SLOVENIAN PARTICLE: NOT A SYNTACTIC CATEGORY

Jakob LENARDIČ[1]
[1]Institute of Contemporary History, Ljubljana

This paper claims that there is insufficient syntactic or semantic evidence to distinguish particles as a separate part of speech from adverbs, a distinction which is otherwise made in mainstream Slovenian descriptive linguistics and which is also used for the morphosyntactic annotations of Slovenian corpora. To this end, the paper presents a corpus investigation which probes into the clausal distribution of frequent words tagged as particles on the one hand and as adverbs on the other, showing that there are no appreciable differences there. The paper also proposes that the well-known *wh*-question test, which is otherwise used to determine sentence elements like subject, object, and adjunct, is a theoretically weak and empirically inconsistent criterion for determining category membership. Furthermore, the paper argues that common interpretative criteria used to distinguish particles from adverbs suffer from conceptual fallacies and are empirically unmotivated.

Keywords: particle, adverb, corpus linguistics, theoretical linguistics, syntax

1 INTRODUCTION

The aim of this paper is to challenge the traditional viewpoint of Slovenian descriptive linguistics according to which the words in (1) belong to a different syntactic category (or part of speech) from the prototypical adverbs in (2) – namely, to the syntactic category of the so-called particles. This viewpoint, which I will argue against, is advocated for by the Slovenian descriptive grammar *Slovenska slovnica* (Toporišič, 2000) and is maintained in the work of many other authors as well (e.g., Jakop, 2001; Krvina & Žele, 2018; Žele, 2014; Žele, 2015).

(1) Particles
   a. *morda* ‘possibly’
   b. *tudi* ‘also’

c. *ne* ‘not’
d. *naj* ‘should’

(2) Adverbs

a. *včeraj* ‘yesterday’
b. *danes* ‘today’
c. *zlahka* ‘easily’
d. *tam* ‘there’

The question of how to classify these words, and whether to make a division between them, is not only relevant for theoretical purposes, but also (and perhaps even more so) for language resource creation. The linguistic annotations of the major Slovenian corpora, such as the written reference corpus *Gigafida 2.0* (Krek et al., 2019), the *Corpus of academic Slovene KAS 2.0* (Žagar et al., 2022), and the computer-mediated-communication corpora of the *JANES* family (Fišer et al., 2020), follow the MULTEXT-East specifications (Erjavec, 2017),[1] which recognise particles as a category separate from adverbs. Similarly, the annotated corpora of the *ParlaMint* project (see Erjavec et al., 2023, for the latest, annotated, version), whose creation is spearheaded by the Slovenian CLARIN.SI consortium,[2] make use of the Universal Dependencies schema (De Marneffe et al., 2021), which likewise posits that particles constitute their own category (see also Dobrovoljc et al., 2023, for a description of the Slovenian Universal Dependencies guidelines).

In this paper, I will claim that there is insufficient linguistic evidence – either syntactic or semantic – that would warrant treating words like those in (1) as being appreciably different grammatically from the words in (2). In other words, there is no good reason not to assume simply that the words in both (1) and (2) are anything other than adverbs.

The paper is structured as follows. In Section 2, I first present a corpus investigation using the *Gigafida 2.0* corpus with which I try to determine if words tagged as particles substantially differ in their clausal distribution from words tagged as adverbs. In Section 3, I criticise the well-known *wh*-question test ac-

[1] See http://nl.ijs.si/ME/V6/msd/html/msd-sl.html for the current – i.e., sixth – version of the Slovenian specifications.
[2] https://www.clarin.si

cording to which particles unlike adverbs cannot be targeted for *wh*-question formation; I also argue that particles are not a semantically well-defined category either, and that many particles are in fact quite similar to adverbs interpretatively. In Section 4, I end the paper with a discussion which highlights the pitfalls of relying on mostly interpretative rather than syntactic criteria for determining parts of speech and argue for a syntax-first approach.

2 THE SYNTACTIC DISTRIBUTION OF PARTICLES

2.1 A tentative syntax for particles

In *Slovenska slovnica* 'The Slovenian Grammar', particles are defined as follows:

> Particles are an uninflected syntactic category: they are used to establish connections to the wider discourse; they express certain semantic flavours of individual words, parts of the clause or entire sentences; they are also used to derive syntactic modalities. Some particles are functionally similar to conjunctions, others to adverbs. Particles are not constituents of the clausal structure, in the framework of which they occur; rather, they are fragments that substitute elided clauses which can express additional contextual meanings. For instance, *Sosedovi imajo* ***samo*** *enega otroka* 'The neighbours have only one child → *Sosedovi imajo enega otroka, imeli pa bi jih lahko več* 'The neighbours have one child, but they could have more of them.' (Or: *Navadno je v družinah več kot en otrok* 'There is usually more than one child in a family'). About one quarter of the particles are homophonous with adverbs and conjunctons.'
>
> (Toporišič, 2000, p. 445; translation by JL)

Toporišič (2000) then goes on to propose various ways of classifying particles semantically, drawing up classes like *pozivni členek* 'particle of addressing' and

vrednotenjski členek 'particle of evaluation'.[3] In general, this approach and the definition above is primarily semantic in nature, with the only ostensibly syntactic characteristic being the idea that particles are not constituents of clausal structure, to which point I will return in Section 3.1.

If particles indeed constitute a different part of speech than adverbs as per the definition above, they should differ from one another in some obvious syntagmatic way, perhaps in the way that they are distributed in the clause.

On the face of it, many canonical particles, like *samo* 'only' in the definition above, do indeed seem to show a different clausal distribution from typical adverbs like *hitro* 'fast', primarily in relation to the clause initial position. In this respect, Marušič and Žaucer (2010) observe that the particle *že* 'already' is a phonologically weak element that cannot by itself host the so-called Wackernagel clitic cluster (Anderson, 1993). The cluster occupies the second syntactic position in the clause and has to follow a phonologically non-weak word to ensure well-formedness.[4] To see this, compare the grammaticality of the constructed examples (3a) and (4a),[5] in which the clitic cluster is typeset in bold and is directly preceded by an adverb, with the ungrammaticality of the examples in (3b) and (4b), in which the clitic cluster is directly preceded by the particle *že* 'already'.

(3) a. Že velikokrat **ji** **je** stopil na prste.
already manytimes she.Dat is stepped on toes
'He has stepped on her toes many times.'

