Swedish prepositions are not pure function words

Lars Ahrenberg
Department of Computer and Information Science
Linköping University
`lars.ahrenberg@liu.se`

Abstract

As for any categorial scheme used for annotation, UD abound with borderline cases. The main instruments to resolve them are the UD design principles and, of course, the linguistic facts of the matter. UD makes a fundamental distinction between content words and function words, and a, perhaps less fundamental, distinction between pure function words and the rest. It has been suggested that adpositions are to be included among the pure function words. In this paper I discuss the case of prepositions in Swedish and related languages in the light of these distinctions. It relates to a more general problem: How should we resolve cases where the linguistic intuitions and UD design principles are in conflict?

1 Introduction

The Universal Dependencies Project, henceforth UD, develops treebanks for a large number of languages with cross-linguistically consistent annotations. To serve this mission UD provides a universal inventory of categories and guidelines to facilitate consistent annotation of similar constructions across languages (Nivre et al., 2016). Automatic tools are also supplied so that annotators can check their annotations.

Important features of the UD framework are linguistic motivation, transparency, and accessibility for non-specialists. This is a delicate balance and it would be hard to claim that they always go hand in hand.

As for any categorial scheme used for annotation, UD abound with borderline cases. The main instruments to resolve them are the guidelines, which in turn rests on the UD design principles. UD makes a basic distinction between content words and function words and adopts a policy of the primacy of content words. This means that dependencies primarily relate content words while function words as far as possible should have content words as heads. Thus, multiple function words related to the same content word should appear as siblings, not in a nested structure. However, as content words can be elided and (almost) any word can be negated, conjoined or be part of a fixed expression, some exceptions must be allowed.

It is suggested that there is a class of function words, called 'pure function words', with very limited potential for modification. This class 'includes auxiliary verbs, case markers (adpositions), and articles, but needs to be defined explicitly for each language' (UD, 2017c). The choice of candidates is motivated by the fact that some languages can do without them, or express corresponding properties morphologically. In the case of prepositions in Germanic languages the similarity with case suffixes in languages such as Finnish or Russian is pointed out. As noted in (Marneffe et al., 2014) this is a break with the earlier Stanford Dependencies framework, one of the UD forerunners. Thus, a special dependency relation, *case*, is associated with prepositions in their most typical use (see Figures 1 and 2).

It is not clear from the characterization of pure function words if the identification must be made in terms of general categories or in terms of individual words. Given the examples it seems that a combination of part-of-speech categories and features may suffice. We can note that subjunctions are not listed among the pure function words. On the contrary, the UD guidelines include an example, 'just when you thought it was over' where 'just' is analysed as an adverbial modifier, *advmod*, of the subjunction 'when' (UD, 2017c).

The aim of this paper is to discuss how well Swedish prepositions fit the category of pure func-

tion words. In the next section I will present data illustrating the range of uses for Swedish prepositions and review the current UD guidelines for their analysis. In section 3, the actual analysis of prepositions in Swedish and other Germanic treebanks will be reviewed. Then, in section 4, the data and the analyses will be discussed with a view to the problems of fuzzy borders in UD, in particular as regards the classification of words. Finally, in section 5, the conclusions are stated.

2 Prepositions in Swedish

Consider the following Swedish sentences, all containing the preposition *på* (on, in):

(1) Max säljer blommor på torget
    Max sells flowers in the market-square
(2) Max sätter på kaffet
    Max is making coffee
(3) Max litar på Linda
    Max trusts Linda

A traditional descriptive account of these sentences goes as follows: In (1) 'på torget' is an adverbial, providing an answer to the question 'Var säljer Max blommor?' (Where does Max sell flowers). It may be moved to the front of the clause as in 'På torget säljer Max blommor'. In (2) 'på kaffet' is not a constituent; it does not answer a question about the location of something, and it cannot be moved to the front. Instead, it is construed with the verb as a verb particle, which means that it will receive stress in speech, and the word 'kaffet' is analysed as an object of the complex *sätta på*. In (3) 'på Linda' is a prepositional (or adverbial) object of the verb. It can be moved to the front but it does not express a location, and the preposition is not stressed.

