

Tanulmány

Rákosi György, Laczkó Tibor & Csernyi Gábor

On English phrasal verbs and their Hungarian counterparts: from the perspective of a computational linguistic project¹

Abstract

The paper gives an overview of the grammar of particle verb constructions in English and Hungarian, and outlines the framework of an LFG-theoretic analysis and its computational implementation that tackles both the lexical and the syntactic properties of these constructions.

Keywords: particle verbs, English, Hungarian, Lexical-Functional Grammar, computational implementation

1 Introduction

Particle verb constructions (a.k.a. *phrasal verbs* in English or *preverb+verb* constructions in Hungarian) show many intriguing properties both in English and in Hungarian which provide a challenge for theoretical analyses as well as for computational implementation. In this article, we overview the core grammar of these constructions, paying attention to both the similarities and the differences between the two languages. After a brief survey of the linguistic literature, we discuss the outlines of an analysis in the framework of *Lexical-Functional Grammar*, and show how this analysis can be implemented in a computational grammar of English and Hungarian.

The main reason why particle verb constructions are so challenging is that despite their obvious syntactic complexity, they often show traits of being *lexically complex* units. Consider the following English sentence in (1a) and its Hungarian equivalent in (1b):

(1) a. *John gave up the fight.*

¹ The authors acknowledge that the research reported here has been supported, in part, 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 University 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.

- b. *János fel-ad-ta* *a* *harc-ot.*
 John up-give-past.3sg the fight-acc
 ‘John gave up the fight.’

The particle verb is non-compositional in both languages, there is no reasonable semantic computation that derives the meaning ‘abandon’ for *give up* from the meaning of the base verb and the particle. Furthermore, particle verb formation changes the argument structure of the verbal stem: whereas *give* is a three-place predicate, *give up* is diadic, the dative beneficiary argument having been removed from the argument structure. These facts strongly point towards a lexical analysis. In addition, Hungarian particle verbs very productively serve as input to lexical processes, cf.:

- (2) *a fel-ad-hatatlan* *harc*
 the up-give-able.not fight
 ‘the fight that cannot be given up’

It is also the case that the verb and the particle that immediately precedes it are frequently spelled as one word (cf. 1b) in standard Hungarian orthography. This is, strictly speaking, not a linguistic issue, but it still reflects the underlying intuition that the construction is a lexical unit. In addition, this orthographical convention *is* an issue in the computational implementation, where the particle must always be linked up with the verb.

Nevertheless, there are also reasons to treat particle verbs as syntactically complex units. Notice, first of all, that a sizeable subset of them is fully compositional both in English and Hungarian:

- (3) a. *John pushed them up/down.*
 b. *János fel-/le- tol-ta őket.*
 John up-/down- push-past.3sg them
 ‘John pushed them up/down.’

More importantly, the particle can be separated from the verb in both languages irrespective of whether the particle verb is compositional or not. Compare (1) with (4):

- (4) a. *John did not give the fight up.*
 b. *János nem ad-ta (fel) a harc-ot (fel).*
 John not give-past.3sg up the fight-acc up
 ‘John did not give the fight up.’

Upon, for example, sentence negation, the particle can be separated from the verb in Hungarian and can occupy a range of positions (4b). In English, non-heavy objects can generally separate the particle from the verb (4a).

The examples in (4) thus illustrate the problem in both languages. On the one hand, *give up* needs to be stored in the lexicon as a separate entry mostly because of its non-predictable semantic and argument structure properties. On the other hand, we must store this unit in such a way that the verb and the particle still stay separable in syntax. Though the grammar of particle verb constructions may vary beyond this point, the basic challenge is constant: we

must capture both the lexical and the syntactic properties of these constructions *at the same time*. In this article, we make an attempt at outlining an analysis that is capable of achieving this dual goal.

2 Previous approaches to English and Hungarian particle verbs

English and Hungarian particle verbs (just like particle verbs in other languages like German) pose serious questions for morphological and syntactic theories respecting widely accepted principles and assumptions. Furthermore, an adequate analysis must also capture the fact that there are some significant differences in behaviour between English and Hungarian particle verbs. In this section we offer a brief overview of some basic types of possible approaches to these complex phenomena.

2.1 On the treatment of English particle verbs

Given that English particle verb constructions behave as syntactically complex units in some of their crucial features, there are, in the literature, many syntactic accounts. They come in two major varieties.

