

Tanulmány

Károly Bibok & Martina Katalin Szabó

A corpus analysis of negative emotive words in Hungarian (with special focus on their discursive functions)¹

Abstract

The present paper proposes a detailed scheme of the meaning structure of the semantic and pragmatic functions of two Hungarian negative emotive words: *durva* ‘rough’ and its adverbial derivative *durván* ‘roughly’ occurring within the three corpora which represent three different text domains. Then it presents the results of testing this scheme by annotating those corpora. Besides unproblematic cases, the paper also highlights ambiguous occurrences concerning the possible semantic and pragmatic values of *durva* ‘rough’ and *durván* ‘roughly’ as sentiment words and the differentiation between the discourse marker function and the sentiment one.

Keywords: negative emotive word, sentiment values, intensifier, discourse marker, corpus analysis

1 Aims of the current study

In this paper, we present a pilot study whose aims are twofold. First, we want to systematize various semantic and pragmatic functions of two Hungarian negative emotive words (henceforth: NEWs²), which are words with a prior lexical-semantic content associated with negative emotion. These words are the adjective *durva* ‘rough’ and its adverbial derivative *durván* ‘roughly’.³ They were analyzed through a thorough examination of the data in three corpora, representing three different text domains:

¹ We owe many thanks to the anonymous reviewer for their invaluable comments. We would also like to express our gratitude to Thomas A. Williams for the proofreading.

² The term *negative emotive word* was coined on the basis of Jing-Schmidt (2007) in Szabó’s work (see, e.g., Szabó 2018), but it was extended to other occurrences than the case of emotive intensification.

³ As to the English equivalent(s) of *durva* and *durván*, we provide *rough(ly)* in single quotation marks, which seems to correspond to the negative literal sense of the Hungarian words under discussion. However, as the subsequent text attests, there are several other equivalents even in connection with the main (group of) meanings. Here, we wish to draw the reader’s attention to only one of several possible translations, *rude(ly)*, which is crucial from the perspective of negative emotive meaning.

- (i) formal news texts published on a web portal between 2002 and 2018 (Szabó et al. 2023: 1,167,398 news items, 78,578,470 tokens without punctuation),⁴
- (ii) informal texts of tweets (Szabó & Guba 2022: 40,880 tweets, 730,337 tokens without punctuation), and
- (iii) spontaneous oral discourse (Vincze et al. 2021: HuTongue Corpus, 1,149,457 tokens without punctuation).

The following number of occurrences of *durva* and *durván* were automatically filtered out of the corpora:

- (i) News texts
1,563 occurrences of *durva* and 1,436 occurrences of *durván*: 2,999 in total
- (ii) Tweets
77 occurrences of *durva* and 26 occurrences of *durván*: 103 in total
- (iii) Oral discourse
169 occurrences of *durva* and 30 occurrences of *durván*: 199 in total

Second, we discuss the results from the manual annotation of the corpus data based on the categories we formed by systematizing various semantic and pragmatic functions of the two Hungarian NEWs with a focus on the domain dependency of NEWs across various types of texts in our corpora. At the same time, we offer recommended improvements for future work with NEWs. This will concern extending the range of NEWs to be investigated and the stock of corpora as well as identifying domain-specific properties of NEWs and comparing corpus data both qualitatively and quantitatively.

2 Novelties of the current study

We wish to highlight the following special features of our investigation:

- (i) We examine a wide spectrum of different meanings and functions of the NEWs under scrutiny, which has not yet been comprehensively discussed.
- (ii) We use a large amount of data from three electronic corpora of different text domains in our investigation.
- (iii) We offer a new systematization of these meanings and functions, thus expanding our previous findings.
- (iv) We not only discuss our results from the preliminary annotation using the categories of the novel system; we also detail our concerns on how to separate the different functions of NEWs from each other. In doing so, we found that one may heavily rely on the distinction between conceptual and procedural meanings in Relevance Theory (cf. Clark 2013) or the distinction between Sentence Grammar (SG) and Thetical Grammar (TG) in Discourse Grammar (cf. Kaltenböck et al. 2011; Heine et al. 2013).

⁴ The corpus was provided by the Institute for Political Science, Center for Social Sciences, HUN-REN, Budapest, Hungary.