In UD, the distinction between adverbials and adverbial objects is not made. Thus, for both sentences (1) and (3) the preposition will be assigned the dependency *case* in relation to its head nominal, which in turn will be an oblique dependent (*obl*) of the main verb, as depicted in Figures 1 and 2. The prime motivation is this: *The core-oblique distinction is generally accepted in language typology as being both more relevant and easier to apply cross-linguistically than the argument-adjunct distinction* (UD, 2017a). This position then assumes that we should only have a binary
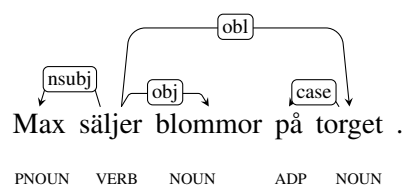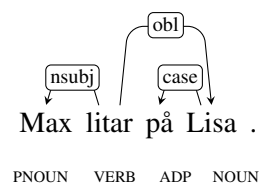

Figure 1: Partial analysis of sentence (1)

Figure 2: Partial analysis of sentence (3)

division of the nominal dependents of a verb[1].

Although its dependency is different in (2), *på* can be tagged as an ADP there as well. Alternatively, given the possibility of (4), it may be regarded as an adverb (ADV).

(4) Kaffet är på
    Coffee is on i.e., in the making

Arguably, in (4) 'på' must be analysed as the root as the verb 'är' (*is*) is a copula here, another type of function word in UD. Hence it shouldn't be a function word, specifically not a pure function word. Whatever decision we take on 'på' in these examples they illustrate that the border between content words and function words is not always clear-cut.

Another difference between (1) and (3) is the interpretation of the preposition. In (1) it seems to have more semantic content than in (3). It has a lexical meaning which is independent of the main verb and which can be used in construction with any nominal that refers to an object with a horizontal surface held parallel to the ground. This independence is also shown by the possibility of using a phrase of this type as the title of a story or an image caption. This lexical meaning is not present in (3) where the occurrence and interpretation is wholly dependent on the verb. Thus, (1) can answer a question such as 'Vad händer på torget?' (What's happening in the market square?), whereas (3) cannot answer the question 'Vad händer på Linda?' (What's happen-

---

[1]Formally, further specification could be implemented via the :-technique, but this is reserved for relations assumed to be language-specific

ing on Linda?).

The fact that we can distinguish between (1) and (3) does not necessarily mean that we must do so for any occurrence of a phrase that UD considers to be oblique. In other cases the classification can be much more difficult. UD, however, demands of us to make many other distinctions at the same level of difficulty, including that between pure and non-pure function words and, as we just saw, that between adpositions and adverbs. Other fundamental distinctions concern clauses vs. noun phrases, which have distinct sets of dependencies and largely distinct sets of constituents. For example, adpositions are regarded as different from subjunctions, presumably because typically adpositions are found with noun phrases while subjunctions are found with clauses.

For a scheme such as UD with its 17 part-of-speech categories and some 37 dependency relations there are many borderline cases. The design principles can help us decide, but sometimes they are not informative enough, and may be in conflict with linguistic intuitions. It can also be observed that the design principles are much more developed for syntax than for parts-of-speech.

2.1 Recommended UD analyses

Sentences (1)-(4) showed us different uses of Swedish prepositions. We have seen how UD treats (1) and (3). For (2) the relation *compound:prt* is used, telling us that we are dealing with a specific subtype of multiword expressions. For (4), it was suggested that the preposition is root, although possibly re-categorized as an adverb.

Sentences (5)-(8) illustrates other uses of prepositions in Swedish. They may be stranded as in (5a), or isolated as in (5b) and (5c). They may introduce a VP as in (6a), or a clause as in (6b). They may be modified as in (7) and they may follow after an auxiliary verb, as in (8).

UD has recommendations for all of these cases. For stranded and isolated prepositions the recommendation is *When the natural head of a function word is elided, the function word will be promoted to the function normally assumed by the content word head.* Thus, in (5a) 'på' will relate to 'lita' via the *obl* relation. (5b) and (5c) can be treated in the same way.

Another option used in both the English treebank and Swedish_LinES is to let the moved nom-

(5a)	Vem kan man lita på?		
	Who can you trust?		
(5b)	Max fick en bild att titta på.		
	Max got a picture to look at.		
(5c)	Max sitter därborta och Linda sitter bredvid.		
	Max sits over there and Linda sits beside.		
(6a)	Max tröttnade på att vänta.		
	Max got tired of waiting		
(6b)	Max litar på att Linda ringer.		
	Max trusts that Linda will call.		
(7)	Max blev träffad mitt på näsan.		
	Max was hit right on his nose.		
(8)	Jag måste i nu.		
	I have to get in(to the water) now.		

inal be the head of the *case* dependency. An English example (from UD_English item 0029) similar to (5a) is shown in Table 1. This solution can be applied in (5a) and (5b) but not in (5c) except via an enhanced dependency. In (5c), the lack of an NP after the preposition can be attributed to the discourse context.