(A) It is a relatively popular view to assume that, in the case of transitive predicates, the direct object and the particle make up a constituent: there is a kind of a subject-predicate relationship between them, and this constituent is the complement of the verb. In Chomskyan terms, for instance, such an analysis typically holds that the verb subcategorizes for a *small clause* complement (cf., for example, den Dikken 1995 and Kayne 1998):

(5) [VP *call* [SC [NP *Mary*] [Part *up*]]]

(B) The alternative syntactic analysis postulates that it is the verb and the particle that form a unit, and this constituent is combined with the direct object. For a typical Chomskyan account, see Radford (1997).

There are, however, several proponents of a morphological/lexical approach. They assume that the verb and the particle combine *presyntactically*, this complex predicate is inserted below a V^0 node in the syntax, and they offer various transformational techniques for capturing the syntactic separability of the verb and the particle, cf. Johnson (1991) and Koizumi (1993).

2.2 On the treatment of Hungarian particle verbs

Given the many lexically sensitive aspects of the behaviour of Hungarian particle verbs, it is not surprising that there have been several morphologically oriented accounts proposed in the Hungarian generative literature, see, for instance, É. Kiss (1987) and Ackerman-Webelhuth (1998). These approaches, although differing in significant details, all assume that the verb and the particle combine in the lexical component of the grammar, and some special rules ensure their separability in the syntax. The bottom line here is that, contrary to rather widely accepted assumptions (especially in the generative literature), the lexicon has to be allowed to handle the combination of syntactically separable elements: the verb and the particle. For

example, Ackerman-Webelhuth (1998) argue that, in addition to ordinary synthetic lexical items (words), taken to be syntactic atoms, as a marked option, analytic word combinations must also be admitted in this component of grammar. Alternatively, É. Kiss (1987) assumes that particle verbs (*particle+verb* clusters) are of category V^0 ; however, they are exempt from morphological bracket erasure, therefore their pieces can be syntactically visible and active. It is a very strong argument in favour of this lexicalist line that, as we pointed out in Section 1, Hungarian particle verbs absolutely productively serve as input to genuine derivational, that is lexical, processes.

The extremely free (although principally predictable) syntactic dissociation of the verb and the particle has naturally triggered various kinds of syntactically biased analyses in nature, again, differing in significant details. See, for instance, Piñón (1992) and É. Kiss (2005). The basic idea is that the verb and the particle are always inserted from the lexicon as separate syntactic atoms. This even holds for cases in which the particle immediately precedes the verb, and standard Hungarian orthography writes them as one word. These approaches postulate a special, distinct syntactic position for the immediately preverbal particle as well. The challenge here is to capture the lexical properties of particle verbs (including their frequent non-compositionality as well as their being input to productive derivational processes).

3 Outlines of a uniform Hungarian and English analysis from a computational linguistic perspective

3.1 Hungarian particle verbs

Our Lexical-Functional Grammar Research Group works on developing an LFG grammar of the Hungarian language and on implementing it on an LFG-based computational linguistic platform in the context of an international collaboration. (On LFG, see Bresnan 2001; on the international program called *Parallel Grammar, ParGram*, for short, see <http://pargram.b.uib.no>; on our Hungarian grammar project, *HunGram*, for short, see <http://hungram.unideb.hu>; and on the implementational platform, see Butt et al. 1997.)

It is generally desirable both from a theoretical and from an implementational point of view to aim at as uniform a treatment of the phenomena under investigation as possible. As far as our analysis of Hungarian particle verbs that we are currently developing is concerned, at this stage our objective is a uniform syntactic approach in terms of both theory and implementation, on the basis of the following considerations.

(A) There are several strong empirical arguments for an overall syntactic analysis, cf. Section 1. For detailed argumentation, see the references given in section 2.2.

(B) It is of special importance that the postulation of a distinct preverbal syntactic position for the particle even when it immediately precedes the verb, despite the very strong one-word-spelling tendency in Hungarian orthography, provides a straightforward tool for accounting for the similar behaviour and complementarity of particles and other elements (typically: reduced arguments of the verbal predicate), collectively called verbal modifiers (for details, see Komlósy 1992, for instance). Compare the following examples:

- (6) a. *János el-olvas-ta a könyv-et.* (el = particle)
John away-read-past.3sg the book-acc
'John read the book.'
- b. *János nem olvas-ta el a könyv-et.*
John not read-past.3sg away the book-acc
'John did not read the book.'
- (7) a. *János könyv-et olvas-ott.* (könyvet: reduced object argument)
John book-acc read-3sg
'John was reading a book.'
- b. *János nem olvas-ott könyv-et.*
John not read-3sg book-acc
'John was not reading a book.'

