. Nijmegen: University of Nijmegen.

Wilson, Stephen. 1999. *Coverbs and Complex Predicates in Wagiman.* Stanford: CSLI Publications.

RELATIONAL-REALIZATIONAL PARSING

REUT TSARFATY
UPPSALA UNIVERSITY

Statistical parsing models aim to assign accurate syntactic analyses to natural language sentences based on the patterns and frequencies observed in human-annotated training data. State-of-the-art statistical parsers to date demonstrate excellent performance in parsing English, but when the same models are applied to languages different than English, they hardly ever obtain comparable results. In this talk I present a new parsing model, called the *Relational-Realizational (RR)* model [11], that is designed to effectively cope with parsing languages that allow for flexible word-order and richer morphological marking. The model is developed based on two principles: the first is *form-function separation* in the representation of constituents, and the second is *typological decomposition* of the syntactic spell-out rules. I show an application of the RR model to parsing the Semitic language Modern Hebrew, obtaining significant improvements over previously reported results.

Many state-of-the-art statistical parsing models to date have been developed with English data in mind, utilizing the Wall-Street Journal Penn treebank [5] as the primary, and often the sole, resource. However, English is quite unusual in its fairly *configurational* character [2]. The main challenge associated with parsing languages that are *less configurational* than English, such as German, Arabic, Hebrew or Warlpiri, is the need to model and to statistically learn complex correspondence patterns between functions, e.g., sets of abstract grammatical *relations*, and their morphological and syntactic forms of *realization*.

Whereas grammatical relations are largely universal, realization is known to vary greatly. Different means of realization involve the interaction of (at least) two typological parameters, one associated with word order [4], and another associated with word-level morphology [8, 3]. In order to adequately model complex form-function correspondence patterns that emerge from such interactions, I firstly consider morphological models that map grammatical properties of words to the surface formatives that realize them, and I follow up on recent studies in singling out the paradigmatic, realizational approach as an adequate strategy for modeling complex form-function correspondence [1]. I then extend the *inferential, realizational* principles of mapping grammatical properties to surface words [10] to *relational, realizational* principles of mapping grammatical relations to surface constituents.

In the resulting RR model, constituents are organized into syntactic paradigms [6]. Each cell in a paradigm separates the function, a Relational Network [7] and a set of grammatical properties, from the form of the individual constituent. The form of a constituent in a specific paradigm cell emerges from the (i) internal grouping, (ii) linear ordering, and (iii) morphological marking of its subconstituents. The RR decomposition of spell out rules for specific paradigm cells into parameters thus separates the *functional, configurational* and *morphological* dimensions. Subconstituents may belong to different paradigms and may

be associated with relational networks of their own, and the process continues recursively until fully-specified morphosyntactic representations map to words. This 3-phased spell-out gives rise to a recursive generative process that can be used as a probabilistic model and its probabilistic parameters can be estimated from data based on relative frequencies.

The resulting statistical model is empirically evaluated by parsing sentences in the Semitic language Modern Hebrew on the basis of a small annotated treebank [9]. Through a series of experiments we report significant improvements over the state-of-the-art Head-Driven (HD) alternative on various measures, without paying any computational costs. The typological characterization of the RR statistical distributions also suggests itself as a starting point for the development of quantitative methods that would facilitate typological classification learned directly from natural language corpus data.

Acknowledgements

The research reported in this talk constitutes a part of a 4-year project entitled "Integrated Morphological and Syntactic Ambiguity Resolution for Modern Hebrew" (principal invetigator: Reut Tsarfaty) funded by the Dutch Science Foundation (NWO), grant number 017.001.271. This work was supervised by Dr. Khalil Sima'an and Prof. Remko Scha.

References

[1] Stephen R. Anderson. *A-Morphous Morphology*. Cambridge University Press, 1992.
[2] Joan Bresnan. *Lexical-Functional Syntax*. Blackwell, 2000.
[3] Joseph H. Greenberg. A quantitative approach to the morphological typology of language. In R. F. Spencer, editor, *Method and Perspective in Anthropology*. University of Minessota Press, Minneapolis, 1954.
[4] Joseph H. Greenberg. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg, editor, *Universals of Language*, pages 73–113. MIT Press, 1963.
[5] Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. Building a Large Annotated Corpus of English: The Penn Treebank. *Computational Linguistics*, 19:313–330, 1993.
[6] Kenneth L. Pike. Dimensions of grammatical constructions. *Language*, 38(3):221–244, 1962.
[7] Paul M. Postal and David M. Perlmutter. Toward a universal characterization of passivization. In *BLS 3*, 1977.
[8] Edward Sapir. *Language: An Introduction to the Study of Speech*. Brace and company, New York, 1921.
[9] Khalil Sima'an, Alon Itai, Yoad Winter, Alon Altman, and Noa Nativ. Building a Tree-Bank for Modern Hebrew Text. In *Traitement Automatique des Langues*, 2001.
[10] Gregory T. Stump. *Inflectional Morphology: A Theory of Paradigm Structure*. Number 93 in Cambridge Studies in Linguistics. Cambridge University Press, 2001.
[11] Reut Tsarfaty. *Relational-Realizational Parsing*. PhD thesis, University of Amsterdam, 2010.

