PARTICLE VERBS IN COMPUTATIONAL LFGS: ISSUES FROM ENGLISH, GERMAN, AND HUNGARIAN

Martin Forst†, Tracy Holloway King†, and Tibor Laczkó‡

Microsoft Corp.†, University of Debrecen‡

Proceedings of the LFG10 Conference

Miriam Butt and Tracy Holloway King (Editors)

2010

CSLI Publications

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

Abstract

We present the ways in which particle verbs are implemented in two relatively mature computational grammars, the English and the German ParGram LFGs, and we address the issues that arise with respect to particle verbs in the development of a computational LFG for Hungarian. Considerations concerning the ParGram LFG implementation of productive Hungarian particle + verb combinations raise questions as to their treatment in the other two grammars. In addition to providing analyses for English, German, and Hungarian particle verbs, we use these phenomena to highlight how constraints on available lexical resources can affect the choice of analysis and how detailed implementations of related phenomena in typologically different languages can positively guide the analyses in all of the languages.

1 Introduction

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; Booij, 2002).[1] Particle verbs are verbs whose meaning and argument structure depend on the combination of a (base) verb and a particle. Often the meaning and argument structure of a particle verb are 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 such a particle verb in English, German, and Hungarian[2] is shown in (1).

(1) a. He **gave up** the fight. (English)

b. Er **gab** den Kampf **auf**.
he gave the fight up
'He gave up the fight.' (German)

c. Ő **fel#adta** a küzdelmet.
he up#gave the fight
'He gave up the fight.' (Hungarian)[3]

[1]We would like to thank the audience of LFG10 for their comments on the abstract and presentation of this paper. Tibor Laczkó gratefully acknowledges that his research reported here has been supported by OTKA (Hungarian Scientific Research Fund), grant number: K 72983, by the Research Group for Theoretical Linguistics of the Hungarian Academy of Sciences at the Universities of Debrecen, Pécs and Szeged, and by the TÁMOP 4.2.1./B-09/1/KONV-2010-0007 project, which is implemented through the New Hungary Development Plan, co-financed by the European Social Fund and the European Regional Development Fund. The authors are listed alphabetically.

[2]In Hungarian, what we refer to here as "particles" are often referred to as "preverbs" in the linguistic literature. Since that term is not adequate for the particles in English phrasal verbs, and for ease of exposition, we use the "particle" terminology.

[3]The '#' sign is not part of the regular orthography. We use it to indicate the boundary between particles and base verb forms when these are spelled as one word, following a convention used in a number of computational language resources.

However, particle verbs can also be compositional, as shown in (2) for some directional particles, and highly productive, which is a challenge for the coverage of computational grammars (Villavicencio, 2003).

(2) a. **Push** them **up/in/out**. (English)
b. **Push up/in/out** the boxes. (English)

In this paper, we present the ways in which particle verbs are implemented[4] in two relatively mature computational grammars, the English and the German ParGram LFGs (Butt et al., 1997, 2002), and we address the issues that arise with respect to particle verbs in the subsequent 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, especially as regards the analysis of highly productive particle verbs. An additional interesting phenomenon brought to light by Hungarian is a set of particles which exhibit inflectional properties; we outline an LFG analysis of this phenomenon which is similar to that of incorporated pronouns, e.g. with Welsh prepositions (Sadler, 1999).

Thus, in this paper we provide analyses for English, German, and especially Hungarian particle verbs. We also use these phenomena and our experience creating computational grammars which account for them to highlight how constraints on available lexical resources can, sometimes negatively, affect the choice of analysis and how detailed implementations of related phenomena in typologically different languages can positively guide the analyses and implementations in all of the languages.

2 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. This is not surprising given that particles are always written as separate words in English and short NPs can intervene between base verbs and particles, as in (3).

(3) a. They **threw** it/the trash **out**. (English)
b. They **cut** them/the onions **up**. (English)
c. They **bandied** it **about**. (English)

In German and Hungarian, however, particle + verb combinations are generally spelled as a single word when the particle immediately precedes the verb. However, variation with respect to the spelling as one or two words can be observed with semantically compositional particle + verb combinations. This particle + verb

[4]Throughout the paper we use standard LFG notation for rules and lexical entries. The implemented grammars use the XLE notation described in Crouch et al. (2010), which is a variant of the LFG notation which uses the ascii character set.

order is in a way the default order, since only clearly definable conditions (V1 and V2 in German; focus, negation, imperatives, etc. in Hungarian) cause particles to appear in positions other than the immediately preverbal one.[5] Even German and Hungarian verbs that do not exist on their own, but only in combination with particles, e.g. *aus#flippen* 'to flip/freak out' (German; **flippen*) and *be#fejez* 'to finish' (Hungarian; **fej-ez*), appear with the particle separate from the verb in these conditions, as in (4) and (5).

(4) a. . . . weil er immer so schnell **aus#flippt**.
. . . because he always so quickly out freaks
'. . . because he is always freaking out so quickly.' (German)

b. Er **flippt** immer so schnell **aus**.
he freaks always so quickly out
'He is always freaking out so quickly.' (German)

(5) a. János **be#fej-ez-te** a könyv-et.
John.NOM PV#head-VSUFF-PAST.3SG.DEF the book-ACC
'John finished the book.' (Hungarian)

b. János nem **fej-ez-te** **be** a könyv-et.
John.NOM not head-VSUFF-PAST.3SG.DEF PV the book-ACC
'John did not finish the book.' (Hungarian)

As a result of this behavior, 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 along the lines of Piñón (1992) and É. Kiss (1992, 2005) (for Hungarian), and Lüdeling (2001) (for German) across the LFG implementations for the three languages, and offer analyses that nevertheless capture the lexical properties of particle verbs in a principled manner.

3 Current Implementations in the ParGram LFGs

In this section we present the current analyses used in the broad coverage English (Riezler et al., 2002) and German (Dipper, 2003; Rohrer and Forst, 2006) ParGram grammars. Both grammars aim to capture the often-idiosyncratic meaning of the particle verbs by forming a composite PRED, while allowing for the particle and verb to appear separated in the c-structure, an analysis which is enabled by the LFG projection architecture. However, due to differences in the morphologies of the two languages, in certain constructions, namely when the particle immediately precedes the verb in German, the analyses diverge at the lexical and hence at the c-structure level.