[3] When exemplifying the various classes, Toporišič (2000) identifies phrasal particles as well, which to my mind further underscores the idea that particles taxonomically are not to be treated as a part of speech, but rather as a special interpretation of individual words, which need not be just adverbs, or of phrases. When for instance Smolej (2004) talks about particles, she uses the term to refer to phrases such as *konec koncev*, which are idiomatically used for discourse-organisational purposes; in terms of parts of speech, however, this is simply a noun complex in which the first noun is in nominative (i.e., *konec* '(the) end') and the second (i.e., *koncev* 'of ends') in genitive.

[4] See also Marušič (2008) for further observations on the seeming second-position placement of the clitic cluster.

[5] I will try to make it as clear as possible when I am talking about constructed or about corpus examples. Additionally, all the examples are glossed using the abbreviations of the Leipzig Glossing Rules (https://www.eva.mpg.de/lingua/resources/glossing-rules.php) – so for instance Dat in the interlinear glosses of example (3a) stands for Dative, 1Sg in example (5) for first person singular, and so on.

b. *Že **ji je** velikokrat stopil na prste.

(4) a. Že od včeraj **mi je** Peter dolžen 5 Eur.
already since yesterday I.Dat is Peter indebted 5 Eur
'Peter has owed me 5 Eur already since yesterday.'
b. *Že **mi je** Peter dolžen 5 Eur.

(Marušič & Žaucer, 2010, examples (2)–(3))

Similarly, the negator *ne* is also typically analysed as a phonologically weak element that in finite declarative clauses forms a syntactic unit with the finite verbal element (Ilc, 2008), together with which it can play host to the clitic cluster:

(5) [Ne dam] **ti ga**.
Neg give.1Sg you.Dat him.Acc
'I won't give it to you.'

(Milojević Sheppard & Golden, 2000, example (11))

Unlike the particle *že* 'already' in examples (3)–(4), prototypical adverbs can be left-adjacent to the second-position clitic cluster, as shown by the following constructed example.

(6) Danes **ga je** videl.
today him.Acc is saw
'Today he saw him.'

On the basis of these examples, it seems that particles could be – by inductive reasoning – defined as in (7).

(7) **(Tentative) particle definition**
A phonologically weak adverb, which cannot host a second-position clitic.

If it turns out that Slovenian particles differ from adverbs systematically in their clausal distribution along these lines (i.e., particles never directly precede the

clitic cluster), then a case can indeed be made for their constituting a distinct syntactic category. It is worth noting that in English, words identified as particles do have a unique distribution that is tied to their being phonologically weak or clitic-like. Such English particles are words like *up*, which differ from prototypical adverbs in that they are invariably constituents of the verb phrase: *John looked the information* ***up*** or *John looked* ***up*** *the information*, but not ****Up*** *John looked the information* (Den Dikken, 1995, p. 1).

2.2 Corpus investigation

In order to test (7), I have conducted a corpus investigation that checks how Slovenian particles pattern with the Wackernagel clitic cluster. According to the constructed examples (3b) and (4b), the prediction is that particles should not be able to directly precede the cluster, as there should necessarily be a phonologically strong element like an adverb that intervenes between the particle and the cluster. The corpus used for the investigation is the deduplicated version of *Gigafida 2.0* (Krek et al., 2019).

Using CLARIN.SI's noSketch Engine concordancer,[6] I have first created a frequency list of the top 30 lemmas tagged as particles in the deduplicated *Gigafida 2.0*. This frequency list is shown in Table 1.

Many of the most frequent particles in this Table are words like *tudi* 'also', *ne* 'not', *še* 'yet', and *že* 'already', which intuitively adhere to the clitic-adjacency restriction described in the previous section. There are, however, several words, like *naj* 'should' and *morda* 'possibly', which a priori seem to show no such restriction and can host the clitic cluster; compare the constructed example in (8) with (3b) and (4b) from the previous section.

(8) Morda **ga** **je** videl.
possibly him.Acc is saw
'He has possibly seen him.'

So it seems rather that only a subset of the words tagged as particles in *Gigafida 2.0* adhere to (7).

[6] https://www.clarin.si/ske/

Table 1: The 30 most frequent particles in the *Gigafida 2.0* corpus.

Lemma	Frequency	Lemma	Frequency
tudi 'also'	8,478,793	*morda* 'possibly'	468,348
ne 'not'	6,734,994	*niti* 'not even'	458,790
še 'yet'	5,787,718	*sploh* 'really/even'	419,675
že 'already'	3,735,419	*šele* 'let alone'	388,144
le 'only'	2,165,776	*pač* 'but'	294,972
naj 'should'	1,598,002	*zgolj* 'just'	254,436
prav 'precisely'	1,208,290	*zlasti* 'especially'	242,203
sicer 'otherwise'	1,078,123	*ravno* 'just'	247,650
samo 'only'	940,204	*pravzaprav* 'actually'	198,500
predvsem 'mostly'	844,382	*vsekakor* 'most certainly'	159,819
seveda 'of course'	725,852	*no* 'well'	142,860
celo 'even'	646,664	*najbrž* 'likely'	123,679
skoraj 'almost'	528,454	*menda* 'evidently'	132,347
več 'more'	520,542	*koli* 'ever'	107,068
vsaj 'at least'	507,625	*češ* 'as if'	87,039

To see if this is indeed the case, I have also queried each particle from Table 1 with a CQL search string like the one in (9), which is concretely defined for the word form *Tudi* 'also'.

(9) [word = "Tudi" & tag = "L"][tag = "Z.*k" & word != "mi"]

This search string yields bigrams in which the first element corresponds to a capitalised word belonging to the particle class, which is specified by the MSD tag L. Capitalisation is used as a heuristic to ensure that the particle is sentence initial, in parallel to the sentence-initial particle in the examples in (3)–(4). The second element is any pronominal clitic (defined by the MSD tag Z.*k). The word *mi* is filtered out because many cases of the plural nominative pronoun *mi* 'we' are incorrectly tagged with the same MSD as the syncretic singular dative pronoun *mi* 'to me'. For instance, the form *mi* in the *Gigafida 2.0* examples in (10) is tagged in both cases as a dative clitic, concretely with the tag Zop-ed--k or the English equivalent Pp1-sd--y (Pronoun Type=personal Person=first Number=singular Case=dative Clitic=yes).