Reversing F-Structure Rewriting for Generation from Meaning Representations

Sina Zarrieß Jonas Kuhn
University of Stuttgart, IMS, Germany

1. Introduction

The standard meaning construction approach for LFG grammars implemented in the XLE framework is the transfer semantics system developed by (Crouch and King, 2006). Their system exploits the XLE transfer module to map LFG f-structures to shallow meaning representations on the basis of an ordered list of term-rewrite rules. A commonly mentioned disadvantage of this transfer approach is the fact that the XLE rewrite rules are not reversible (Crouch and King, 2006).

In this paper, we present a method to automatically induce a transfer grammar which maps meaning representations back to f-structures. The method converts sequences of seen instances of "forward" transfer rules into their approximated "backward" counterparts. The reversal method can be applied to any comparable transfer semantics.The ultimate goal of this work is to have an LFG-based generation system that can generate more grammatical paraphrases than f-structure based generation (e.g. voice alternations). We basically pursue the same strategy as former work in LFG-based generation (e.g. Cahill et al. (2007)), where all possible linguistic realisations of an abstract input representation are first generated and the final output sentence is then selected by a realisation ranking model which is based on linguistic experience and/or trained on a corpus.

2. Main Idea

Intuitively, it is problematic to reverse the f-structure rewriting because the XLE transfer rules can arbitrarily delete information from the input f-structure. As an example, the topmost f-structure in figure 1 contains features like CASE, PERS, or TOPIC which do not appear at all in the meaning representation. Moreover, the node embeddings in meaning representation differ from the f-structure so that we have to be very careful to reintroduce the right f-structure nodes in the reverse mapping.

One of the main ideas of our approach is that we do not need to generate full-fledged f-structures from the meaning representations because the XLE generator can handle underspecified input (Crouch et al., 2004). If the generator does not find a CASE feature in a place where it would expect one, it can freely add all possible values of that feature. The output of the f-structure generation is the complete set of sentences compatible with its (more or less abstract) input and the constraints encoded in the grammar and lexicon.

Thus, the main question is whether we can split up the f-structure features into classes of (a) core meaning features and (b) syntax-internal features; such that only the former need to be present in the semantic input representation. Of course, in principle, it cannot be guaranteed that all the core features are recoverable because the transfer rules may perform arbitrary deletions on the input. However, empirically, our experiments confirm the intuition that the meaning representation generally preserves information about a) the "lexical" features, such that PRED values and argument functions can be reconstructed, b) the semantic features like tense of a verb or definiteness of a noun. An interesting exception to b) is a class of implicit default features in the semantics which are explicit in the f-structure. For instance, in the meaning representation, the clause type is only marked if it is not declarative, whereas the f-structure always marks the clause type. This means that the clause type feature cannot (automatically) be reconstructed for declarative sentences - but if we leave this feature underspecified in the f-structure input, we will always generate questions and declarative sentences for a declarative semantic input. Currently, we deal with this problem by manually configuring the XLE generator in a way such that it only adds certain default values for a class of features which are not specified in the semantic input.

Our implementation reduces the task of inducing a mapping from meaning representations to f-structures to two subproblems: 1) determine the set of features that can actually be recovered from the original f-structure 2) determine the rewrite sequence between a recoverable f-structure term and its corresponding term in the meaning representation. Both of these subproblems cannot be solved by just looking at the "forward" transfer grammar. The reason is that the exact application of the transfer grammar on a given f-structure depends on the "feeding and bleeding" of the single rules. However, given a concrete instance of a <f-structure, meaning representation> pair defined by the transfer mapping, one can exactly tell which f-structure features have been deleted and which features have a correspondence in the meaning representation (in section 4, we will define what "correspondence" formally means). Moreover, it is possible to run the XLE transfer system in a mode that outputs each rule application of the rewrite process. This means that we will be able to recover the sequence of rules which mapped a concrete term in the f-structure to a term in the meaning representation.