3 A novel model of semantic and pragmatic occurrences of Hungarian *durva*

Considering the traditional lexicographic descriptions of *durva* (cf. Bárczi & Országh 1959–1962; Pusztai 2003), one can establish three main (groups of) meanings:

- (i) ‘rough’ in the literal sense, referring to an uneven, unsmooth surface, but, more broadly, ‘harsh’, i.e., ‘jarring to the senses’,
- (ii) ‘poorly made; unfinished; lacking certain elements’ in reference either to physical objects, descriptions and ideas, or even calculations and guesses,
- (iii) ‘rude’, meaning ‘using too much force and not enough gentleness; offending others’ sensitivities’.

Clearly, the third meaning enables us to treat the word *durva* as a NEW. The literature mainly discusses NEWs, such as *durva*, as intensifiers, without highlighting other pragmatic aspects. The present authors and their colleagues have not only re-interpreted the function of intensifiers as causing a loss in polarity (NEWs lose their negative polarity) within a framework of lexical pragmatics; we have also pointed to several other pragmatics-related functions (Szabó 2018; Szabó & Bibok 2019; Szabó & Otani 2022; Szabó et al. 2023). NEWs that express a positive evaluation (*polarity shifters* in which the negative polarity of a NEW turns into a positive one) were considered to represent a specific type of enantiosemym. In addition, NEWs may indicate the striking or strange quality of certain information or the surprise of the speaker. What is more, being markers of positive politeness, evidentiality, and attention, they are able to fulfill a function called interjection by Szabó & Otani (2022).

Based on the experiences of our current analysis, we now offer the following system of semantic and pragmatic functions that can also be mapped onto the similar categories of tags to annotate our corpora:⁵

1. ‘rough; harsh’ – *durva szivacs* ‘rough sponge’, *a sziréna durva hangja* ‘the harsh sound of the siren’
2. ‘poorly made; unfinished; lacking certain elements’ – *durva hamisítvány* ‘cheap forgery’, *durva vázlat* ‘first/rough draft’, *durva becslés* ‘rough estimate’
3. ‘rude’ in connection with an emotive-evaluative semantic content

A word with this content is called a *sentiment word* in computational linguistics. It may express the speaker’s negative or positive evaluation (with or without an emotive component) of several pieces of information indicating that the information in question is desirable or undesirable

⁵ The anonymous reviewer conceives of *durva* ‘rough’ and *durván* ‘roughly’ as words encoding less negative emotional polarity than, for example, *csúnyán* ‘uglyly, nastily’ and *rettenetesen* ‘awfully, terribly’, as well as the positive emotionality of the former appears more frequently than with words with strongly negative meanings. However, our analysis does not focus on a comparison of several words alongside the negative–positive scale, but on a wide spectrum of negative, emotional, sentiment (whether negative, positive or neutral) etc. meanings that *durva* and *durván* occur with in various contexts. Cf. also the term *polarity shifter*. Furthermore, although the reviewer admits that the authors have fulfilled their aims, set at the beginning of the paper, (s)he thinks that some stylistic considerations would have been in order. In this respect, we think it seems to be sufficient enough to refer the reader to the passages here and below where individual perspectives of speakers are highlighted.

from the viewpoint of the speaker. As for the negative and positive polarity, consider, e.g., *durva ember* ‘rude man/woman’, *durva modor* ‘rude manners’ (*durva* as a NEW), on the one hand, and, on the other, *durva motor* ‘motorcycle of high quality’, *a buli durva volt* ‘the party was terrific = very good’ in corresponding contexts (cf. “polarity shifters”). Furthermore, there is a third, non-negative, non-positive, and thus neutral, i.e., in-between, use of a NEW as a sentiment word when it indicates that certain information is surprising, unexpected, worthy of attention, etc., for the speaker (Szabó 2018).