[5] For an overview of the most important instances of this separation in German and Hungarian, see Piñón (1992).

3.1 English

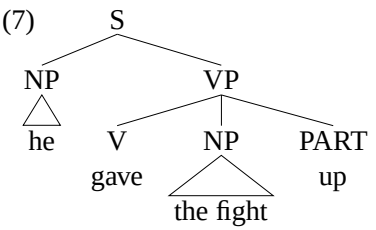
As verb particles are always spelled as separate words in English, particle verbs receive a syntactic analysis in the English ParGram LFG. The lexical entries of verb particles contribute a feature called PRT-FORM, which records the form of the respective particle. The lexical entries of base verbs introduce 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 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-occurring with the appropriate particle. In (6) are the lexical entries involved in the analysis of the English example sentence *He gave the fight up*.[6] The verb *give* can appear with a SUBJ and OBJ only if there is a PRT-FORM *up*, which will be provided by the particle. The PRED value for the verb is formed by the CONCAT template which takes the lexical verb form (%stem = *give*), a hash mark (*#*), and the PRT-FORM (*up*) and concatenates them to create a new form (%NewPred = *give#up*). The resulting c- and f-structures are shown in (7).

(6) give V (↑ PRED) = ′%NewPred<(↑ SUBJ)(↑ OBJ)>′
(↑ PRT-FORM) =$_c$ up
@(CONCAT %stem # (↑ PRT-FORM) %NewPred).

up PART (↑ PRT-FORM) = up.

(7)



$$\begin{bmatrix} \text{PRED} & \text{′}\textbf{give\#up}\text{<SUBJ,OBJ>′} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{′he′} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{′fight′} \\ \text{SPEC} & \begin{bmatrix} \text{DET} & \begin{bmatrix} \text{PRED} & \text{′the′} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

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 uses of particle verbs.

3.2 German

In German V1 and V2 clauses, particle verbs are spelled as separate words, as shown in (8a). In these contexts, the German ParGram LFG thus treats them in

[6]The actual lexical entry calls a template which expands to (6). For more on templates in XLE and LFG see Crouch et al. (2010), Dalrymple et al. (2004), and Asudeh et al. (2008).

the same way as its English counterpart. In verb-final clauses and in headed VPs, however, particle verbs are usually spelled as single words, as in (9).

(8) a. Er **lud** seine Kusine **ein**.
he loaded his cousin in
'He invited his cousin.' (German)

b. *Er **ein#lud** seine Kusine.
he in#loaded his cousin
'He invited his cousin.' (German)

(9) Er wird seine Kusine **ein#laden**.
he will his cousin in#load
'He will invite his cousin.' (German)

This spelling difference is important because it is reflected in the finite-state morphologies used to process the lexical items (Kaplan et al., 2004). In particular, the DMOR finite-state morphology (Schiller, 1995) currently used by the German ParGram LFG outputs analyses like that in (10) for particle verbs written as a single lexical item.

(10) einlud $\Longleftrightarrow$ 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 a single +V unit: the morphology does not analyze it as a particle followed by a verb.

In contrast, when the particle is separate from the verb, as in (8a), each form is passed separately to the finite-state morphology to be analyzed. The result is shown in (11), where the verb is analyzed as a +V unit and the particle as a +VPRE.

(11) a. lud $\Longleftrightarrow$ laden +V .13 .Sg .Past .Ind
b. ein $\Longleftrightarrow$ ein +VPRE

As a result, the grammar must analyze spelled-together particle verbs as morphological objects, while the spelled-apart ones comprise two morphological objects, like their English counterparts. This has the unfortunate consequence that the lexical information for the German particle verb *ein#laden* 'invite/load in' must be listed both under the base verb lemma (as in the English ParGram LFG), i.e. *laden*, and under the particle verb lemma, i.e. *ein#laden*. Fortunately, the CONCAT template makes it possible to project similar f-structures regardless of whether a given particle verb is spelled together or as separate words. So, the f-structure in (12) is the same for the sentences in (8) and (9), modulo the tense marking.

(12)
$$\begin{bmatrix} \text{PRED} & \text{'}\textbf{ein\#laden}\text{<SUBJ,OBJ>'} \\ \text{SUBJ} & \begin{bmatrix}\text{PRED} & \text{'er'}\end{bmatrix} \\ \text{OBJ} & \begin{bmatrix}\text{PRED} & \text{'Kusine'} \\ \text{SPEC} & \begin{bmatrix}\text{POSS} & \begin{bmatrix}\text{PRED} & \text{'er'}\end{bmatrix}\end{bmatrix}\end{bmatrix} \\ \text{TENSE} & \text{past/future} \end{bmatrix}$$

The German system exhibits the same limitations as the English one with respect to productively formed combinations and to the inability to differentiate compositional from non-compositional forms. Furthermore, the treatment of particle verb forms as syntactic atoms makes it necessary to use a lexically specified feature that records the fact that a given verb form contains (or does not contain) a particle, as verb forms with particles are disallowed in V1 and V2 position. Since this feature has no semantic relevance and only serves the purpose of ensuring morphosyntactic wellformedness, it is declared as a so-called CHECK feature; its name (and f-structure "path") is (CHECK _VMORPH _PARTICLE). Particle verb forms introduce this feature with the value ATTACHED, while finite verb forms in the Cbar rule are annotated with the following equation:

(13) Cbar → V[v,fin] VP
↑=↓ ↑=↓
(↓ CHECK _VMORPH _PARTICLE) ≠ ATTACHED

Consequently, verb forms that include a particle are excluded from the V1 and V2 positions. However, this exclusion is obtained via an otherwise unmotivated annotation in the Cbar rule, and it requires that this CHECK feature be introduced in the verb lexicon for each particle verb lemma.