The algorithm we implemented approximates the reverse of a given "forward" transfer grammar by operating on instances of transfer rules which have been encountered in a concrete rewrite sequence. If we run it on a large set of <f-structure, rewrite sequence, meaning representation> triples (coming from testsuites or corpus sententences) and take the union[1] of the reversed rules, we can expect all the relevant rules to be recovered. In this way, the method is generally applicable to comparable transfer grammars.

[1] Technically, alternative rules for the same input are allowed by making the rewrite rule applications optional (or at least a subset of them).

3. Example

Once we have induced the reverse transfer grammar, the generation system works as is illustrated in figure 1.

First, an input corpus sentence is parsed and mapped to a flat semantic representation. Note that the subject of the passive f-structure is mapped to a "semantic object" in the meaning representation. If our reversal method has encountered rewrite sequences for active and passive sentences, it will know that both subjects of active f-structures and objects of passive f-structures are mapped to "semantic objects". In the reverse mapping from meaning representation to f-structures, it will thus produce an f-structure chart that, besides the original f-structure, contains its paraphrase f-structure in active voice.

'Tom was discovered by Chomsky.'

↓ *XLE parsing*

Original F-Structure:

PRED	'discover < (↑ SUBJ)(↑ OBL-AG) >'
SUBJ	PRED 'Tom'; NTYPE proper-name; PERS 3
OBL-AG	PRED 'Chomsky'; ...
TOPIC	['Tom']
VTYPE	main
PASSIVE	+

↓ *transfer meaning construction*

HEAD (discover)
PAST (discover)
ROLE (sem-subj,discover,Chomsky)
ROLE (sem-obj,discover,Tom)

↓ *reverse transfer mapping*

Output F-Structure Chart:

PRED	'discover < (↑ SUBJ)(↑ OBJ) >'
SUBJ	PRED 'Chomsky'
OBJ	PRED 'Tom'
PRED	'discover < (↑ SUBJ)(↑ OBL-AG) >'
SUBJ	PRED 'Tom'
OBL-AG	PRED 'Chomsky'

↓ *XLE generation*

"Tom was discovered by Chomsky."
"Chomsky discovered Tom."

↓ *surface realisation ranking*

"Tom was discovered by Chomsky."

Figure 1: Generation via meaning representations

Also note that the generated f-structures are much more underspecified than the original one because e.g. TOPIC or CASE cannot be recovered from the semantic representation. The XLE generator adds these missing features and outputs all strings which are proper linguistic realisations of the generated f-structure chart.

The resulting set of strings is passed to a generation ranking module which selects the most appropriate paraphrase realisation in the given context (Cahill et al., 2007).

4. Instance-based Transfer Reversal

In section 2, we pointed out that we can partially reverse instances of f-structure rewrite *sequences*. We will briefly sketch this intuition here, presupposing some basic familiarity with the transfer mechanism.

The XLE transfer system represents f-structures as sets of two-place predicates. The predicate's name represents the f-structure attribute, the first argument its f-structure node and the second argument the embedded f-structure node or feature value (e.g. CASE(var(1),acc) or OBJ(var(0),var(1))). We define a rewrite mapping to be reversable if the predicate name and the atomic values of the left-hand term have a correspondence in the right-hand term. For instance, "CASE(%Var,%Value) ==> noun(%Var)" is not reversable because the second argument of the CASE predicate is deleted. On the other hand, "PRED(%Var,%Word) ==> word(%Word)" is reversable, because it only deletes the f-structure node. In this way, we allow the reverse transfer rules to reintroduce f-structure nodes. Now, an input f-structure predicate x can be defined as reversable if there is a sequence of reversable mappings between x and an output semantic clause y.

To prevent the reverse transfer rules from introducing arbitrary f-structure nodes, we have to add conditions to the left-hand side of the rules. These conditions can also be inferred automatically from seen instances of rewrite sequences. For instance, given the reverse instances "word('do') ==> PRED(var(0),'do'), past('do') ==> TENSE(var(0),past)", we know that the PRED clause has to be a condition for the second rule. Otherwise, the rule sequence "word(%Word) ==> PRED(%Word,%Var), past(%Word) ==> TENSE(%Var,past)" would introduce two independent f-structure nodes.

References

Cahill, Aoife, Martin Forst, and Christian Rohrer. 2007. Stochastic realisation ranking for a free word order language. In *Proceedings of the 11th European Workshop on NLG*, pp. 17–24, Saarbrücken, Germany.

Crouch, Dick, and Tracy Holloway King. 2006. Semantics via F-Structure Rewriting. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG06 Conference*.

Crouch, Richard, Tracy Holloway King, John T. Maxwell III, Stefan Riezler, and Annie Zaenen. 2004. Exploiting F-structure Input for Sentence Condensation . In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG04 Conference*, University of Canterbury.