Table 2: Phonologically weak particles followed by a pronominal clitic.

Particle	Capitalised word + clitic F.	Relative F.	Lemma F.
tudi 'also'	775	91.4	8,478,793
ne 'not'	939	139.4	6,734,994
še 'yet'	679	117.3	5,787,718
že 'already'	477	127.7	3,735,419
le 'only'	41	18.9	2,165,776
prav 'precisely'	695	575.2	1,208,290
samo 'only'	177	188.3	940,204
celo 'even'	16	24.7	646,664
več 'more'	0	0.0	520,542
vsaj 'at least'	12	23.6	507,625
niti 'not even'	431	939.4	458,790
šele 'let alone'	9	23.2	388,144
pač 'but'	66	223.8	294,972
zgolj 'just'	0	0.0	254,436
ravno 'just'	363	1,465.8	247,650
zlasti 'especially'	603	2,489.6	242,203
no 'well'	17	119.0	142,860
koli 'ever'	0	0.0	107,068
češ 'as if'	0	0.0	87,039
Σ	5,300	160.9	32,949,187

(10) a. Tudi **mi** jih bomo "morali" poslušati.
also we.Nom them will.1Pl must.Pl listen.Inf
'We too will have to listen to them.'

b. Tudi **mi** ne gre več.
also I.Dat Neg goes anymore.
'I am no longer able to do it.'

But in actuality, *mi* in (10a) is a nominative pronoun, which in contrast to the dative form in (10b) is crucially not a clitic and is therefore irrelevant for the present investigation.

The results are given in Tables 2 and 3, which are organised as follows. The second column, labelled as Capitalised word + clitic F., gives the absolute fre-

quency of the bigrams retrieved with search strings like (9) (relevantly modified for each particle separately). The column Relative F. provides the per-million-token frequency of the bigram relative to the absolute frequency of each lemma, which is reported in the last column.

The division of the particles between Tables 2 and 3 intuitively corresponds to their phonological strength. All the particles in Table 2 are those that are expected to pattern with the constructed examples in (3)–(4) in disallowing direct adjacency to a second-position pronominal clitic, while the particles in Table 3 are those that are intuitively expected to pattern with regular adverbs in allowing strict adjacency with the clitic, as in (6) and (8).

Interestingly, most of the particles in Table 2 can indeed be followed by a pronominal second position clitic, in contrast to what is expected by the assumed ungrammaticality of examples (3b) and (4b). In this Table, there are 775 examples of capitalised (therefore sentence-initial) *tudi* followed by a second-position clitic; 939 examples of capitalised *ne* followed by such a clitic; 679 examples of *še*; and so on. The following is a corpus example of capitalised *že* followed by the dative feminine clitic *ji* 'to her'; recall from the reported judgement of the constructed example (3b) that this string is expected to be ungrammatical. Indeed, the attestastion of such examples seems to conform to the idea that the placement of clitics in Slovenian is not fixed in any true sense, so their occurring second-position is merely a tendency (Marušič et al., 2024), albeit a strong one.

(11) Že ga je v mislih videla, kako stopa izza
already him.Acc is in thoughts saw.F how steps from-behind
grma in ji hiti naproti.
bush.Gen and her.Dat hurries towards
'In her thoughts she already saw him step out of the bush and hurry towards her.'

There is an important difference, however, in the distribution of the particles: most of the particles in Table 3 are much more frequent in this sentence-initial position followed by a clitic in comparison to the particles of Table 2. For instance, the relative frequencies of 6 out of the 11 capitalised particle + clitic bigrams in Table 3 – that is, the bigrams containing capitalised *seveda*

Table 3: Phonologically non-weak particles followed by a pronominal clitic.

Particle	Capitalised word + clitic F.	Relative F.	Lemma F.
naj 'should'	12,076	7,556.9	1,598,002
sicer 'otherwise'	6,150	5,704.4	1,078,123
predvsem 'mostly'	4,406	5,218.0	844,382
seveda 'of course'	21,650	29,827.0	725,852
skoraj 'almost'	730	1,381.4	528,454
morda 'possibly'	18,218	38,898.4	468,348
sploh 'really/even'	2,414	5,752.1	419,675
pravzaprav 'actually'	6,320	31,838.8	198,500
vsekakor 'most certainly'	5,014	31,373.0	159,819
menda 'evidently'	3,353	25,334.9	132,347
najbrž 'likely'	3,812	30,821.7	123,679
Σ	84,143	13,404.6	6,277,181

'of course', *morda* 'maybe', *pravzaprav* 'actually', *vsekakor* 'definitely', *menda* 'maybe', and *najbrž* 'probably' – are above 10,000 tokens per million, while all of the bigrams in Table 2 – save for those with capitalised *ravno* 'just' and *zlasti* 'mostly' – have relative frequencies below 1,000 tokens, many of them even below 100 tokens per million (e.g., *tudi* 'also', *le* 'only', *celo* 'even'). Taken together, the bigrams of Table 2 have an overall relative frequency of 160.9 tokens per million while the bigrams of Table 3 have an overall relative frequency of 13,404.6 tokens per million, which is roughly 83 times as many.

Let us now see how such particles compare to some of the most frequent lexemes tagged as regular adverbs in *Gigafida 2.0*. Table 4 shows how frequently 10 such adverbs play host to a second-position pronominal clitic. The main thing to notice in this Table is that the relative frequencies practically mirror those of the "non-weak" bigrams in Table 3. Six out of 10 bigrams (those with the capitalised adverbs *lahko* 'possibly/easily', *danes* 'today', *potem* 'then', *letos* 'this year', *nato* 'then', and *včeraj* 'yesterday') show relative frequencies above 10,000 per million tokens, while the overall relative frequency – that is, 13,748.6 per million – is barely higher than that of Table 3, which is 13,404.6 tokens per million.

To test these overall differences statistically, I have used the *Calc: Corpus Calculator* tool (Cvrček, 2021), which calculates χ^2 values (among other such test statistics) for the differences in pairwise absolute frequencies across differently sized sets of words. Aside from this, *Calc* also provides for each difference a DIN value, which is an effect size metric (Fidler & Cvrček, 2015).

Let us first compare the difference between the overall frequency of the "weak" particle bigrams (5,300 tokens out of 32,949,187; relative frequency = 160.9) in Table 2 and the "strong" particle bigrams (84,142 tokens out of 6,277,181; relative frequency = 13,404.6) in Table 3 on the other. This difference has a χ^2 score of 40,6508.45 and a p value of < .00001, which is statistically significant at the .05 cut-off point.