4 A Uniformly Syntactic Analysis

In this section, we argue that a uniformly syntactic analysis as implemented in the English ParGram grammar is desirable in German and Hungarian, too, even though orthography and the semantic opacity of many particle verbs seem to suggest otherwise. We concentrate on why it is desirable to treat particles as separate c-structure nodes regardless of whether they are spelled as separate words or not.[7]

4.1 A Uniformly Syntactic Analysis for German Particle Verbs

In subsection 3.2, we described the current implementation of particle verbs in the German ParGram grammar, arguing that it is uneconomical in terms of lexical

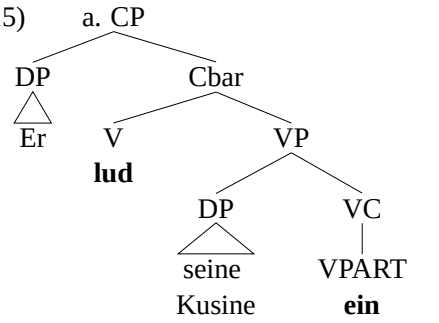
[7]We do not address the question of which category (or categories) the morphemes belong to. Many particles, or particles in general (see Toivonen (2001)), may not belong to a separate "particle" category but instead may be analyzed as adverbs, adjectives, nouns, or intransitive prepositions; here we simply assume a category called "(verb) particle".

entries and requiring a stipulative feature. Nevertheless, it is easy to understand why this analysis was chosen: With a morphology that outputs a single lemma for particle verb forms that are spelled as one word, other analyses are not an option. In order to allow for an implementation in which spelled-apart and spelled-together configurations receive more similar analyses, the output of the finite-state morphology would have to separate the particle from the verb. This is possible to do with finite-state morphologies. For example, SMOR (Schmid et al., 2004) encodes such an analysis; the resulting morphological analysis is shown in (14).[8]

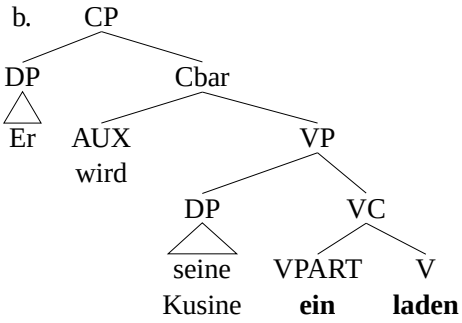
(14) einlud ⟺ ein <VPART> laden <+V> <13> <Sg> <Past> <Ind>

With such a morphological analysis, it becomes possible to treat forms like *einlud* as two c-structure nodes rather than one, which in turn makes it possible to consider the particle part of the verb complex (but not of the verb itself). Under this assumption, the c-structures for the sentences in (8) and (9) look as follows:

(15) a.



b.



While (15a) is basically the c-structure the German ParGram LFG produces for (8a), (15b) crucially differs from the c-structure that the grammar produces for (9) in that *ein* and *laden* are separate c-structure nodes. As a consequence of this, one of the lexical entries for the particle verb, namely the *ein#laden* form, becomes unnecessary; with the new analysis, only the subcategorization information listed under the base verb lemma is needed. Also, the stipulative feature (CHECK _VMORPH _PARTICLE) can be abolished because the distribution of verb particles is controlled by the c-structure rules: there is no slot for a VPART as a daughter of Cbar. Finally, the treatment of verb particles as separate c-structure nodes allows for a straightforward analysis of coordinations like (16).

(16) An der nächsten Haltestelle werden viele Leute [**ein-** und **aus**]#**steigen**.
at the next stop will many people in and out#step.
'At the next stop, there will be a lot of people getting on and off.' (German)

[8]The fact that the morphological tags are surrounded by angled brackets in this morphology instead of preceded by a full stop as in the other is unimportant. The difference in form is merely technical but is included for completeness.

A uniformly syntactic analysis of particle verbs that treats the spelled-apart and spelled-together variants similarly is thus more economical and systematic in as far as the lexicon is concerned, it gives a more parsimonious and less stipulative account of the word order facts observed with respect to particle verbs, and it makes it easy to account for coordinations of verb particles even when the second one is spelled as part of a complex verb.

4.2 A Uniformly Syntactic Analysis for Hungarian Particle Verbs

As with German particle verbs, Hungarian particle verbs are also sometimes written together and sometimes apart. As with the German case, we argue that regardless of which written form is used, the f-structure analysis should be similar. Since the finite-state morphology of the Hungarian ParGram LFG encodes an analysis that splits the particle from the base verb, as in (17), it is possible to have a more uniform c-structure analysis, thereby simplifying the Hungarian lexicon so that verbs are listed only under their stem form.

(17) belép $\Longleftrightarrow$ be +Prefix+ lép +Verb +PresInd +Indef +Sg +3P

The conditions under which Hungarian particles appear together and apart from the verb are discussed in Piñón (1992) and É. Kiss (1992, 2005). The conditions are not important for this paper, but we provide some examples in (18).

(18) a. János **fel#ad-ta** a küzdelm-et.
John.NOM up#give-PAST.3SG.DEF the fight-ACC
'John gave up the fight.'

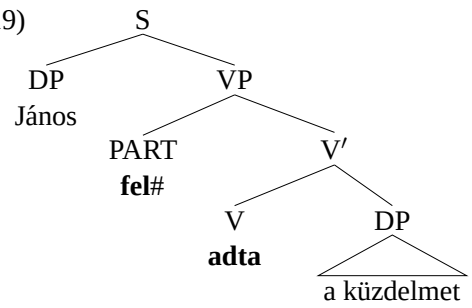
b. JÁNOS **ad-ta** **fel** a küzdelm-et.
John.NOM give-PAST.3SG.DEF up the fight-ACC
'It was JOHN who gave up the fight.'

c. János nem **ad-ta** **fel** a küzdelm-et.
John.NOM not give-PAST.3SG.DEF up the fight-ACC
'John didn't give up the fight.'

d. **Ad-d** **fel** a küzdelm-et!
give-IMPER.2SG up the fight-ACC
'Give up the fight!'