1	He	PRON	3	nsubj
2	obviously	ADV	3	advmod
3	had	VERB	0	root
4	no	DET	5	det
5	idea	NOUN	3	obj
6	what	PRON	9	obl
7	he	PRON	9	nsubj
8	was	AUX	9	aux
9	talking	VERB	3	ccomp
10	about	ADP	6	case

Table 1: Example analysis from the UD_English treebank.

For (6a) and (6b) the recommendation is to use the *mark* relation, which is otherwise typically used for subjunctions. The issue was discussed in the UD forum (issue #257) with the conclusion that the preposition could keep its POS tag while being assigned the relation *mark*. The main argument was that if the relation would be *case*, it would not be possible to reveal an occurrence of an ADP with the relation *case* to a VERB as an error automatically. This is a good way to promote consistency of annotation, but is quite arbitrary from a linguistic point of view.

For (7) the UD recommendation is that the ad-

verb 'mitt' (right) modifies the head noun 'nä]san' (nose) rather than the preposition. This is in line with the desired constraint that function words should not have dependents, but is contrary to semantic intuitions and the fact that the phrase '*träffad mitt näsan' is ungrammatical. Given that subjunctions are allowed to have adverbs as modifiers, just as predicates, the asymmetry in analyses seems unmotivated. Note that 'på' in 'mitt på' has spatial lexical meaning, in the same way as 'när' (when) has a distinct temporal meaning in a phrase such as 'just när' (just when).

Sentence (8) is similar to (4) involving an auxiliary rather than a copula. The solution can be the same, i.e., analysing the token *i* (in) as root but tagged as adverb rather than as adposition.

3 Adpositions in UD treebanks

As expected, the most common dependency relation assigned to adpositions in Germanic treebanks is *case*. Other alternatives fall far behind. Overall relative frequencies for the treebanks of the Scandinavian languages, English, and German from the v2.0 release (Nivre et al., 2017) are shown in Table 2. The counts are based on the train- and dev-treebanks joined together. Note that Table 2 does not show all alternatives.

Treebank	case	mark	cmp:prt
Danish	0.842	0.110	0.009
English	0.923	0.005	0.041
German	0.952	0.008	0.029
No-Bokmaal	0.801	0.115	0.063
No-Nynorsk	0.795	0.108	0.067
Swedish	0.851	0.059	0.027
Sw-LinES	0.970	0.000	0.009

Table 2: Relative frequencies for some dependency relations of ADP tokens in seven UD v2.0 treebanks.

Of the five Scandinavian treebanks all except Swedish-LinES frequently give an ADP the relation *mark*. Swedish-LinES only has one example; English and German have only a few instances,

The rightmost column of Table 2 shows that there are marked differences in relative frequency for the relation *compound:prt* as applied to ADP tokens in these treebanks. This difference gets its explanation when we look at what parts-of-speech these particles are assigned, shown in Table 3. English and the Norwegian treebanks have a clear

dominance for adpositions, German a clear majority, while adverbs are in the majority in the Swedish and Danish treebanks .

Treebank	ADP	ADV	other
Danish	92 (0.25)	269	0
English	802 (0.98)	16	3
German	901 (0.61)	562	3
No-Bokmaal	2308 (1.00)	0	0
No-Nynorsk	2585 (1.00)	0	0
Swedish	236 (0.35)	411	37
Sw-LinES	64 (0.10)	500	62

Table 3: Frequencies for parts-of-speech assigned the relation compound:prt in different treebanks.

Given the close relationship between these languages it is hard to believe that the differences are solely due to language differences. It is more likely that the annotators have followed different principles both as regards parts-of-speech and dependency relations. For instance, the Norwegian treebanks analyse as ADP a number of verb particles that in the Swedish treebanks are analysed as adverbs, such as (from the Norwegian-Bokmaal treebank) *bort* (away), *hjem* (home), *inn* (in), *opp, oppe* (up), *tilbake* (back), *ut* (out).