Let us now turn to the markedly smaller difference between the overall frequency of the "strong" bigrams (84,142 tokens out of 6,277,181; relative frequency = 13,404.6) in Table 3 and the adverbs (173,823 out of 12,642,922 tokens; relative frequency = 13,748.6) in Table 4. This difference has a χ^2 score of 36.95 with a p value of < .00001, which is also statistically significant at the .05 cut-off point.

As pointed out by Fidler and Cvrček (2015), the issue here is that when comparing large sample sizes, even differences that intuitively seem very small and possibly irrelevant quickly turn out to be statistically significant, which is what seems to be happening with this second difference between the "strong" particle and adverb bigrams. This is where the DIN effect-size metric comes into play. The DIN score for the overall difference between the "weak" and "strong" particle bigrams is $\pm$97.62, while the DIN score for the difference between the "strong" particle bigrams and the adverb bigrams is $\pm$1.26.

DIN values are to be interpreted as follows.

(12) a. A value of 0: the word occurs equally often in both corpora.
b. A value of $\pm$100: the word is present only in one of the two corpora.
(Adapted from Fidler & Cvrček, 2015, p. 230)

According to (12), the first DIN score of $\pm$97.62 indicates that the "strong" particles really do show a marked preference for being adjacent to second position

Table 4: Adverbs followed by a pronominal clitic.

Adverb	Capitalised word + clitic F.	Relative F.	Lemma F.
lahko 'easily/possibly'	24,988	7,163.8	3,488,074
tako 'thus'	56,520	17,182.5	3,289,402
vedno 'always'	7,722	7,182.4	1,075,127
dobro 'well'	6,186	6,824.8	906,394
danes 'today'	19,339	21,365.6	905,146
potem 'then'	22,895	28,107.2	814,559
skupaj 'together'	1,954	3,179.6	614,538
letos 'this year'	10,259	16,907.8	606,763
nato 'then'	18,328	31,121.5	588,918
včeraj 'yesterday'	5,632	15,909.6	354,001
Σ	173,823	13,748.6	12,642,922

clitics in comparison to the "weak" particles (even though such weak particles are attested in such positions), while the DIN score of ± 1.26 indicates that the "strong" particles are just as likely to directly precede second position clitics as the regular adverbs.

To my mind, there are two ways to interpret this in light of the present paper's main claim. One interpretation takes these differences in effect sizes seriously, whereby it is possible to claim something like the following: while certain Slovenian particles, namely those of Table 2, indeed differ from the adverbs distributionally, there exists a subset of particles, namely those of Table 3, that does not differ from them at all. According to this point of view, the particle category syntactically overlaps with the adverbs. But this of course raises the question why not consider this overlapping part as simply belonging to regular old adverbs – one answer is perhaps due to non-syntactic but interpretative factors, but see Section 3.2 for an argument against this as well.

The other way of looking at this is that the corpus data just show a tendency for the "weak" particles to avoid positions where they host second position clitics, but that this is not a hard and fast rule – if it were, then no such "weak" particles would be attested in this position, contrary to fact. According to this interpretation, then, particles are not distributionally different from the adverbs at all,

given that both can technically precede the clitics, albeit at different frequencies.

No matter which interpretation ultimately turns out to be correct (the answer might boil down to one's attitude towards the interpretation of word frequency in corpora, and how grammaticality is related to attestation, a non-trivial issue), neither seems to serve as a strong argument for the treatment of particles as a syntactically coherent category that is appreciably different from the adverbs.

3 OTHER DIAGNOSTICS FOR CATEGORY MEMBERSHIP

3.1 Wh-question formation

One classic (and very informal) argument in favour of treating particles as belonging to a different category from adverbs is related to the formation of *wh*-questions (Toporišič, 2000). Notice that adverbs can freely undergo *wh*-question formation, as in (13), where *kdaj* 'when' in the *wh*-question (13b) "substitutes" the adverb *včeraj* 'yesterday' in (13a). Particles, by contrast, are unable to undergo *wh*-question formation; descriptive Slovenian grammarians also typically tie this seemingly syntactic constraint to the idea that particles differ from adverbs semantically, in which sense they are closer to functional rather than lexical words (Jakop, 2001) or perhaps they are interpretatively something else entirely (see Section 3.2).

(13) a. **Včeraj** smo šli domov.
yesterday Aux.1Pl went home
'We went home yesterday.'

b. **Kdaj** smo šli domov?
when Aux.1Pl went home
'When did we go home?'

The *wh*-question test is used to identity sentence elements. The idea goes that since particles cannot be targeted by *wh*-question formation, they do not constitute a sentence element in contrast to adverbs, where for instance *Včeraj* in (13) would be analysed as an adjunct. Recall from the beginning of Section 2.1 that *Slovenska slovnica* considers particles not to be "constituents of clausal

structure" (Toporišič, 2000, p. 445). What seems to be meant by this is that particles are not sentence elements like adjuncts.

But relying on sentence elements is inappropriate for determining category membership in any case. Notice that verbs are also strictly speaking unable to undergo *wh*-question formation (yet they are traditionally considered to constitute their own syntactic category). What is possible with verbs is to form a *wh*-question that substitutes both the lexical verb (and possible verbal auxiliaries) and its possible object complement (14), but there is no operation of *wh*-question formation that would target just the lexical verb (to the exclusion of everything else), in parallel to the *wh*-question that substitutes just the adverb in (13). Notice also that with forming *wh*-questions out of verb phrases there is no 1-to-1 substitution, but the supporting verb *delati* has to be invoked aside from the *wh*-word *kaj* 'what', which is quite unlike the simple *wh*-substitution with the adverb in (13).

(14) a. Moj prijatelj **igra računalniško igro**.
my friend plays computer.Adj.Acc game.Acc
'My friend is playing a video game.'

b. **Kaj dela** moj prijatelj?
what.Acc does my friend
'What is my friend doing?'

In terms of sentence elements, example (14a) would be parsed as in (15).