4. Function of intensifiers (cf. “polarity losers”) – not only in expressions, such as *durva/durván jó* (lit. ‘rough(ly) good’ – ‘mercilessly good’), but also *durva fejlődés* (‘high degree of (the process of) development = ‘very fast development’), *durva tempó* (‘high degree of speed’ = ‘very high pace’)
5. Function of discourse markers (instead of Szabó and Otani’s term *interjective function* from their unpublished 2022 paper)

The present authors join the trend of using the term *discourse marker* as an umbrella term to cover both textual (discourse) functions and (inter)subjective (pragmatic) functions (cf. Furkó 2020: 2). Here it is worth noting that there is also an attempt to terminologically distinguish *pragmatic marker* and *discourse marker* (Farkas et al. 2020). Discourse markers (used here in a broad sense) may have multiple functions both in the total entirety of a functional class and as individual lexical items. For the time being, we will not deal with the questions “of whether different uses of a given marker are to be considered incidental and unrelated (maximalist approach) or motivated and related (minimalist approach) and whether there is an invariant “core meaning” of discourse markers that is context-independent and preserves some component of the lexeme’s original semantic meaning” (as formulated by Furkó 2020: 9). In the present paper, we focus on the distinction between the discourse marker use and other types of non-discourse marker use of the items in question. In addition, the term *discourse marker* is an umbrella term in another sense: it involves words from different parts of speech, mainly connectives and particles, if one considers their original grammatical/semantic functions.

One of the main consequences of the heterogeneity of discourse marker functions is that discourse markers cannot be characterized universally and that the characteristic features are not applicable to each discourse marker to the same extent (cf. Furkó 2020: 6–14). Many researchers take non-propositionality, non-truth-conditionality, as a *sine qua non* for discourse marker status. In other words, discourse markers are optional from the perspective of truth-conditional utterance meaning. However, one should take into consideration another peculiarity, following relevance theorists (Wilson & Sperber 1993): the procedural, i.e., non-conceptual, nature of their meaning imposing a constraint or instruction on the pragmatic inferential phase of utterance interpretation. While propositionality, or truth-functionality, can be tested within the scope of connectives, such as *if* and *because*, the distinction between procedural and conceptual meaning is subject to the so-called compositionality test. What is even more important here is that not only non-propositional (non-truth-conditional) and/or non-conceptual, i.e., procedural, items fall within the class of discourse markers. On the one hand, the expressions *then* and *after that* seem to be propositional; on the other, *frankly* is conceptual, although it is not truth-conditional. Furthermore, being synonymous with their non-discourse marker counterparts, *seriously* and *in other words* can also be considered conceptual.

Consequently, none of the distinctions under discussion is crucial or conclusive for the notion of discourse marker. The same seems to hold for the distinction between Thetical Grammar and Sentence Grammar (cf. conceptual theticals in Kaltenböck et al. 2011). Thus, discourse markers can be defined as a class of lexical items with a wide spectrum of discourse and intersubjective functions, as noted above when the term itself was introduced.

4 Corpus analysis

Let us continue with a discussion of the corpus analysis results. First of all, some meanings were easy to separate from others, but still others were not. The latter cases need to be investigated further. Based on our experience, we will present the unambiguous categories first and then discuss the problematic categories and our dilemmas.

4.1 Results and discussion of categories 1, 2, and 4

As a first step in the investigation, all the occurrences with *durva* or *durván* were automatically filtered out of each corpus (cf. Section 1 above). The number of occurrences of the meaning types that could easily be distinguished from other meanings and functions is presented in Table 1, where tags for the manual annotation of *durva* and *durván* in the three corpora correspond to the meaning structure scheme in Section 3:

Tag	Meaning of tag	Number of occurrences
1	‘rough; harsh’	11
2	‘poorly made; unfinished; lacking certain elements’	686
4	intensifier	1518

Table 1. Occurrences of *durva* and *durván* with tags 1, 2, and 4

As we can see, the intensifier function proves to be the most frequent of these three semantic categories. Now let us see the results broken down by the three corpora and the two words *durva* and *durván* (with the total number of occurrences from Section 1):

	Category	News	Total number of occurrences	Tweets	Total number of occurrences	Oral discourse	Total number of occurrences
<i>durva</i>	1	10	1,563	—	77	1	169
	2	33		—		—	
	4	788		10		—	
<i>durván</i>	1	—	1,436	—	26	—	30
	2	646		4		3	
	4	686		21		13	

Table 2. Occurrences of *durva* and *durván* with tags 1, 2, and 4 in the three corpora

Although the numbers of occurrences of NEWs functioning as sentiment words and discourse markers are not presented here (those cases will be discussed later), based on the data in Table 2, we already notice some important differences between the three domains under examination. Let us take a closer look at these differences along with corpus examples.