(C) In the flexible architecture of LFG, the lexical component can easily handle the introduction of distinct lexical entries for the base verb and the related particle. Moreover, it is straightforwardly possible technically to provide them with appropriate information, in LFG terms: with appropriate annotations, that will capture the relevant functional and semantic properties of the given *particle+verb* combination, even if the two elements are inserted separately into the syntactic component (sometimes they can even occur in two different clauses). The annotations to be used can range from the most specific, idiomatic to the most general, productive, depending on the nature of individual particle verbs.

(D) The initial implementational steps in our HunGram project have been successful. One of the crucial issues was whether the output of our computational program's morphological analyzer, which, naturally, takes orthography seriously, treats a *particle+verb* combination as a single word, can be manipulated in such a way that, as a rule, the two elements should occupy distinct syntactic position in our LFG-style constituent-structure representation. This turned out to be possible, because the analyzer has a partially different morphological-informational symbol (in more technical terms: *tag*) associated with the particle when it is part of a complex word. Compare (8a), the tag-set for the simple verb *megy* 'go' with (8b), the tag-set for the one-word particle verb *ki-megy* 'go out'.

- (8) a. *megy*: *megy* +Verb +PresInd +Indef +Sg +3P
b. *kimegy*: *ki* +Prefix+ *megy* +Verb +PresInd +Indef +Sg +3P
c. *ki*: +Prefix

The important point here is that when a particle itself is analyzed as a separate word, it gets the tag show in (8c), while when it is part of a single complex verb (word), its tag contains an extra "+" symbol. (Notice that this is the marked, exceptional case: all the other tags in (8) have only one (initial) + symbol). This special tag-representation for the "part-of-a-word" use of the particle enables us to write a syntactic rule with two designated positions: one for the particle and another for the verb, which we have tested implementationally.

3.2 *English particle verbs*

As we pointed out above (cf. Sections 1. and 2.2), the behavior of English particle verbs even more easily and straightforwardly motivates a syntactic analysis than that of their Hungarian counterparts. The two most important factors are as follows: (i) the verb and the particle are always spelt as two distinct words in English, (ii) in English, these combinations much less frequently serve as input to derivational processes. This explains why there are several fundamentally syntactically oriented accounts of English particle verbs in the generative literature.

As far as the implementational side is concerned in the ParGram project, the analysis in the English grammar is basically also syntactic, with the obviously necessary support from the lexical component in the entries for the verb and the particle.

In actual fact, it was this syntactically biased approach to English particle verbs both theoretically and implementationally that additionally motivated the direction in which we set out to develop our HunGram account: uniformity not only in the analysis of a complex set of phenomena within one language, but also uniformity across accounts of comparable phenomena in various languages. In the ParGram project, these particle verbal implementational issues have been addressed recently with respect to English, Hungarian and German. For a detailed discussion of the present state of affairs, see Forst-King-Laczkó (2010). They offer a novel, uniformly syntactic implementational approach, one of whose most significant advantages is that it efficiently and elegantly captures the productive use of particle verbs.

In the next subsection, we can only confine ourselves to outlining the most salient general and implementational aspects of the analysis of the major types of English and Hungarian particle verbs. For reasons of space-limitation, the discussion and the representations below are considerably simplified and sketchy, and we refer the reader to Forst-King-Laczkó (2010) for further details of this analysis.

3.3 *Main aspects of the analysis*

3.3.1 *Structural issues*

We illustrate the basic LFG-style grammatical representation of particle verb constructions via the examples in (9) and (10):

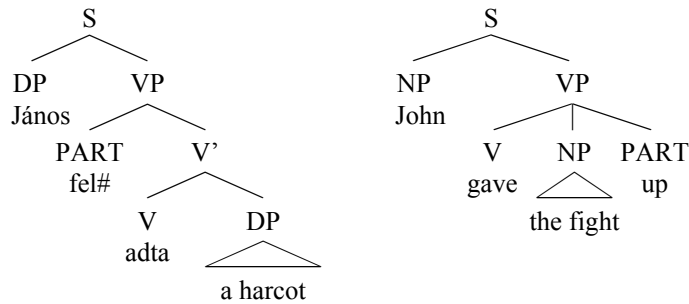
- (9) *János fel-ad-ta a harc-ot.* (=1b)
 John up-give-past.3sg the fight-acc
 ‘John gave up the fight.’

- (10) a. *John gave the fight up.*
 b. *János nem ad-ta a harc-ot fel.* (=4b)
 John not give-past.3sg the fight-acc up
 ‘John did not give the fight up.’