It is also interesting to look at cases where a preposition has been analysed as the head of a dependency. As expected we find many instances of *fixed* and *conj* in most of the treebanks, but the distributions are not at all similar, as shown in Table 4. The differences are probably due both to differences in treebank-specific guidelines, and to errors in applying them. An interesting fact, however, is that all of them have instances of *advmod*, and just not for negations. Some of these are likely to be errors, but some reasonable examples are the following: *omedelbart efter lunch* (Sw-LinES; immediately after lunch), *langt fra hele* (No-Bokmaal; far from all), *andre ting kan han gå sterkt imot* (No-Nynorsk; other things he may go strongly against), *up to 40 rockets* (English), and *genauso wie in Portugal* (German; just as in Portugal). All of them also have instances of prepositions governing a copula.

Table 5 shows data on part-of-speech assignments for all words that have been analysed as an adposition at least once. Note that the treebanks may have errors so that the figures should not be taken as exact, but the differences in distributions are nevertheless interesting. The first col-

Treebank	fixed	conj	advmod	other
Danish	0.34	0.03	0.03	0.61
English	0.65	0.03	0.03	0.28
German	0.02	0.04	0.06	0.87
No-Bokmaal	0	0.12	0.14	0.74
No-Nynorsk	0	0.10	0.14	0.76
Swedish	0.85	0.02	0.05	0.08
Sw-LinES	0.61	0.08	0.02	0.29

Table 4: Relative frequencies for dependency relations headed by ADP tokens in seven UD v2.0 treebanks.

umn shows the number of words that have ADP as their only part-of-speech tag. We can see that the Norwegian treebanks are markedly different from the others in recognizing a very high number of non-ambiguous adpositional words.

The two Swedish treebanks have different distributions. This may partly be due to differences in genre, but also to differences in the specific guidelines used. In both treebanks, however, we will find tokens that are found in both ADP and ADV, other tokens that are sometimes ADP and sometimes SCONJ and yet others having all three tags.

Treebank	Instances			
	1	2	3	≥ 4
Danish	20	20	13	1
English	34	29	31	35
German	48	48	26	28
No-Bokmaal	116	34	6	3
No-Nynorsk	121	33	13	4
Swedish	44	20	8	2
Sw-LinES	35	39	11	3

Table 5: Number of different POS assignments for words that have been tagged as ADP at least once in different UD v2.0 treebanks.

4 Discussion

The previous section shows with no uncertainty that the UD guidelines and design principles are not followed uniformly by all treebank developers. There are natural explanations for this, such as differences in the original pre-UD guidelines for the different treebanks, and a lack of time for reviewing treebank data. I suspect, however, that there is also a certain conflict between the UD principles and linguistic intuitions of treebank annotators.

4.1 UD part-of-speech categories

The POS tags used in UD are 17. They form an extension of the 12 categories of (Petrov et al., 2012). We note in particular the addition of the category SCONJ for subjunctions.

A treebank is allowed not to use all POS tags. However, the list cannot be extended. Instead, more fine-grained classification of words can be achieved via the use of features (UD, 2017b). In annotations only one tag per word is allowed, and it must be specified.

The basic division of the POS tags is in terms of open class, closed class, and other. The point of dividing them this way is unclear, since most of the syntactic principles referring to POS tags as we have seen use the categories content words and function words[2].

The POS tags that will be of interest here are: ADP(ositions), ADV(erbs), and SCONJ (subordinating conjunctions). These classes are singled out as they are hard to separate consistently and account for a large share of the homonymy found in Table 4. Moreover, many linguists such as (Bolinger, 1971; Emonds, 1985; Aarts, 2007) have questioned the linguistic motivations behind the distinctions. It may also be a problem for users who may find the distinctions less than transparent. The definitions of these parts-of-speech in the UD documentation are not always of help as they focus on the most common and prototypical examples. The categories ADP and SCONJ have partly overlapping definitions. An SCONJ is described as typically incorporating what follows it as a subordinate clause, while an ADP, in addition to NP:s, may have a clause as its complement, when it functions as a noun phrase[3].

4.2 Prepositions and subjunctions

As many other prepositions 'på' can in itself not be used as a subjunction. Removing the infinitive marker from (6a) or the subjunction 'att' from (6b) results in ungrammaticality. There are, however, in Swedish, as in English, many words that can be used both ways, specifically those expressing temporal or causal relations, and comparisons. Common examples are *sedan* (since), *på grund av* (because of), *än* (than), *efter* (after), *innan* (before)

---

[2]A wish for an exact definition was expressed by Dan Zeman in the UD forum issue #122

[3]This description may refer to constructions headed by a gerund, free relatives, and the like, but is fairly non-transparent for a non-linguist.

and *före* (before). The latter two are the subject of a constant debate on grammatical correctness in Swedish where purists would hold that one is a preposition and the other a subjunction, but speakers tend not to follow suit.