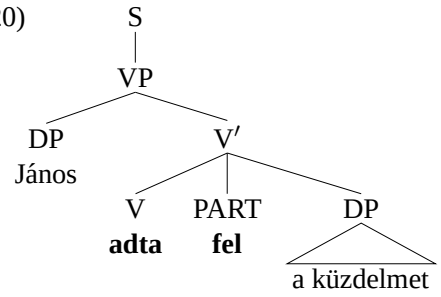
The c-structure and the f-structure for (18a) are given in (19), while those for (18b) are given in (20).

(19)



$$\begin{bmatrix} \text{PRED} & \text{'}\textbf{fel\#ad}\langle \text{SUBJ,OBJ}\rangle\text{'} \\ \text{TOP} & \left[\text{PRED} \quad \text{'János'}\right]_1 \\ \text{SUBJ} & [\ ]_1 \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'küzdelem'} \\ \text{DEF} & + \end{bmatrix} \end{bmatrix}$$

(20)



$$\begin{bmatrix} \text{PRED} & \text{'}\textbf{fel\#ad}\langle \text{SUBJ,OBJ}\rangle\text{'} \\ \text{FOC} & \left[\text{PRED} \quad \text{'János'}\right]_1 \\ \text{SUBJ} & [\ ]_1 \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'küzdelem'} \\ \text{DEF} & + \end{bmatrix} \end{bmatrix}$$

5 Compositional and Productively Formed Particle Verbs

The analyses described in Section 3 for English and German 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 texts and because the regular character of these combinations is not captured. For example, the particles *by* (English), *hinterher* 'after' (German) and *rá* 'onto' (Hungarian) can combine with a wide variety of verbs (mostly motion verbs), and generally contribute the same meaning. Following Toivonen (2001), we assume that particles in compositional particle + verb combinations can contribute aspectual information, thereby potentially affecting the telicity of the base verb, or that they can have a grammatical function subcategorized for by the base verb. In addition, we will see that certain particles can modify the argument structure of the base verb in a predictable way.

Productive, compositional particles play a number of roles relative to the verb they combine with. They can be adverbial ((21)), resultative ((22)), and aspectual ((23)). English, German, and Hungarian all have particle verbs of these types.

(21) a. She was rescued by Boris Johnson, who was **cycling by**. (English)

b. Einer Frau, die **vorbei#radelt**, johlen einige hinterher.
a.DAT woman who by#cycles hoot some after
'Some hoot after a woman who is cycling by.' (German)

c. Az ütközés után a fiú **tovább#kerékpároz-ott**.
the collision after the boy.NOM further#cycle-PAST.3SG
'After the collision the boy cycled on.' (Hungarian)

(22) a. He **sanded off** the paint from the timber ceiling. (English)

b. Er läßt sich die Rüsselnase **ab#operieren**.
He lets/makes himself the trunk nose off#operate
'He has his trunk-like nose surgically removed.' (German)

c. János **le#smirgliz-te** a festék-et a gerendá-ról.
John.NOM down#sand-PAST.3SG.DEF the paint-ACC the timber-FROM
'John sanded off the paint from the timber.' (Hungarian)

(23) a. She **painted on** for half an hour or so. (English)

b. Jetzt muß in der ganzen DDR **weiter#getrommelt** werden.
Now must in the entire GDR on#drummed be
'Now the drumming has to continue throughout the entire GDR.' (German)

c. János **meg#ír-ta** a level-et egy óra alatt.
John.NOM perf#write-PAST.3SG.DEF the letter-ACC one hour under
'John wrote the letter in one hour.' (Hungarian)

In the remainder of this section, we show how the different types of compositional and productively formed particle verbs can be accounted for in LFG. We demonstrate the technical details of the proposed analyses using mostly German examples, but these accounts naturally and straightforwardly carry over to Hungarian and English.

5.1 Adverbial Particles

Sentence (24) exemplifies a particle + verb combination where the particle adds adverbial information. Adverbials expressing a path tend to directly precede the verb in German; this may be the reason why a number of "path" adverbs that routinely occur in this position are considered particles by prescriptive grammars and hence spelled together with the verb, even though both spelled-together and spelled-apart variants can be found in corpora.

(24) Einer Frau, die **vorbei#radelt**, johlen einige hinterher.
a.DAT woman who by#cycles hoot some after
'Some hoot after a woman who is cycling by.' (German)

The analysis of this category treats the particle as a separate word with its own PRED, which either fills one of the argument positions subcategorized for by the base verb or acts as an ADJUNCT modifier of the verb. Here we do not concern

ourselves with the vexing question of whether the particle becomes an ADJUNCT or an OBL-DIR (see Zaenen and Crouch (2009) on differentiating adjuncts and obliques, especially in implemented grammars), but instead focus on how these analyses can be realized in LFG.

The lexical entries for the adverbial use of the particle *vorbei* 'by' and the lexical entry for the verb *radeln* 'cycle' are shown in (25).

(25) a. vorbei VPART (↑OBL-DIR PRED)='vorbei'
b. radeln V (↑PRED)='radeln<SUBJ, OBL-DIR>'

The VC rule, in which VPART is introduced, then combines these just like it combines verbs with particles that only introduce a PRT-FORM feature for an idiomatic particle + verb combination.[9]

(26) VC ⟶ (VPART) (V)
↑=↓ ↑= ↓

The f-structure this rule and these lexical entries produce is shown in (27).

(27)
$$\begin{bmatrix} \text{PRED} & \text{'Frau'} \\ \text{ADJUNCT} & \left\{ \begin{bmatrix} \text{PRED} & \text{'}\mathbf{radeln}\text{<SUBJ, OBL-DIR>'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'pro'} \end{bmatrix} \\ \text{OBL-DIR} & \begin{bmatrix} \text{PRED} & \text{'}\mathbf{vorbei}\text{'} \end{bmatrix} \end{bmatrix} \right\} \\ \text{SPEC} & \begin{bmatrix} \text{DET} & \begin{bmatrix} \text{PRED} & \text{'eine'} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Similarly, the adverbial particle *mit* 'with', illustrated in (28a), adds an ADJUNCT to the base verb. We therefore assume the lexical entries in (29) for the particle *mit* and the verb *fahren* 'go/drive'. These give rise to the analysis in (30).