(15) Moj prijatelj (Subject) igra (Predicator) računalniško igro. (Direct object)

But the impossibility to form a *wh*-question out of just the verb in (14) shows that this test is not sensitive to this kind of division;[7] rather, *wh*-question formation is sensitive to the phrase structure in (16), as it can only target the entire VP

[7] Determining what is and is not a sentence element is also quite arbitrary and depends on the particular grammatical tradition of language analysis – for instance, in English, words like *possibly* (a close equivalent for Slovenian *morda*) are in the descriptive *Comprehensive Grammar of the English Language* (Quirk et al., 1985) indeed treated as sentence elements, namely as subjuncts, which are a subtype of adverbials along with adjuncts and disjuncts. Furthermore, sentence elements are simply metalinguistic descriptors for common syntactic patterns of argument structure

igra računalniško igro or the NP *računalniško igro* 'computer game' contained within it, but not the V *igra* 'plays' itself, which is a syntactic head rather than a phrase.[8]

(16) [$_{vP}$ [$_{NP}$ Moj prijatelj] [$_{VP}$ igra [$_{NP}$ računalniško igro]]]

Furthermore, it is not strictly speaking true that it is impossible to form questions that output particles. Such questions just are not *wh*-ones, as shown by the pair in (17).

(17) a. Ali je on to naredil?
Q is he this.Acc made
'Has he done this?

b. Morda.
possibly
'Possibly.'

Consequently, the fact that particles cannot undergo *wh*-question formation is thus not a syntactic constraint, but is rather the result of a simple lexical gap according to which no *wh*-word exists in the Slovenian lexicon (or perhaps universally) that is interpretatively akin to words identified as the particles. Relying on such lexical gaps is not, by itself, a sufficient criterion for determining membership in an ultimately *syntactic* category; indeed, apart from verbs, other categories such as conjunctions and interjections also do not have *wh*-word equivalents, yet they are typically considered to be members of their own category all the same (precisely because of their syntactic/distributional characteristics, which distinguish them from one another). But, as was shown in Section 2.2, syntactically particles can occupy the same clausal positions as adverbs.

3.2 The semantics of particles and adverbs

Krvina and Žele (2018) distinguish particles from adverbs on semantic grounds, claiming that particles are neither lexical nor functional words. This idea of

realisation rather than real linguistic features, so they are only indirectly related to syntactic categories.

[8] In (16), *v*P is the phrase in which the external argument – that is, the NP *moj prijatelj* – is introduced; see (Larson, 2014).

particles supposedly going beyond the lexical–functional divide is tied to the claim by Žele (2014, pp. 9–10) that particles are not part of (the logical content of) the proposition but that they merely modify it.[9]

Let us return to the particles in Table 1. I would here like to propose that this Table actually contains 2 groups of words from the perspective of semantics. In fact, it seems that the semantic division between the two groups that I will propose shortly more or less aligns with the division between Tables 2 (the phonologically "weak" particles) and 3 (the "strong" particles). Note that a similar division is proposed by Žele (2014), who claims that the two main groups of particles in Slovenian correspond to discourse particles (Table 2) and modal particles (Table 3). Contra Žele (2014), however, I will now claim that these two groups are substantially different from each other semantically, and cannot be subsumed under her umbrella category of "propositional modifiers" (*ibid.*, p. 10).

In Table 2, there are words like *tudi* 'also', *že* 'already', and *samo* 'only'. The main semantic characteristic of these words is that they preserve the truth conditions of the proposition they modify – according to Grosz (2020), such particles "simply constrain the context in which an utterance is felicitous by acting as truth-conditionally vacuous presupposition triggers" (see also Kaufmann, 2010). For instance, the presence of *tudi* in example (18) (taken again from *Gigafida 2.0*) does not change the truth conditional meaning of the prejacent proposition – that is, regardless whether the particle is present or not, the example conveys that what one will be able to see is a set of many colour slides; the particle merely lexicalises the presupposition that other things aside from the slides will be on display as well. These words, then, are presuppositional rather than propositional modifiers.

(18) Na ogled bodo **tudi** številni barvni diapozitivi.
on display will-be.3Pl also many colour.Adj.Nom.Pl slides.Nom
'One will also be able to see many colour slides.'

[9]This claim is made on functionalist rather than formalist grounds, and crucially for Žele (2014), only sentence elements make up the propositional content of an utterance. But the classification of particles as not sentence elements is made on shaky ground in any case; see the criticism of the *wh*-question test in the previous Section as well as Footnote 7.

In Table 3, there are words like *morda* 'maybe', *seveda*, *naj* 'should', and *najbrž* 'possibly'. These words are semantically completely different from the presuppositional modifiers, as they do not preserve the truth conditions of the prejacent proposition. They correspond to modal expressions and as such they shift the interpretation of the prejacent proposition into the realm of possibility and necessity, so that the sentence modified this way no longer speaks of what holds in the actual world (von Fintel, 2006). Unlike the presuppositional modifiers, these expressions have an evaluative component to them – a word like *morda* 'possibly' is an epistemic modal (Gomboc Čeh, 2021, pp. 27–28), which means that the possibility expressed is tied to the speaker's world knowledge. What is then conveyed by *morda* 'possibly' is not only that the proposition corresponds to a possibility, but that the speaker is uncertain about it as well.

For the particles in this second group, it is quite unclear to me why they should be distinguished from prototypical adverbs at all, save of course from possible pretheoretically understood differences in degrees of abstraction. One could speculate, following Jakop (2001), that the particles correspond to functional words whereas adverbs are lexical words. But it is not the case that all adverbs semantically correspond to lexical words, which means that a lexeme being a lexical word cannot be a necessary condition for its membership in the adverb category either.

Adverbs like *takrat* 'at that time', *potem* 'then' are not referring expressions that pick out and crucially provide encyclopedic labels for events or individuals, which is how lexical words are usually conceptualised with regard to their semantics (Cann, 1999), but play a wholly functional role; *takrat* 'at that time', for instance, is a definiteness operator that picks out a temporal interval or point before the time of utterance, so it is similar to the article *the* in English, a functional word that picks out a unique individual in the context. Similarly, an adverb of frequency like *občasno* 'sometimes' semantically corresponds to a quantifier over temporal intervals/points (Lepore & Ludwig, 2007), just like *najbrž* 'probably' corresponds to a quantifier over possibilities (or possible worlds). It seems, then, that a word like *morda* 'possibly' is considered to be a particle rather than an adverb just because modality is conceptually more abstract than temporality. Apart from that, however, a particle like *morda* 'pos-

sibly' and an adverb like *občasno* 'sometimes' operate on the proposition in a completely parallel way.