Almost half the cases were intensifiers in news texts: 1474 occurrences out of 2,999. If we break down the numbers by *durva* and *durván*, the results are the following: 788 cases out of 1,563 for *durva* and 686 cases out of 1,436 for *durván*. Thus, we can see that the proportion for the two forms, *durva* and *durván*, is quite similar. Consider a pair of corpus examples.

- (1) a. Ugyanakkor a hosszabb trend is **durva** visszaesést mutat [...]

‘At the same time, the long trend also shows a serious (lit. ‘rough’ = ‘notable, significant’) decline [...]
- b. [...] ha a termékárak nem esnek **durván** vissza a hét hátralévő részében.

‘[...] if product prices don’t fall seriously (lit. ‘roughly’ = ‘notably, significantly’) for the rest of the week.’

Then, we also realize that the intensifier function was far less frequent in the other two domains in the case of *durva*: there were ten occurrences in tweets out of a total of 77, but *durva* did not occur in this function in oral discourse. As for *durván* in tweets and oral discourse, the case is different from *durva*. In the vast majority of occurrences, *durván* is an intensifier in tweets: 21 occurrences out of 26. Cf., for example,

- (2) [...] mennyit önthettek bele a rejtélyes szerből, ami ennyire **durván** érződik a második emeleten is.

‘[...] how much of the mysterious substance must have been poured into it so that it should have such a strong (lit. ‘rough’ = ‘intense’) smell even on the second floor.’

In oral discourse, the frequency distribution of *durván* as an intensifier is similar to that of news texts: 13 cases out of a total of 30.

As for categories 1 and 2, they were generally underrepresented in our corpora, with one noteworthy exception: in the case of *durván*, category 2 was common in news texts (646 occurrences out of 1,436). Based on a manual analysis of the corpus examples, *durván* meant ‘approximately’ (< ‘poorly made; unfinished; lacking certain elements’) in these cases. Consider, e.g., (3).

- (3) Ez a megkötés **durván** másfél négyzetméterrel növeli a lakások alapterületét.

‘This restriction increases the floor space of the apartments by roughly one and a half square meters.’

4.2 On the results and dilemmas concerning categories 3 and 5

Let us move on to categories 3 and 5. The SG–TG dichotomy of Discourse Grammar will be used as a starting point for our discussion.⁶ As for the sentiment function, cases of negative sentiment (in other words, *durva* and *durván* with a negatively evaluative meaning) may be relatively easy to distinguish. Consider the following examples from news texts (4a–e) and tweets (5):

- (4) a. **durva** diákcsinny ‘rude student prank’
 b. **durva** bánásmód ‘rude treatment’
 c. **durva** kijelentés ‘rude statement’
 d. **durván** viselkedik ‘(sy) behaves rudely’
 e. **durván** beszél ‘(sy) speaks rudely’
- (5) [...] nem kell mentőt hívni. Kimentem, megnéztem, mindenki jól van, de ez **durva** volt.
 ‘[...] no need to call an ambulance. I went out and checked, everyone was fine, but it was rough (= ‘terrible/awful’).’

We claim with some degree of certainty that these examples are mostly unambiguous cases of negative sentiments both syntactically and semantically. In these cases, *durva* and *durván* are grammatically dependent and add a negative semantic component to the propositional meaning of the utterances. Thus, these cases are related to SG (cf. Kaltenböck et al. 2011; Heine et al. 2013). However, there are also some problems we inevitably encounter when we examine some examples further. To begin with, when *durva* or *durván* functions as a sentiment word, sometimes its sentiment value (whether it is negative, positive, or neutral) may be difficult to determine based on the particular context. Consider the following example from the tweets:

- (6) (About an action movie.)
 És mivel elit kommandósokat képeznek az első film második felétől, elég **durva** kiképzős jelenetek.
 ‘And since elite commandos are trained from the second half of the first movie, [there are] pretty rough (= ‘brutal (–) / impressive (+) / amazing (0)’) training scenes.’