(28) a. Wer will nach Norwegen **mit#fahren**?
who wants to Norway with#go
'Who wants to go to Norway with us?' (German)

b. Wer will **mit** nach Norwegen **fahren**?
who wants with to Norway go
'Who wants to go to Norway with us?' (German)

(29) a. mit VPART (↑ADJUNCT ∈ PRED)='mit'
b. fahren V (↑PRED)='fahren<SUBJ, OBL-DIR>'

[9] V and PART themselves are nodes with internal structure as dictated by the output of the morphology (see section 4). See Kaplan et al. (2004) on sublexical rules and the interaction of morphology and syntax in LFG.

(30) $\begin{bmatrix} \text{PRED} & \text{'wollen<XCOMP>SUBJ'} \\ \text{SUBJ} & \left[\text{PRED} \quad \text{'pro'}\right]_1 \\ \text{XCOMP} & \begin{bmatrix} \text{PRED} & \text{'}\mathbf{\text{fahren}}\text{<SUBJ, OBL-DIR>'} \\ \text{SUBJ} & [\ ]_1 \\ \text{OBL-DIR} & \begin{bmatrix} \text{PRED} & \text{'nach<OBJ>'} \\ \text{OBJ} & \left[\text{PRED} \quad \text{'Norwegen'}\right] \end{bmatrix} \\ \text{ADJUNCT} & \left\{\left[\text{PRED} \quad \text{'}\mathbf{\text{mit}}\text{'}\right]\right\} \end{bmatrix} \end{bmatrix}$

This analysis of the particle *mit* also has the advantage of projecting f-structures with identical verbal PREDs and argument structures for sentences (28a) and (28b), which are semantically equivalent. Finally, note that it is a directional PP that can intervene between the adverb/particle *mit* and the clause-final verb; the adverbial particle *mit* and the directional PP seem to compete for the position directly to the left of the verb.

5.2 Resultative Particles

Verb particles often take part in resultative constructions, as in (31).

(31) Er läßt sich die Rüsselnase **ab#operieren**.
He lets/makes himself the trunk nose off#operate
'He has his trunk-like nose surgically removed.' (German)

Resultative particle verbs are probably the category of (semi-)compositional particle + verb combinations that has received the most attention and that has motivated a number of Chomskyan syntacticians to claim that particles head small clauses.

In LFG terms, this corresponds to an XCOMP-PRED analysis, and so we posit the lexical entry in (32a) for the particle *ab* 'off' as used in (31). For the verb *operieren* 'operate', we assume the lexical entry in (32b).[10]

(32) a. ab VPART (↑XCOMP-PRED PRED)='ab<SUBJ>'
b. operieren V (↑PRED)='operieren<SUBJ,OBJ,OBJ$_\theta$,XCOMP-PRED>'
(↑XCOMP-PRED SUBJ)=(↑OBJ)

These lexical entries in combination with the regular VC rule shown in (26) associate (31) with the following f-structure.

[10]The resultative subcategorization frame for *operieren* 'operate' is needed for particle and full phrase resultative predicates. The OBJ$_\theta$ in the subcategorization frame is a dative external possessor construction, which is also independent of the occurrence of a particle.

(33)
$$\begin{bmatrix}
\text{PRED} & \text{'lassen<SUBJ, XCOMP>OBJ'} \\
\text{SUBJ} & \left[\text{PRED} \quad \text{'pro'}\right] \\
\text{OBJ} & [\text{ die Rüsselnase }]_1 \\
\text{XCOMP} & \begin{bmatrix}
\text{PRED} & \text{'}\mathbf{operieren}\text{<NULL, SUBJ, OBJ}_\theta\text{, XCOMP-PRED>'} \\
\text{SUBJ} & [\ ]_1 \\
\text{OBJ}_\theta & \left[\text{PRED} \quad \text{'pro'}\right] \\
\text{XCOMP-PRED} & \begin{bmatrix}\text{PRED} & \text{'}\mathbf{ab}\text{<SUBJ>'} \\ \text{SUBJ} & [\ ]_1\end{bmatrix} \\
\text{PASSIVE} & +
\end{bmatrix}
\end{bmatrix}$$

This structure relies on a resultative subcategorization frame for *operieren* 'operate' that can be provided by a lexical rule in the spirit of Simpson (2006), for example. However, since the question of the resultative use of verbs needs to be addressed independently of their co-occurrence with particles, we believe that our implementational framework could, in theory, accommodate a purely syntactic treatment as well, for instance, along the lines of Alsina (1996). He proposes that the predicate and the resultative expression bring about their special joint argument structure in the syntax. This is based on his assumption that the argument structure of a predicate may be different from that of the clause it occurs in.

5.3 Aspectual Particles

For aspectual particles, examples of which were shown in (23), Toivonen (2001) suggests introducing aspectual features or a separate PRED for the particle. The lexical entry of an aspectual particle can thus look like (34). (35) shows the f-structure for *She painted on (for days and days)*. Here the aspectual information is contributed by the particle *on* as a simple feature for continuous aspect; see Toivonen (2001) and references therein for analyses of aspect within LFG.

(34) on PART (↑ TNS-ASP ASPECT)=continuous
(↑ PRT-FORM)=on

(35)
$$\begin{bmatrix}
\text{PRED} & \text{'}\mathbf{paint}\text{<SUBJ>'} \\
\text{SUBJ} & \left[\text{PRED} \quad \text{'she'}\right] \\
\text{TNS-ASP} & \begin{bmatrix}\text{ASPECT} & \mathbf{continuous} \\ \text{PRT-FORM} & \mathbf{on}\end{bmatrix}
\end{bmatrix}$$

5.4 Argument-Changing Particles

As seen above, particles can combine with verbs in a compositional way without altering the argument structure. However, at least in German and Hungarian, there are also productive particles that can add arguments, as pointed out, e.g. by

Stiebels (1996). An example is the German particle *hinterher* 'after', which in (36) subcategorizes for the dative argument *dem Glück* 'the happiness'.[11]

(36) **Lauf** dem Glück nicht länger **hinterher**!
run.IMP.2SG the-DAT happiness not longer after
'Don't run after happiness any longer!' (German)

This type of behavior where the addition of a morpheme, here the particle, results in a predictable modification of the argument structure is reminiscent of causatives and complex predicates. Butt et al. (2003) and Butt and King (2006) use restriction to analyze these constructions and provide a way to implement them in XLE for the ParGram LFG grammars. We propose that certain compositional particle verbs should be analyzed similarly.