In sum, only a subset of the lexemes understood as particles in contemporary Slovenian grammar substantially differ from adverbs in interpretation, this being the presuppositional modifiers of Table 2. But not only is the second group – that of Table 3 – similar to adverbs in its clausal distribution (see Section 2.2), we have just shown that it is also similar to them semantically in its altering of the truth conditions of the prejacent proposition. The particle category thus does not seem to be robustly defined on semantic grounds either.

4 DISCUSSION AND CONCLUSION

The main problem that goes beyond the putative particle–adverb distinction (but of which this distinction is a symptom) lies in what mainstream Slovenian grammarians like Toporišič (2000) consider as defining criteria for determining category membership. Rooted in general functionalist approaches to linguistic description rather than formal ones, such grammarians use mainly interpretative criteria for distinguishing syntactic categories, so for instance nouns are defined as those expressions that typically refer to persons, animals, and inanimate things,[10] adjectives are those linguistic expressions that head phrases which semantically characterise nouns, and so on.

There are two massive drawbacks of relying on interpretative criteria for defining category membership, however, especially if interpretation is the only criterion that is really used, as is the case of particles and adverbs in *Slovenska slovnica* (Toporišič, 2000). The first is that lexemes that are syntactically wholly different from one another quickly end up getting lumped together under the same category because of similar semantics. Let us for instance consider the modals *lahko* 'can' and *treba* 'must', which denote logical possibility and necessity, respectively (see Lenardič & Fišer, 2021, for a discussion). In the *Gigafida*

[10] The Universal Dependencies provides precisely such a semantic definition of nouns:

> Nouns are a part of speech typically denoting a person, place, thing, animal or idea.

(https://universaldependencies.org/u/pos/NOUN.html).

2.0 corpus, whose tagset follows the *MULTEXT-East* specifications (Erjavec, 2017), both words are consistently tagged as adverbs. Similarly, in the Slovenian part of the annotated *ParlaMint 4.0* corpus (Erjavec et al., 2023), whose morphosyntactic tags follow a different tagset – that of the Universal Dependencies framework (Dobrovoljc et al., 2023) –, both words are also in all cases defined as adverbs.[11]

However, *treba* in no way syntactically behaves like *lahko*; while the latter is syntactically indeed an adverb, the former behaves like a predicative adjective that, on the one hand, complements a copular verb (Rothstein, 1999), and on the other, combines in standard Slovenian with an infinitival verbal phrase (and rarely with a finite clause according to Uhlik, 2016, p. 55). (Note that non-finite complementation of adjectives is cross-linguistically common; see Rickman & Rudanko, 2018, for English examples. Additionally, such adjective+non-finite-verb combinations typically have a modal, evaluative, component to them, just like *treba*.) Since it is *treba*, a predicative adjective, that syntactically governs the type of the verbal phrase it combines with (i.e., the infinitival *povedati* in example (19a)), its omission from the sentence results in ungrammaticality, while *lahko*, being an adverb, is always syntactically omissible (though of course dropping it changes the meaning of the sentence). This is shown by the pair of examples in (19) from the Slovenian subset of the *ParlaMint 4.0* corpus, where the non-optionality of *treba* is indicated by *(.

(19) a. Hkrati pa je *(treba) povedati, da se študentje
simultaneously but is must tell.Inf that Refl students
vedno vključujejo ...
always join-in.3Pl
'But at the same time it *(must) be said that students always join in...'

b. Tako da je (lahko) študent tudi neodvisen.
thus that is possibly student also independent
'Thus the student is also (possibly) independent.'

[11] The Slovenian Universal Dependencies do recognize the particle class, but consider it to be a legacy category inherited from the *JOS* schema (Erjavec et al., 2010); see here https://universaldependencies.org/sl/pos/PART.html.

The second problem is that using semantics as a criterion for determining syntactic category membership can quickly result in fuzzily demarcated groups of words which contain members that are so to speak neither fish nor fowl interpretatively. As previously mentioned, Žele (2014) in her *Dictionary of Slovenian Particles* (see also Žele, 2015) defines two groups of particles in terms of interpretation – the modal particles, which modify the proposition and crucially for Žele (2014) have an evaluative component to them (e.g., *morda* 'possibly' conveys the lack of speaker's truth commitment; *skoraj* 'almost' is an approximator and thereby hedges the statement insofar as it denies a possible universal reading of the thing it modifies); and the discourse particles, which are used to organise the discourse but do not modify the prejacent presupposition and lack the evaluative component of the modals or approximators.

However, there is (at least) one word belonging to the group of phonologically weak adverbs (Table 2) which has interpretative properties of both groups – this is the negator *ne* 'not'. Like the discourse particles and unlike the modals, it has no inherent evaluative component to it (its use simply entails that a proposition is not the case); like the modals but unlike the discourse particles, it is a propositional rather than a presuppositional modifier (compare what the use of *ne* 'not' does to the truth conditions with what *tudi* 'also' does to them; i.e., only the latter preserves them). Thus, divvying up groups of words on the basis of informally defined semantic categories immediately runs into the problem of categorical fuzziness.

In light of this, I end this paper with a very brief proposal that is aimed both at descriptive Slovenian grammarians and at the developers of Slovenian corpora, who have to rely on some kind of theoretical basis when annotating the corpus data for parts of speech or even syntactic dependencies. The proposal is simply that syntactic category membership should first and foremost be defined on syntactic rather on semantic grounds (and never exclusively on the basis of semantics).

Adverbs could in this sense be defined as follows:

(20) **Adverb – a syntactic definition**
A word that heads a phrase which does not syntactically interact (e.g.,

by undergoing syntactic agreement or syntagmatically determining categorical selection) with the rest of the clausal structure in any way.

Virtually all other categories, save from possibly interjections, do interact with the rest of the clausal structure, by which I mean that they take part in observable and/or inferable syntactic dependencies. For instance, nouns head phrases which occupy clausal positions in which they receive case and, if in nominative, agree with the verb phrase. Lexical verbs determine transitivity, show Voice contrasts (active vs. passive vs. perhaps middle), agree with the noun phrase in nominative, etc. Auxiliary verbs need to be complemented by a participle. Attributive adjectives undergo syntactic agreement with the headword of the noun phrase in which they are embedded. Predicative adjectives like *treba* in (19a) determine the type of clause that syntactically complements them. Prepositions require nominal complementation and also determine the case value of the complement. And so on.