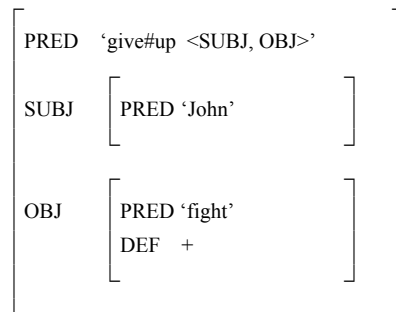
As far as the placement of the particle is concerned, these examples represent two basic patterns: the particle is (left-)adjacent to the verb (9), or it is separated from the verb (10). The

respective *constituent structures* are in (11a). The simplified *functional structure* of both (9) and (10a) is in (11b). These are the two core syntactic levels of representation in LFG.

(11) a.



b.



Economy considerations force us to add only the following comments on the representations in (11). See Bresnan (2001) for a detailed overview of LFG.

(A) In Lexical-Functional Grammar c(onstituent)-structures (cf. 11a) and f(unctional)-structures (cf. 11b) are two distinct, parallel levels of syntactic representation. The former are designed to capture surface properties of specific constructions like word order and constituency. The examples in (10) and the one in (9) differ in terms of linear ordering, and that difference is reflected in the corresponding c-structures. F-structure is the storage place for grammatically relevant features and predicate-argument relations, both of which may be relatively invariant across languages, as is the case with our English and Hungarian examples. (9) is the functional and the semantic equivalent of the English (10a), therefore their f-structure representation is going to be identical (at least in details that are important for us now, we disregard irrelevant differences).

(B) As (11a) shows, we assume that, despite the standard Hungarian orthographical convention, the particle occupies a separate syntactic position even when it immediately precedes the verb.

(C) Furthermore, at the present stage of our research, we postulate that ordinary particles, reduced arguments (see 3.1) and focussed elements compete for the same preverbal position. We argue for such an analysis in detail in a forthcoming publication.

3.3.2 Major types of particle verbs

Theoretical and implementational considerations necessitate a differential treatment of the compositional and the non-compositional particle verb constructions. There are in principle two options in the case of non-compositional particle verbs like *give up* and its Hungarian

counterpart *felad*. Either we represent the particle verb in the lexicon by creating a single entry for it, or we list both the particle and the verb as separate entries with additional information on their co-occurrence as a syntactic construction. The latter solution is more desirable for reasons of economy and for the purposes of the implementation. It has to be indicated in the verbal entry that the particle can co-occur with it, and the argument structure of the resulting new predicate is to be specified (cf. the third line of (12a)).

- (12) a. *ad*: V { (↑PRED) = ‘ad <(↑SUBJ) (↑OBJ) (↑OBL)>’
 |@(CONCAT (↑PRT-FORM) `# %stem %NewPred)
 (↑PRED) = ‘%NewPred <(↑SUBJ) (↑OBJ)>’
 (↑PRT-FORM) = c fel }.
- b. *fel*: PART (↑PRT-FORM) = fel.

The lexical entry in (12a) shows that the Hungarian verb *ad* (as well as the English verb *give*) can either be a triadic or a diadic predicate. The latter is the case when the verb forms a complex unit with the particle *fel* ‘up’. This concatenation of the particle and the verb is achieved with the help of variables in the lexicon (cf. the second line of (12a)). But since the verb and the particle are listed separately in the representation, it also becomes possible to model the syntactic independence of the particle.

The functional-structure in (11b) and the complex entry in (12a) both show that the particle does not contribute any grammatical or semantic features in this sort of particle verb formation. Compositional particle verb constructions require a differential treatment. Here we discuss two basic types, and refer the reader to Forst-King-Laczkó (2010) for a more detailed discussion.

Aspectual particles like *away/el* in (13) change the semantics of the verbal predicate predictably by specifying a given aspectual value.

- (13) a. *John was typing away.*
 b. *János el-gépelget-ett.*
 John away-type-past.3sg
 ‘John was typing away.’

This grammatical feature is encoded in the lexical entry of the particle:

- (14) *away/el*: PART (↑TNS-ASP ASPECT) = continuous
 (↑PRT-FORM) = away/el.

In any other respects, the particle in (13) is semantically empty. Its sole function is to contribute the +continuous aspectual feature to the f-structure of the clause projected by the verb. (15) represents a different case:

- (15) a. *After the accident the driver drove on.*
 b. *A baleset után az autós tovább-hajt-ott.*
 The accident after the motorist.nom further-drive-past.3sg
 ‘After the accident the motorist drove on.’