The lexical entries for a complex-predicate analysis of (36) are shown in (37).

(37) a. hinterher VPART (↑ PRED)='hinterher<%ARG1,(OBJ$_\theta$)>'
b. laufen V (↑PRED)='laufen<SUBJ>'

The rules that require additional annotations for proper predicate composition are the VC rule, whose "regular" version was presented in (26), as well as the Cbar rule, which introduces the finite verb of a clause when it is in V2 position.

(38) a. VC ⟶ (VPART) (V)
↑=↓ ↑/PRED/OBJ$_\theta$ = ↓/PRED
(↑PRED ARG1) = (↓ PRED)

b. Cbar ⟶ V (VP)
↑/PRED/OBJ$_\theta$ = ↓/PRED ↑=↓
(↑PRED ARG1) = (↓ PRED)

These rules and lexical entries produce the f-structure in (39) for (36). Note that the particle and the verb together make up the PRED of the top f-structure, both contributing to the argument structure of the combination.

(39)
$$\begin{bmatrix} \text{PRED} & \text{'}\mathbf{hinterher}\text{<'}\mathbf{laufen}\text{<SUBJ>'},\text{OBJ}_\theta\text{>'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'pro'} \end{bmatrix} \\ \text{OBJ}_\theta & \begin{bmatrix} \text{PRED} & \text{'Glück'} \\ \text{SPEC} & \begin{bmatrix} \text{DET} & \begin{bmatrix} \text{PRED} & \text{'die'} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{ADJUNCT} & \left\{ \begin{bmatrix} \text{PRED} & \text{'nicht'} \end{bmatrix}, \begin{bmatrix} \text{PRED} & \text{'lang'} \end{bmatrix} \right\} \end{bmatrix}$$

[11] The fact that the sentential negation intervenes between *dem Glück* 'the happiness' and *hinterher* 'after' makes it impossible to analyze the particle as a postposition. Also compare (24), where *hinterher* is separated from its argument by even more material.

5.5 Corpus Frequency of Compositional Particle Verbs

To quantify the importance of a productive rather than a list-based treatment of compositional particle + verb combinations, we did a small-scale corpus study of German particle verbs. Knowing that the particles *entgegen* 'towards' and *hinterher* 'after' combine with base verbs almost exclusively in a compositional fashion, we extracted all verb forms whose lemma starts with one of these particles from the so-called Huge German Corpus of the University of Stuttgart, and sorted them by frequency. Then we determined which of these particle verbs are in the German ParGram verb lexicon.[12]

The result in terms of tokens is the following: Of the 11,652 verb forms starting with *entgegen*, 11,067 correspond to a lexical entry in the grammar's verb lexicon; 585 or 5.0% of the verb tokens starting with the particle *entgegen* are not covered by the verb lexicon. For the particle *hinterher*, the situation is even clearer. Of the 1,164 verb forms starting with *hinterher*, only 604 correspond to a lexical entry in the verb lexicon, which leaves 542 or 47.3% of the tokens of this type uncovered.

Looking at types reveals that, for both kinds of particle + verb combinations, more than half of the types correspond to a single token, which suggests that these combinations are truly productive. In the case of verb forms starting with *entgegen*, the proportion of hapax legomena (words which occur only once in a given corpus) is 117/223=52.5%; in the case of verbs starting with *hinterher*, it is 78/148=52.7%. Only two of the *entgegen* and one of the *hinterher* hapax legomena are listed in the verb lexicon; overall, 85.7% of the *entgegen* + verb combination types and 91.2% of the *hinterher* + verb combination types are unknown to the verb lexicon.

5.6 Compositional vs. Non-compositional Particle Verbs

To conclude this section, we return to the issue of how best to distinguish compositional from non-compositional particle verbs in an LFG analysis. We propose that compositional particle verbs be analyzed by means of a lexical entry for the particle where, rather than a PRT-FORM feature, it contributes a PRED that, in certain cases, may subcategorize for an argument it introduces. A predicate composition rule involving restriction then fills argument slots of argument-changing particles.

Although we argue that compositional particle verbs are best accounted for using the analyses proposed above, the non-compositional particle verbs should be captured by an analysis similar to that outlined in Section 3. Under this analysis, the verb has a lexical entry which lists the particle it co-occurs with as well as its argument structure. In addition, the predicate is formed by concatenation, i.e. prt#verb, so that a unique PRED is formed, reflecting the non-compositional nature of the particle verb combination. In this way, compositional and non-compositional particle verbs are easily distinguished both in the grammar implementation and in the resulting f-structures.

[12] We only verified whether there was any lexical entry at all for a given particle + verb combination; we did not verify whether the lexical entry was adequate.

(40) **Non-compositional:**

$$\begin{bmatrix} \text{PRED} & \text{'prt\#verb<GFS>'} \\ \text{GFS} & [\ ] \\ \text{PRT-FORM} & \text{prt} \end{bmatrix}$$

(41) **Compositional: adv. oblique**

$$\begin{bmatrix} \text{PRED} & \text{'verb<GFS,OBL>'} \\ \text{GFS} & [\ ] \\ \text{OBL} & \begin{bmatrix} \text{PRED} & \text{'prt'} \end{bmatrix} \end{bmatrix}$$

Compositional: adv. adjunct

$$\begin{bmatrix} \text{PRED} & \text{'verb<GFS>'} \\ \text{GFS} & [\ ] \\ \text{ADJUNCT} & \left\{ \begin{bmatrix} \text{PRED} & \text{'prt'} \end{bmatrix} \right\} \end{bmatrix}$$

Compositional: resultative

$$\begin{bmatrix} \text{PRED} & \text{'verb<GFS,XCOMP-PRED>'} \\ \text{GFS} & [\ ] \\ \text{XCOMP-PRED} & \begin{bmatrix} \text{PRED} & \text{'prt<SUBJ>'} \end{bmatrix} \end{bmatrix}$$

Compositional: aspectual

$$\begin{bmatrix} \text{PRED} & \text{'verb<GFS>'} \\ \text{GFS} & [\ ] \\ \text{TNS-ASP} & \begin{bmatrix} \text{ASPECT} & \text{prt} \end{bmatrix} \end{bmatrix}$$

Compositional: predicate composition

$$\begin{bmatrix} \text{PRED} & \text{'prt<'verb<GFS>',PRTGF>'} \\ \text{GFS} & [\ ] \\ \text{PRTGF} & [\ ] \end{bmatrix}$$

As a final point, these productive particles may still be semantically restricted so that they cannot occur with verbs with incompatible meanings (e.g. argument structure and aspectual incompatibilities).