For these words the meaning is quite the same whether what follows is a noun phrase or a clause. This fact has been taken as an argument that the prepositions and subjunctions are sufficiently similar to be regarded as one category. The difference can be seen as one of complementation which, in the case of verbs, is not sufficient to distinguish two part-of-speech categories (Emonds, 1976). In UD it may be seen as logical to distinguish the two, given the emphasis on the distinction between noun phrases and clauses. On the other hand, this leads to certain oddities of the kind that allow these words to have dependents when they are subjunctions, but not when they are adpositions.

Sentence (7) illustrated the fact that Swedish prepositions may be modified. There are a number of adverbs that can modify prepositions and some of them can modify prepositions and subjunctions alike. Examples are *alldeles*, (just) *rakt*, *rätt* (both meaning 'right'), *precis* (exactly). Examples are given in (9) and (10)[4]. Spatial and temporal prepositions may actually be modified by noun phrases indicating distance in space and time, as in (11)-(12). Thus, the potential of Swedish prepositions to be modified is quite equal to that of subjunctions.

(9) Hon kom precis före (oss).
     She came just before us.
(10) Hon kom precis innan (vi kom).
     She came just before we did.
(11) Hon kom en timme före (oss).
     She came an hour before us.
(12) De sitter två rader bakom (oss).
     They're sitting two rows behind (us).

4.3   Adpositions as *mark*

As noted above, the current recommendation for the analysis of (6b) in UD is that the preposition 'på' should be assigned as a dependent of the head of the following clause, in this case the verb 'ringer' (call). This is so because 'på' is not a verb particle in (6b) and it can only attach to the main verb if it functions as a particle. Then there are two

relations to choose from: *case* or *mark*. None of them is ideal; if we choose *case* we add a dependency to clauses which seems to be rare in other languages and which blurs the distinction between clauses and noun phrases; if we choose *mark* we add a property to adpositions that make them more similar to subjunctions.

Another issue is the relation assigned to the clause itself. In the Swedish treebank it is *advcl*. Normally, a clause introduced by the subjunction 'att' would be *ccomp*. However, in the same way as a noun phrase introduced by a preposition would be *obl* it is logical to use *advcl*. Then again, this makes the difference between prepositions and subjunctions fuzzier.

Now, the clause 'att hon ringer' is as independent in (6b) as 'Linda' is in (3). It can be moved to the front ('Att Linda ringer litar Max på'), it can be the target of a question ('Vad litar du på?') and it can be focused: ('Är det något jag litar på är det att Linda ringer'). Thus, there is an NP-like flavour of these verb-headed structures, just as for prepositional objects, suggesting that *ccomp* may be a viable alternative nevertheless.

An interesting observation is the UD recommendation for the analysis of comparative subjunctions such as 'än' (than) and 'som' (as). Since they can virtually combine with phrases of any kind, including clauses, prepositional phrases, and noun phrases, and may use nominative pronouns in the latter case, the question arises as how they should be analysed[5]. In a phrase such as 'än Max' (than Max), 'än' could be an ADP with relation 'case', or it could be an SCONJ with relation 'mark'. However, given the new option why not an ADP with relation 'mark'? We may note that the most detailed analysis of Swedish grammar regards all instances 'än' as subjunctions and allows 'subjunction phrases' (Teleman et al., 2010). We may even regard a phrase such as 'than on Sundays' to have two case markers. The choice seems arbitrary.

The UD decision not to make a distinction between complements and adjuncts serves the interests of transparency and ease of annotation. It means, however, that the UD analyses do not make all the differences that can be made so that what is arguably different phenomena gets identical analyses. When it comes to part-of-speech annota-

---

[4]The brackets indicate material that can be left out without loss of grammaticality.

[5]In the UD_Swedish treebank these two words are actually tagged CCONJ.

tion the situation is a bit different. Extensions are not allowed, but sub-categorization is required (or requested) for some parts-of-speech, specifically pronouns and determiners via features.