Here the particle *on/tovább* contributes its basic spatial meaning to the predication, and it essentially functions as an oblique argument of the base verb. The respective lexical entries are as follows:

- (16) a. *on/tovább*: PART (↑OBL PRED) = ‘on’.
 b. *drive/hajt*: V (↑PRED) = ‘drive <SUBJ, OBL>’.

This (use of the) particle is predicative (16a), and it spells out the directional argument of the base verb in (16b).

The constructions represented in (13) and (15) are productive. The particle in (16a) can combine with any motion predicate with the meaning ‘continue moving along a path’, and the particle in (14) quite freely combines with agentive predicates to change their lexical aspect. Thus, unlike in the non-compositional cases, cf. (12), it seems motivated not to list these *particle+verb* combinations individually in the lexicon but to form them directly in the syntax. As is argued in Forst-King-Laczkó (2010), a syntactic analysis requires no specific devices in these cases. In particular, there is no need to call the concatenation template (12) since no new lexical entry is formed in (13) and (15).

4 Conclusion

In this paper, we have presented an LFG-theoretic framework which is capable of handling both the lexical and syntactic properties of different types of particle verb constructions in English and Hungarian. We have offered an overview of the current state of analysis of these constructions in our computational grammar writing project. We intend to extend this initial setup in especially the following two directions: (i) the grammar of verbal modifiers in general and their complementarity with particles in particular, (ii) a proper treatment of the fact that particle verbs in Hungarian systematically feed various word formation processes.

References

- Ackerman, Farrell & Webelhuth, Gert (1998): The composition of (dis)continuous predicates: Lexical or syntactic? *Acta Linguistica Hungarica* 44, 317-340.
- Bresnan, Joan (2001): *Lexical-Functional Syntax*. Oxford: Blackwell.
- Butt, Miriam, King, Tracy Holloway, Niño, Maria-Eugenia & Segond, Frédérique (1997): *A Grammar Writer’s Cookbook*. Stanford: CSLI Publications.
- den Dikken, Marcel (1995): *Particles. On the Syntax of Verb-Particle, Triadic, and Causative Constructions*. Oxford: Oxford University Press.
- É. Kiss, Katalin (1987): *Configurationality in Hungarian*. Budapest: Akadémiai Kiadó.
- É. Kiss, Katalin (2005): First steps towards a theory of the verbal particle. In: Piñón, C. & Siptár, P. (eds.): *Papers from the Düsseldorf Conference. Approaches to Hungarian* 9. Budapest: Akadémiai Kiadó, 57-88.

Rákosi György, Laczkó Tibor & Csernyi Gábor:
*On English phrasal verbs and their Hungarian counterparts:
from the perspective of a computational linguistic project*
Argumentum 7 (2011), 80-89
Debreceni Egyetemi Kiadó

- Forst, Martin, King, Tracy H. & Laczkó, Tibor (2010): Particle verbs in computational LFGs: Issues from English, German, and Hungarian. In: Butt, M. & King, T.H. (eds.): *Proceedings of the LFG '10 Conference*. Stanford: CSLI Publications, 228-248.
Online: <http://csli-publications.stanford.edu/LFG/15/lfg10.pdf>
- Johnson, Kyle (1991): Object positions. *Natural Language and Linguistic Theory* 9, 577-636.
- Kayne, Richard S. (1998): Overt vs. covert movement. *Syntax* 1, 128-191.
- Koizumi, Masatoshi (1993): Object agreement phrases and the split VP hypothesis. In: Phillips, C. & Bobaljik, J.D. (eds.): *Papers on Case and Agreement I. MIT Working Papers in Linguistics* 18, 99-148.
- Komlósy, András (1992): Régensek és vonzatok. In: Kiefer, F. (ed.): *Strukturális Magyar Nyelvtan I. Mondattan*. Budapest: Akadémiai Kiadó, 299-528.

Dr. György Rákosi
University of Debrecen
Institute of English and American Studies
Pf. 73
H-4010 Debrecen
rakosi.gyorgy@arts.unideb.hu

Dr. Tibor Laczkó
University of Debrecen
Institute of English and American Studies
Pf. 73
H-4010 Debrecen
laczko.tibor@arts.unideb.hu

Gábor Csernyi
University of Debrecen
Institute of English and American Studies
Pf. 73
H-4010 Debrecen
gabor.csernyi@arts.unideb.hu