Of course, the way adverbs are distributed in a clause is not completely unconstrained, but the limited distribution invariably seems to be semantic rather than syntactic in nature (Ernst, 2007), stemming either from the semantics of the adverbs themselves (e.g., manner adverbs modify the lower parts of the clause in comparison to temporal adverbs or modals, so they tend to occur lower in the clause as well) or from the semantics of the main clause predicate (e.g., *to #deliberately contain something*). Syntax-wise adverbs can of course be constituents of different phrases, like the verbal or adjectival phrase (and possibly many others, see Cinque, 2004), but they never seem to establish any kind of *relational*(="syntagmatic") dependency with any other phrase in the clause.

A reviewer points out that *preveč* 'too much' in a structure like (21a) is involved in an observable dependency – namely, it governs the fact that the nominal *denarja* 'money' appears in genitive case. Traditionally (and in Slovenian corpora), *preveč* in such structures is indeed analysed as an adverb. According to (20), however, it is not, which is correct in my opinion. Syntactically, *preveč* 'too much' completely mirrors the syntactic behaviour of numerals assigning the so-called genitive of quantification (see also Stegovec, 2022), which a prototypical adverb does not do. For another structural parallel with the nu-

meral, notice that when *preveč* occurs in subject position, as in (21a), clausal agreement fails to take place, resulting in the verbal phrase appearing with default agreement features (third-person singular auxiliary and neuter on the main verb in periphrastic tense constructions). This to me implies that such uses of words like *preveč* should be analysed – together with numerals – as their own separate part of speech, possibly as quantifiers/determiners, rather than as belonging to a category like adverbs from whose prototypical members they differ so substantially in terms of grammar.

(21) a. Preveč denarja je bilo v banki.
too-much money.Gen Aux.3Sg was.N in bank
'too much money'

b. Pet ljudi je šlo domov.
five people.Gen Aux.3sg went.N home
'Five people went home.'

The fact that there is no reference to semantics in (20) contrasts with the definition of adverbs in, for instance, the Universal Dependencies (De Marneffe et al., 2021) framework. Like *Slovenska slovnica* (Toporišič, 2000), the Universal Dependencies guidelines rely mainly on semantic criteria for defining syntactic categories – notice the reference to semantic notions such as 'place', 'direction', and 'manner' in the following definition:

> Adverbs are words that typically modify verbs for such categories as time, place, direction or manner. They may also modify adjectives and other adverbs, as in *very briefly* or *arguably wrong*.
>
> (https://universaldependencies.org/u/pos/ADV.html)

Note also the use of *typically* in this definition. I believe that syntactic categories should be defined more robustly than this, and (undefined) exceptions should be avoided.

If a definition like (20) is adopted, the particle class could be retained to refer simply to the subset of those adverbs whose clausal distribution is limited due to phonological considerations, but not as its own category independent of adverbs – this would then parallel the way in which nominal clitics are already

treated as subsets of pronouns in the MULTEXT-East specifications (Erjavec, 2017). In my mind, relying on such syntactic criteria rather than interpretative ones would also simplify any potential work on manual annotation, since annotators would thereby be less likely to run into the aforementioned neither-fish-nor-fowl problem, which quickly arises whenever different parts of speech are stipulated (due to apparently different semantics) for words that in actuality are syntactically indistinguishable, as is the case of the current understanding of the division between particles and adverbs.

5 ACKNOWLEDGMENTS

Work described in this paper has been carried out in the context of the research project *Z6-4616: Slovenian Verbal Valency: Syntax, Semantics, and Use*, as well as under the auspices of the national research programme *P6-0436: Digital Humanities: Resources, Tools and Methods* and *DARIAH-SI: Slovenian Digital Research Infrastructure for the Arts and Humanities*.

REFERENCES

Anderson, S. R. (1993). Wackernagel's revenge: Clitics, morphology, and the syntax of second position. *Language*, *69*(1), 68–98. https://doi.org/10.2307/416416

Cann, R. (1999). Functional versus lexical: A cognitive dichotomy. In *The nature and function of syntactic categories* (Vol. 32, pp. 37–78). Brill.

Cinque, G. (2004). Issues in adverbial syntax: Taking up the gauntlet-adverbs across frameworks. *Lingua*, *114*(6), 683-710. https://doi.org/10.1016/S0024-3841(03)00048-2

Cvrček, V. (2021). *Calc 1.03: Corpus calculator*. Czech National Corpus. https://www.korpus.cz/calc/

De Marneffe, M.-C., Manning, C. D., Nivre, J., & Zeman, D. (2021). Universal dependencies. *Computational linguistics*, *47*(2), 255–308. https://doi.org/10.1162/coli_a_00402

Den Dikken, M. (1995). *Particles: On the syntax of verb-particle, triadic, and causative constructions*. Oxford University Press.

Dobrovoljc, K., Terčon, L., & Ljubešić, N. (2023). Universal dependencies za slovenščino: nove smernice, ročno označeni podatki in razčlenjevalni model. *Slovenščina 2.0: empirične, aplikativne in interdisciplinarne raziskave*, *11*(1), 218–246. https://doi.org/10.4312/slo2.0.2023.1.218-246

Erjavec, T. (2017). MULTEXT-East. *Handbook of linguistic annotation*, 441–462. https://doi.org/10.1007/978-94-024-0881-2_17

Erjavec, T., Fiser, D., Krek, S., & Ledinek, N. (2010). The JOS Linguistically Tagged Corpus of Slovene. In *LREC 2010 proceedings.* http://www.lrec-conf.org/proceedings/lrec2010/pdf/139_Paper.pdf

Erjavec, T., Kopp, M., Ogrodniczuk, M., Osenova, P., Agerri, R., Agirrezabal, M., … Fišer, D. (2023). *Linguistically annotated multilingual comparable corpora of parliamentary debates ParlaMint.ana 4.0*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1860

Ernst, T. (2007). On the role of semantics in a theory of adverb syntax. *Lingua*, *117*(6), 1008-1033. https://doi.org/10.1016/j.lingua.2005.03.015

Fidler, M., & Cvrček, V. (2015). A data-driven analysis of reader viewpoints: Reconstructing the historical reader using keyword analysis. *Journal of Slavic linguistics*, 197–239. https://www.jstor.org/stable/24602151