As indicated in (6), *durva* may be interpreted three ways from the perspective of its sentiment value. First, it may denote a negative evaluation by the author, so the scenes in question are physically brutal or brutish. Second, it may hint at his or her positive evaluation. In other words, he or she finds them interesting, worth watching, and impressive. Third, it may also be interpreted as the author not wishing to express either a negative or positive evaluation but to emphasize that those scenes in the movie are powerful and amazing (without any specific polarity of the evaluation).

⁶ With reference to Section 3, we wish to remind the reader of the fact that the case is somewhat more complex in Relevance Theory because – besides the dichotomy of propositional (truth-conditional) vs. non-propositional (non-truth-conditional) meaning – one should also take into consideration another distinction, that of conceptual vs. procedural meaning.

Let us continue with the discussion of non-negative (positive and neutral) cases. First consider (7).

- (7) A hidakat majdnem mindenhol alulról mossa a Vltava, elég **durva** látvány.
 ‘The bridges are washed from below by the Vltava almost everywhere, a pretty crazy (lit. ‘rough’ = ‘amazing’) sight.’ [Tweets]

Example (7) is an unambiguous case, since being a syntactic modifier of the noun *látvány* ‘sight’, *durva* is modified by an intensifier and expresses that the author finds the situation in question amazing. The utterance does not necessarily express the author’s negative or positive evaluation. Thus, a possible interpretation is that *durva* “simply” denotes a peculiarity, i.e., a noteworthy feature, of the situation being depicted. Consequently, both syntactically and semantically, this case falls within the realm of SG.

Let us take further examples where *durva* is not a modifier; it is a predicate modified by an intensifier.

- (8) a. Annyira **durva**, amennyit beleteszel.
 ‘It’s only as crazy (lit. ‘rough’ = ‘cool’) as what you put into it.’ [Tweets]
- b. Tök **durva**. Akkor a tv2 szeret a mostani zenékben otthon lenni.
 ‘Totally crazy (lit. ‘rough’ = ‘incredible’). So tv2 likes to keep up with the latest music.’ [Tweets]
- c. Nagyon **durva**! Már több mint 200 ezren mondták el a véleményüket a nemzeti konzultációban.
 ‘So crazy (lit. ‘rough’ = ‘incredible’)! More than 200,000 people have already expressed their opinion in the national consultation.’ [Tweets]

In (8a) *durva*, which is now not a modifier but is modified by an intensifier, behaves like a sentiment word again. The difference between (7) and (8a) is that the latter seems to be definitely positive. Examples (8b) and (8c) are somewhat different grammatically, since the two occurrences of *durva* form separate utterances together with their intensifiers. However, on the basis of both the syntactic constituency and semantic meaningfulness of *durva*, these cases are also considered positive sentiment words. Our argument is even supported by the fact that it is possible to create negated replies in both cases, such as in (9) and (10):

- (9) – Tök **durva**. Akkor a tv2 szeret a mostani zenékben otthon lenni.
 ‘A: Totally crazy (= ‘incredible’). So tv2 likes to keep up with the latest music.’
 – Szerintem egyáltalán nem **durva**. Ez elvárható minden csatornától.
 ‘B: I don’t think it’s crazy (= ‘incredible’) at all. That’s what we should expect of all channels.’
- (10) – Nagyon **durva**! Már több mint 200 ezren mondták el a véleményüket a nemzeti konzultációban.
 ‘A: So crazy (= ‘incredible’)! More than 200,000 people have already expressed their opinion in the national consultation.’

- Szerintem meg nem **durva**. Mindenki szereti elmondani a véleményét.
 ‘B: I don’t think it’s crazy (= ‘incredible’). Everyone likes to have their say.’

Consider one more example in (11), similar to (8b) and (8c), if one notices that the Hungarian *de* (lit. ‘but’) can be used as an intensifier (Bárczi & Országh 1959–1962). This allows space for interpreting *durva* as an intensified sentiment word but as a neutral one that expresses an exclamation of surprise, unlike the positive evaluation above.

- (11) (Three men are having a conversation. One of them is talking about his driver’s license, and he mentions that the day he received the license coincides with another important day.)
 – Sőt elmondom nektek, hogy pont december 11-én lett meg.
 – Eskü?
 – Komolyan. December 11-dike.
 – De **durva**.
 ‘A: In fact, I’m telling you I got it exactly on December 11th.
 B: You swear?
 A: Seriously. December 11th.
 B: Frack (lit. ‘So rough’ = ‘That’s surprising’).’