We have noted that UD principles are fewer and less developed for part-of-speech categories than for syntax. In particular, there are no principles regulating the degree of homonymy. The current framework suggests that Swedish has three homonyms for the word *innan* (before) although they all have the same meaning. If the framework allowed fewer part-of-speech categories, the degree of homonymy would decrease and annotation would be easier. In particular the degree of homonymity of prepositions could decrease (cf. Table 5).

The idea that prepositions must be pure function words due to their presumed equivalence with case endings can be put into question. The sentences (5)-(8) all illustrate uses that are specific to prepositions. Also, when prepositions have clear semantic content they can be modified in various ways, just as corresponding subjunctions. At the same time there are subjunctions, notably the complementizers such as 'att' (that) and 'som' (that, which) that cannot be modified easily, not even negated. This fact speaks against the view that all adpositions should be put into one basket as pure function words whereas subjunctions should be put into another.

4.4 Prepositions and adverbs

The words that need to be regarded as both prepositions and adverbs are numerous. It includes common prepositions as illustrated in (4) and (8) and a number of prepositions expressing spatial relations such as *utanför* (outside), *innanför* (inside), *nedanför* (down, below), *nerför* (down), and *uppför* (up). For the latter the differences in meaning are minimal, whereas for the more common prepositions such as *på* (on) or *i* (in), the meanings may be varied.

We note that when a sequence of a preposition and a noun is lexicalized into a fixed expression, the result is almost always something adverbial. Some examples are *i kväll* (tonight), *i tid* (in time), *på stört* (at once), *på nytt* (again), *på land* (ashore). In some treebanks, notably UD_Swedish, the preposition keeps its part-of-speech while being assigned the dependency 'advmod'.

We can apply the same arguments and counter-arguments in this case as for the previous case. The difference may be seen just as a difference in complementation, where some prepositions can have both a transitive and an intransitive use and still be prepositions. However, if prepositions are pure function words, and adverbs are not, UD forces a distinction to be made.

5 Conclusions

We have observed that the UD design principles are more elaborated for syntax than for parts-of-speech. This could be interpreted as a recommendation not to take parts-of-speech too seriously; we may assume as much homonymy as the syntactic design principles demands of the data. On the other hand, what the non-expert user would consider to be 'the same word' in a given language should also be given some consideration. I would like to see an attempt to define UD parts-of-speech in more detail, preferably as lists of properties, taking gradience into account. (Aarts, 2007) proposes a simple model for this purpose, which may be taken as an inspiration. In the case of some adpositions and subjunctions such as *before*, however, Aarts sees no difference as it is only a question of complementation possibilities.

To join the categories ADP and SCONJ and include some adverbial uses as well would reduce ambiguity in part-of-speech assignment considerably. If it is desirable to maintain the difference it can be done in the feature column by, say, a type feature (AdpType).

Swedish prepositions, and those of other Scandinavian languages, are more varied in their usage than case suffixes. At the same time, they share important properties with subjunctions and adverbs. Semantically they span the same domains and they share typical positions. The current UD principles allow Swedish prepositions to share the relation *mark* with subjunctions, and the relation *compound:prt* with adverbs. In addition, they share those relations that are common to all parts-of-speech, such as *fixed* and *conj*. Linguistically, at least, Swedish prepositions can be modified adverbially when they have semantic content. Thus, Swedish prepositions are not pure function words.

The possibility to take dependents seems to be more on the level of individual words than a property across the board for any part of speech. In

Swedish it can be found also with subjunctions, in particular complementizers such as *att*, (that), conjunctions such as *och* (and), and adverbs such as *ju* (approx. 'you know'). With this in mind, it can be questioned what benefits are gained from dividing function words further into pure and not-so-pure on the basis of parts-of-speech.

References

Bas Aarts. 2007. *Syntactic Gradience: The Nature of Grammatical Indeterminacy*. Oxford University Press, Oxford.

Dwight L. Bolinger. 1971. *The Phrasal Verb in English*. Harvard University Press, Cambridge, MA.

Joseph E. Emonds. 1976. *A Transformational Approach to English Syntax*. Academic Press, New York.

Joseph E. Emonds. 1985. *A Unified Theory of Syntactic Categories*. Foris, Dordrecht.

Marie-Catherine De Marneffe, Timothy Dozat, Natalia Silveira, Katri Haverinen, Filip Ginter, Joakim Nivre, and Christopher D. Manning. 2014. Universal stanford dependencies: a cross-linguistic typology. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, may.

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajič, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, and Daniel Zeman. 2016. Universal dependencies v1: A multilingual treebank collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*.