Fišer, D., Ljubešić, N., & Erjavec, T. (2020). The janes project: language resources and tools for Slovene user generated content. *Language resources and evaluation*, *54*(1), 223–246. https://doi.org/10.1007/s10579-018-9425-z

Gomboc Čeh, K. (2021). Izražanje naklonskih pomenov z glagolom imeti. In D. Zuljan Kumar & H. Dobrovoljc (Eds.), *Škrabčevi dnevi 12. zbornik prispevkov s simpozija 2021* (pp. 23–33). https://www.dlib.si/details/URN:NBN:SI:DOC-SFGF9P4O

Grosz, P. G. (2020). Discourse particles. In *The wiley blackwell companion to semantics.* John Wiley & Sons. https://doi.org/10.1002/9781118788516.sem047

Ilc, G. (2008). O zanikanju in nikalnici v slovenščini. *Jezik in slovstvo*, *53*(2), 65–79. https://www.dlib.si/details/URN:NBN:SI:doc-CVP39DNK

Jakop, N. (2001). Funkcijska delitev členkov: značilnosti naklonskih členkov. *Jezik in slovstvo*, *46*(7–8), 305–316. http://www.dlib.si/details/URN:NBN:SI:DOC-SI4C339V

Kaufmann, S. (2010). *Strong and weak presupposition: German 'ja' under quantifiers.* (Manuscript)

Krek, S., Erjavec, T., Repar, A., Čibej, J., Arhar Holdt, Š., Gantar, P., … Logar, N. (2019). *Corpus of written standard slovene gigafida 2.0*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1320

Krvina, D., & Žele, A. (2018). O členkih, zlasti o njihovih razločevalnih lastnostih: poudarjen slovarski vidik. *Jezik in slovstvo*, *63*(1), 39–64. https://journals.uni-lj.si/jezikinslovstvo/article/view/17767

Larson, R. K. (2014). *On shell structure.* Routledge.

Lenardič, J., & Fišer, D. (2021). Hedging modal adverbs in slovenian academic discourse. *Slovenščina 2.0: empirične, aplikativne in interdisciplinarne raziskave*, *9*(1), 145–

180. https://doi.org/10.4312/slo2.0.2021.1.145-180

Lepore, E., & Ludwig, K. (2007). Temporal adverbials and quantifiers. In *Donald Davidson's truth-theoretic semantics.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199290932.003.0010

Marušič, F. (2008). Slovenian clitics have no unique syntactic position. In *Formal approaches to Slavic linguistics* (Vol. 16, pp. 266–281). Retrieved 20 May 2024, from https://www2.ung.si/~fmarusic/pub/marusic_2007_clitics_fasl.pdf

Marušič, F., Mišmaš, P., & Žaucer, R. (2024). Placement and ordering of the (en)clitics. In D. Šipka & W. Browne (Eds.), *The Cambridge handbook of Slavic linguistics* (p. 365–384). Cambridge University Press. https://doi.org/10.1017/9781108973021.018

Marušič, F., & Žaucer, R. (2010). *An argument against syntactic positioning of Slovenian 2nd position clitics.* https://www2.ung.si/~fmarusic/pub/marusic&zaucer_2010_clitics_third_position.pdf

Milojević Sheppard, M., & Golden, M. (2000). Imperatives, negation and clitics in slovene. In *Razprave XVII* (pp. 93–109). Slovenska akademija znanosti in umetnosti, Razred za filološke in literarne vede.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. Longman.

Rickman, P., & Rudanko, J. (2018). *Corpus-based studies on non-finite complements in recent English*. Springer.

Rothstein, S. (1999). Fine-grained structure in the eventuality domain: The semantics of predicative adjective phrases and be. *Natural language semantics*, *7*(4), 347–420. https://doi.org/10.1023/A:1008397810024

Smolej, M. (2004). Členki kot besedilni povezovalci. *Jezik in slovstvo*, *49*(5), 45–57. https://www.dlib.si/details/URN:NBN:SI:DOC-25RXY8HV

Stegovec, A. (2022). Number in Slovenian. In P. Acquaviva & M. Daniel (Eds.), *Number in the world's languages – a comparative handbook* (pp. 187–260). De Gruyter Mouton. https://doi.org/10.1515/9783110622713-008

Toporišič, J. (2000). *Slovenska slovnica*. Obzorja.

Uhlik, M. (2016). Nekatere značilnosti izražanja nujnosti oz. obveznosti v slovenščini in ruščini. *Jezikoslovni zapiski*, *22*(2), 45–59. https://doi.org/10.3986/JZ.22.2.6973

von Fintel, K. (2006). Modality and language. In D. Borchert (Ed.), *Encyclopedia of philosophy* (pp. 20–27). Macmillan Reference.

Žagar, A., Kavaš, M., Robnik-Šikonja, M., Erjavec, T., Fišer, D., Ljubešić, N., Ferme, M., Borovič,M., Bošković, B., Ojsteršek, M., & Horvat, G. (2022). *Corpus of academic slovene KAS 2.0*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1448

Žele, A. (2014). *Slovar slovenskih členkov*. Založba ZRC.

Žele, A. (2015). *Dictionary of Slovenian particles*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1128

SLOVENSKI ČLENEK NI SAMOSTOJNA BESEDNA VRSTA

V prispevku predložimo, da se členki premalo razlikujejo v skladenjskem in semantičnem smislu od prislovov, da bi jih v nasprotju z obstoječimi pristopi v slovenskem opisnem jezikoslovju lahko razčlenjevali kot samostojno besedno vrsto. V prvem delu prispevka predstavimo korpusno raziskavo, s katero pokažemo, da slovenski členki niso robustno opredeljena skladenjska kategorija oz. da so skladenjsko prekrivni z običajnimi prislovi. V članku tudi predložimo, da tvorba vprašalnih stavkov, ki jo sicer uporabljamo za identifikacijo stavčnih členov, konceptualno in empirično ne služi kot ustrezen preizkus za določanje besednih vrst. Nadalje trdimo, da so splošno sprejeti pomenski kriterij tudi neustrezni za razločevanje med členki in prislovi, saj jih pesti vrsta konceptualnih problemov, razloček pa niti ni ustrezno empirično motiviran.

Keywords: členek, prislov, korpusno jezikoslovje, teoretsko jezikoslovje, skladnja