6 Hungarian Inflected Particles

In addition to the uninflected particles found in Germanic, Hungarian has inflected particles. Many Hungarian particles are etymologically related to postpositions (*alá* 'to.under') or oblique case suffixes (*-ra* 'onto'). Under normal circumstances, when such elements are used in a pronominal context, the standard Hungarian morphological strategy is to take these elements as 'stems' and to add the 'pronominal content' inflectionally, as in (42).

(42) a. alá 'to.under' b. alá-m 'to.under-1SG'
c. -ra 'onto' d. rá-m 'onto-1SG'

When the argument of the particle verb would be pronominal, Hungarian does not use the particle in its 'neutral' form and express the pronominal oblique argument by a separate inflected element, as shown by the ungrammatical (43).

(43) *János **rá#lép-ett** rá-m.
John.NOM onto#step-PAST.3SG onto-1SG
'John stepped onto me.'

Instead, the 'pronominally inflected' particle alone is used to encode this meaning.

(44) János **rá-m** **lép-ett**.
John.NOM onto-1SG step-PAST.3SG
'John stepped onto me.'

Furthermore, when the oblique argument is 3rd person and non-pronominal, the corresponding inflected particle cannot be used; compare the forms in (45).

(45) a. Mari **rá** **lép-ett** a toll-ak-ra.
Mari onto step.PAST.3SG the pen-PL-SUBL
'Mari stepped onto the pens.' (Hungarian)

b. Mari **rá-juk** **lép-ett**.[13]
Mari onto-3PL step.PAST.3SG
'Mari stepped onto them.' (Hungarian)

We propose that these inflected particles are straightforwardly analyzed in LFG by the classic pronoun incorporation analysis. All these facts can be captured by assuming that with inflected particles the inflectional morphology obligatorily introduces the PRED='pro' feature.[14] An example lexical entry is shown in (46) and the resulting f-structure for (45b) is shown in (47).

(46) rájuk (↑ OBL PRED)='rá'
(↑ OBL OBJ PRED)='pro'
(↑ OBL OBJ PERS)=3
(↑ OBL OBJ NUM)=pl

(47)
$$\begin{bmatrix} \text{PRED} & \text{'}\mathbf{lép}\text{<SUBJ,OBL>'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'Mari'} \end{bmatrix} \\ \text{OBL} & \begin{bmatrix} \text{PRED} & \text{'}\mathbf{rá}\text{<OBJ>'} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'pro'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{sg} \end{bmatrix} \end{bmatrix} \\ \text{TENSE} & \text{past} \end{bmatrix}$$

The fact that the PRED 'pro' is obligatory in inflected particles, instead of optional as is the case in many analyses of pro-drop, ensures that the inflected particle cannot be doubled by an overt argument. This is shown by the ungrammaticality of (48) where the inflected particle *rá-juk* is doubled by an overt oblique NP.

(48) *Mari **rá-juk** **lépett** a toll-ak-ra.
Mari onto-3PL step.PAST.3SG the pen-PL-SUBL
'Mari stepped onto them the pens.' (Hungarian)

[13]Note that conventional Hungarian orthography has the uninflected particle + verb combination as one word and the inflected particle + verb combination typically as two words.

[14]For an inventory of Hungarian inflecting particles and a lexicalist analysis, see Ackerman (1990).

7 Conclusions

We established that the implemented LFG analysis of particle verbs for English and German is appropriate and feasible for non-compositional particle constructions. However, the existing analyses had two problems. First, the analysis did not distinguish between idiomatic and compositional particle-verb constructions: both types were listed in the lexicon and created PREDs composed of the particle and the verb. Second, the German analysis required two different c-structure analyses and, as a result, repetition in the lexicon.

We proposed that idiomatic particle-verb constructions be listed in the lexicon and have PRED values which are composed of the particle and the verb. These idiomatic particle verbs may have argument structures which differ significantly from the verb's non-particle counterpart. This analysis is similar to that usually assumed in the LFG literature and implemented in XLE grammars.

We then argued that compositional particle verbs be composed in the syntax. In many cases, the particle fills an argument slot of the base verb, e.g. a resultative XCOMP-PRED or an adverbial OBL, it modifies the base verb by functioning as an adjunct, or it adds aspectual information. In other cases, the particle introduces an additional argument, and restriction is used to create a new PRED for the verb which differs from the original PRED only in the addition of the new argument. In both types, the additional f-structure information is provided by the lexical entry of the particle. Note that although the compositional particle-verb construction can be very productive, there are semantic constraints on the allowable combinations; we leave the investigation of these constraints for future work.

The issue with the repetition of lexical items in the German lexicon was solved by incorporating a different morphological analysis whereby, even when particle-verb combinations are written as a single word, they comprise two different tokens in the c-structure. That is, their sublexical analysis involves the particle, the verb, and the morphemes encoding inflectional information. The c-structure rule for verbs is then minimally modified to allow for this construction.

Hungarian particle verbs can be straightforwardly accounted for with this division between idiosyncratic and compositional particle verbs, as well as a morphological analysis similar to the one proposed for German. In addition, Hungarian has inflected particles. We argued that LFG, and its implementation via XLE, allows for a straightforward pro-drop style analysis of Hungarian inflected particles.