Joakim Nivre, Željko Agić, Lars Ahrenberg, Maria Jesus Aranzabe, Masayuki Asahara, Aitziber Atutxa, Miguel Ballesteros, John Bauer, Kepa Bengoetxea, Riyaz Ahmad Bhat, Eckhard Bick, Cristina Bosco, Gosse Bouma, Sam Bowman, Marie Candito, Gülşen Cebirolu Eryiit, Giuseppe G. A. Celano, Fabricio Chalub, Jinho Choi, Çar Çöltekin, Miriam Connor, Elizabeth Davidson, Marie-Catherine de Marneffe, Valeria de Paiva, Arantza Diaz de Ilarraza, Kaja Dobrovoljc, Timothy Dozat, Kira Droganova, Puneet Dwivedi, Marhaba Eli, Tomaž Erjavec, Richárd Farkas, Jennifer Foster, Cláudia Freitas, Katarína Gajdošová, Daniel Galbraith, Marcos Garcia, Filip Ginter, Iakes Goenaga, Koldo Gojenola, Memduh Gökrmak, Yoav Goldberg, Xavier Gómez Guinovart, Berta Gonzáles Saavedra, Matias Grioni, Normunds Grūzītis, Bruno Guillaume, Nizar Habash, Jan Hajič, Linh Hà M, Dag Haug, Barbora Hladká, Petter Hohle, Radu Ion, Elena Irimia, Anders Johannsen, Fredrik Jørgensen, Hüner Kaşkara, Hiroshi Kanayama, Jenna Kanerva, Natalia Kotsyba, Simon Krek, Veronika Laippala, Phng Lê Hng, Alessandro Lenci, Nikola Ljubešić, Olga Lyashevskaya, Teresa Lynn, Aibek Makazhanov, Christopher Manning, Cătălina Mărănduc, David Mareček, Héctor Martínez Alonso, André Martins, Jan Mašek, Yuji Matsumoto, Ryan McDonald, Anna Missilä, Verginica Mititelu, Yusuke Miyao, Simonetta Montemagni, Amir More, Shunsuke Mori, Bohdan Moskalevskyi, Kadri Muischnek, Nina Mustafina, Kaili Müürisep, Lng Nguyn Th, Huyn Nguyn Th Minh, Vitaly Nikolaev, Hanna Nurmi, Stina Ojala, Petya Osenova, Lilja Øvrelid, Elena Pascual, Marco Passarotti, Cenel-Augusto Perez, Guy Perrier, Slav Petrov, Jussi Piitulainen, Barbara Plank, Martin Popel, Lauma Pretkalnia, Prokopis Prokopidis, Tiina Puolakainen, Sampo Pyysalo, Alexandre Rademaker, Loganathan Ramasamy, Livy Real, Laura Rituma, Rudolf Rosa, Shadi Saleh, Manuela Sanguinetti, Baiba Saulīte, Sebastian Schuster, Djamé Seddah, Wolfgang Seeker, Mojgan Seraji, Lena Shakurova, Mo Shen, Dmitry Sichinava, Natalia Silveira, Maria Simi, Radu Simionescu, Katalin Simkó, Mária Šimková, Kiril Simov, Aaron Smith, Alane Suhr, Umut Sulubacak, Zsolt Szántó, Dima Taji, Takaaki Tanaka, Reut Tsarfaty, Francis Tyers, Sumire Uematsu, Larraitz Uria, Gertjan van Noord, Viktor Varga, Veronika Vincze, Jonathan North Washington, Zdeněk Žabokrtský, Amir Zeldes, Daniel Zeman, and Hanzhi Zhu. 2017. Universal dependencies 2.0. LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics, Charles University.

Slav Petrov, Dipanjan Das, and Ryan McDonald. 2012. A universal part-of-speech tagset. In *Proceedings of the Eigth International Conference on Language Resources and Evaluation (LREC 2016)*, Istanbul, Turkey, may.

Ulf Teleman, Erik Andersson, and Staffan Hellberg. 2010. *Svenska Akademins Grammatik*. Norstedts, Stockholm.

UD. 2017a. Core dependents in ud v2. http://universaldependencies.org/v2/core-dependents.html.

UD. 2017b. Morphology: General principles. http://universaldependencies.org/u/overview/morphology.html.

UD. 2017c. Syntax: General principles. http://universaldependencies.org/u/overview/syntax.html.