Finally, there are also occurrences of *durva* or *durván* forming an utterance alone, and it tends to be interpreted as a discourse marker, not as a sentiment word. We do not have space for a detailed discussion here, but we would like to draw attention to the fact that these words not being integrated syntactically into the rest of the discourse may have several different pragmatic functions in full accordance with multifunctionality as the main characteristic feature of discourse markers (see Furkó 2020, and the literature therein). The set of relevant functions includes those of expressing support for and solidarity with the speaker in (12a) and of an attention marker in (12b) and (12c), indicating that the listener can understand and follow what has been said.

- (12) a. (Two men are talking. One of them is telling the other about his physical problems.)
 – Nem tudtam jobbra fordítani a fejemet. Az nagyon szar volt.
 – **Durva**.
 ‘A: I couldn’t turn my head to the right. It was really shit.
 B: Sorry about that (lit. ‘Rough’).’
 b. (Two men are talking. One of them is telling the other about how women influence him.)
 – Engem motiválnak a nők.
 – Helyes.
 – A visszajel, a visszajelzések nagyon motiválnak, érted.
 – Tudom. **Durva**.
 ‘A: Women motivate me.
 B: That’s good.
 A: The feedback is very motivating, you know.
 B: I know. I see (lit. ‘Rough’).’

- c. (Two men are talking. One of them does not know the meaning of the word *hedonista* ‘hedonist’, so he is asking the other, who is explaining it to him.)
- Ki a „hedonista”?
 - Az az, aki öö. Érzéki örömöket hajszojja.
 - Hát fogalmam sem volt, hogy ez ezt jelenti. A „hedonista”.
 - Habzsojja az életet.
 - Habzsojja az életet? **Durva**.
 - ‘A: What’s a “hedonist”?’
 - B: Someone who... er.... pursues sensual pleasures.
 - A: Well, I had no idea that word meant that. A “hedonist”.
 - B: He really enjoys life.
 - A: He really enjoys life? I see (lit. ‘Rough’).’

Let us take another pair of examples from tweets:⁷

- (13) a. **Durva**. Még szerencse, hogy időben észrevettétek [...] ‘Hell! / Damn! It’s lucky you noticed it in time [...]’
- b. Az egyik ügyintéző szerint csak akkor működik az ingyenes [*typo!*] közösségi média ha mellé van rendelve fizetős adatforgalom a másik szerint ha van pénz az egyenlegemen. **Durva**. Azt sem tudják mi újság.
 ‘According to one of the administrators, free social media only works if it comes with paid data traffic. According to the other, [it only works] if there’s money in my balance. Frack (= ‘Hell! / Damn!’ or ‘Incredible!’). They have no idea what’s going on.’

As the first element in the turn (the first element of the tweet in a tweet chain) and as an independent utterance, once again not integrated syntactically into the rest of the discourse, *durva* in (13a) can be interpreted as an expletive-type reaction to the information received. Thus, *durva* can be paraphrased with the following Hungarian expletive interjections, for instance, *Huh!* ‘Hah’, *Azta!* ‘Whatta!’, *Basszus!* ‘Damn!’, *Baszki!* (vulg.) ‘Fuck’. Hence, here *durva* may be considered a discourse marker and is thus related to TG because it is syntactically autonomous, and its meaning is basically procedural rather than conceptual (Kaltenböck et al. 2011; Ahn & Yap 2022). However, *durva* in (13b) may be interpreted in two different ways, which are indicated in the English translation of the example as well: as a discourse marker as in (13a) or as a negative evaluation of the previous or subsequent utterance. The second case is possible if one accepts the second utterance as being connected to the first (with the Hungarian pronominal subject *ez* ‘this’ being dropped), not as standing syntactically alone, or if the third utterance is subordinated to the second one (with the complementizer *hogy* ‘that’ being missing).