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 *by-standers*, German *Einladung* 'invitation'). An additional derivational morphology issue involving particle verbs is discussed in Booij (2002): he provides a construction grammar analysis of particles used to create verbs out of nouns and adjectives in Dutch and German. We leave these areas for future work.

References

Ackerman, Farrell. 1983. *Miscreant Morphemes: Phrasal Predicates in Ugric*. Ph. D. thesis, UC Berkeley.

Ackerman, Farrell. 1990. The Morphological Blocking Principle and Oblique Pronominal Incorporation in Hungarian. In Katarzyna Dziwirek, Patrick Farrell and Errapel Meijas-Bikandi (eds.), *Grammarical Relations: A Cross-Theoretical Perspective*, pages 1–19, CSLI Publications.

Alsina, Àlex. 1996. Resultatives: A Joint Operation of Semantic and Syntactic Structures. In *Proceedings of the First LFG Conference*, Grenoble, France: Rank Xerox Research Centre.

Asudeh, Ash, Dalrymple, Mary and Toivonen, Ida. 2008. Constructions with Lexical Integrity: Templates as the Lexicon-Syntax Interface. In *Proceedings of LFG08*, pages 68–88, CSLI On-Line Publications.

Booij, Geert. 2002. Separable Complex Verbs in Dutch: A Case of Periphrastic Word Formation. In *Verb-Particle Explorations*, pages 21–42, Berlin, Germany: Mouton de Gruyter.

Butt, Miriam, Dyvik, Helge, King, Tracy Holloway, Masuichi, Hiroshi and Rohrer, Christian. 2002. The Parallel Grammar Project. In *Proceedings of the COLING 2002 Workshop on Grammar Engineering and Evaluation*.

Butt, Miriam and King, Tracy Holloway. 2006. Restriction for Morphological Valency Alternations: The Urdu Causative. In *Intelligent Linguistic Architectures: Variations on Themes by Ronald M. Kaplan*, pages 235–258, CSLI Publications.

Butt, Miriam, King, Tracy Holloway and Maxwell, III, John T. 2003. Complex Predication via Restriction. In *Proceedings of the LFG 2003 Conference*, pages 92–104, CSLI On-line Publications.

Butt, Miriam, King, Tracy Holloway, Niño, María-Eugenia and Segond, Frédérique. 1997. *A Grammar-Writer's Cookbook*. CSLI Publications.

Crouch, Dick, Dalrymple, Mary, Kaplan, Ron, King, Tracy, Maxwell, John and Newman, Paula. 2010. XLE Documentation. Technical Report, Palo Alto Research Center, Palo Alto, CA, www2.parc.com/istl/groups/nltt/xle/doc/xle_toc.html.

Dalrymple, Mary, Kaplan, Ronald M. and King, Tracy Holloway. 2004. Linguistic Generalizations over Descriptions. In *Proceedings of LFG05*, pages 199–208, CSLI On-Line Publications.

Dipper, Stefanie. 2003. *Implementing and Documenting Large-scale Grammars – German LFG*. Ph. D. thesis, IMS, University of Stuttgart, Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung (AIMS), Volume 9, Number 1.

É. Kiss, Katalin. 1992. Az egyszerű mondat szerkezete [The structure of the simple sentence]. In Ferenc Kiefer (ed.), *Strukturális Magyar Nyelvtan 1. Mondattan [Structural Hungarian Grammar 1. Syntax]*, pages 79–177, Akadémiai Kiadó.

É. Kiss, Katalin. 2005. First Steps towards a Theory of the Verbal Particle. In Christopher Piñón and Péter Siptár (eds.), *Papers from the Düsseldorf Conference. Approaches to Hungarian 9*, pages 57–88, Akadémiai Kiadó.

Kaplan, Ronald M., Maxwell, III, John T. Maxwell, King, Tracy Holloway and Crouch, Dick. 2004. Integrating Finite-state Technology with Deep LFG Grammars. In *Proceedings of the Workshop on Combining Shallow and Deep Processing for NLP (ESSLLI)*, pages 11–20.

Lüdeling, Anke. 2001. *On Particle Verbs and Similar Constructions in German*. CSLI Publications.

Piñón, Chris. 1992. The Preverb Problem in German and Hungarian. In *Proceedings of BLS*, pages 395–408.

Riezler, Stefan, King, Tracy Holloway, Kaplan, Ronald, Crouch, Dick, Maxwell III, John T. and Johnson, Mark. 2002. Parsing the Wall Street Journal using a Lexical-Functional Grammar and Discriminative Estimation Techniques. In *Proceedings of the ACL*.

Rohrer, Christian and Forst, Martin. 2006. Improving Coverage and Parsing Quality of a Large-scale LFG for German. In *Proceedings of the Language Resources and Evaluation Conference (LREC-2006)*, Genoa, Italy.

Sadler, Louisa. 1999. Non-Distributive Features in Welsh Coordination. In *Proceedings of LFG99*, CSLI On-Line Publications.

Schiller, Anne. 1995. DMOR Benutzerhandbuch. Technical Report, IMS, University of Stuttgart.

Schmid, Helmut, Fitschen, Arne and Heid, Ulrich. 2004. SMOR: A German Computational Morphology Covering Derivation, Composition and Inflection. In *Proceedings of the IVth International Conference on Language Resources and Evaluation (LREC 2004)*, Lisbon, Portugal.

Simpson, Jane. 2006. Resultatives. In *Lexical Semantics in LFG*, Stanford, CA: CSLI Publications.

Stiebels, Barbara. 1996. *Lexikalische Argumente und Adjunkte: Zum semantischen Beitrag von verbalen Präfixen und Partikeln*. Berlin, Germany: Akademie Verlag.

Toivonen, Ida. 2001. *The Phrase Structure of Non-Projecting Words*. Ph. D. thesis, Stanford University.

Villavicencio, Aline. 2003. Verb-Particle Constructions and Lexical Resources. In *Proceedings of the ACL 2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*.

Zaenen, Annie and Crouch, Dick. 2009. OBLs Hobble Computations. In *Proceedings of LFG09*, CSLI On-line Publications.