Furthermore, there are occurrences of the word in question appearing as a component of a multiword construction, but the meaning of the whole expression cannot be calculated from the meanings of its individual components. We argue that, since these collocations are non-

⁷ In (13b), besides the typo *ingeyenes* (this should read *ingyenes* ‘free’), several commas are also missing compared to the standard rules of Hungarian spelling.

compositional constructions with the function of an interjection, they are thus multiword discourse markers. Similar Hungarian collocations were found with *durva*, e.g., *azta durva* ‘wow!, hah!’ and *azta durva mindenit* ‘wow!, hah!’. There are no examples of these multiword discourse markers in our corpora, but we have found some examples on the web:

- (14) a. Azta **durva**! A kedvezményes jegyekből már csak 100 db a boltokban!⁸
 ‘Wow! There are only 100 discount tickets left in the stores!’
- b. azta **durva** mindenit... véletlenül ide tévedtem és mit láttam??? új bejegyzések és ez kirááááály!!!!⁹
 ‘Wow!... I found my way here by accident and what did I see??? new posts and it’s cool!!!!’

The multiword expressions in (14) are non-compositional, i.e., lexicalized, as reflected by the English translations. Hence, they have no conceptual meaning, only a procedural one.

One must admit that these cases are highly complex both semantically and pragmatically and that it is sometimes difficult to decide whether a particular example of *durva* or *durván* should be considered a sentiment word with a negative, positive, or neutral value or a pragmatic marker. Moreover, their consistent handling becomes challenging during a corpus annotation when the annotators need to make firm and accurate decisions based on a specific annotation scheme. The right solution might be a multiple choice of tags by annotators in such ambiguous cases, with disambiguation only as a second step provided that a context makes any disambiguation possible.

5 Conclusions

In the present paper, we have proposed a detailed scheme of the meaning structure of the semantic and pragmatic functions of the Hungarian adjective *durva* ‘rough’ and its adverbial derivative *durván* ‘roughly’ occurring within the three corpora which represent three different text domains. Then we have presented the results of testing this scheme by annotating those corpora. Besides unproblematic cases, our paper has also highlighted ambiguous occurrences concerning the possible semantic and pragmatic values of NEWs as sentiment words and the differentiation between the discourse marker function and the sentiment one. We have provided insights into dilemmas of our annotation by illuminating corpus examples. As the reader can realize, the authors themselves have conducted an initial annotation of the corpora to evaluate and refine the proposed scheme. In future phases, a broader annotation process will involve multiple annotators to ensure the reliability and validity of the categories and distinctions suggested in this study. In addition, despite the fact that we have still not revealed the frequency distributions of each and every function and meaning of *durva* and *durván* in our corpora, some features seem to be domain-dependent.

⁸ https://www.facebook.com/ultrapusztas/posts/azta-durva-a-kedvezm%C3%A9nyes-jegyekb%C5%91l-m%C3%A1r-csak-100-db-a-boltokban-haszn%C3%A1lj%C3%A1tok-ki-/2382304791997985/?locale=zh_CN – Accessed: July 24, 2024.

⁹ ildisusatrip.blogspot.com/2012/06/24-ora-alatt-4-orszag-also-nap.html – Accessed: July 24, 2024.

The main lesson to learn from this discussion is twofold. First, the meanings and functions of such lexical items as NEWs are far beyond the dictionary descriptions of lexical entries, especially as regards their behavior in various types of discourse. Second, having a detailed scheme of meanings and functions encountered in corpora, annotators of such corpora may face a series of challenges, which point to the need for flexibility in a carefully developed annotation scheme to register ambiguities even in real contexts of discourse.

Acknowledgements

The research by Károly Bibok presented here was supported by the Subprogram for Linguistic Identification of Fake News and Pseudo-scientific Views, part of the Science for the Hungarian Language National Program within the Hungarian Academy of Sciences (MTA). The research by Martina Katalin Szabó was funded by the OTKA Postdoctoral Research Grant, by the Hungarian Research Fund within the National Research, Development and Innovation Office of Hungary (NKFIH, grant number: PD 132312), and by the International Research Fellowship Program within the Japan Society for the Promotion of Science (JSPS, Postdoctoral Fellowships for Research in Japan (Standard)). This research project was partly supported by the Center for Social Sciences, a member of the Hungarian Research Network (HUN-REN) and a center of excellence of the Hungarian Academy of Sciences.

References

- Ahn, M. & Yap, F. H. (2022): On the evolution of a multifunctional discourse marker: A Discourse Grammar analysis of Korean *com*. *Journal of Pragmatics* 195, 31–47. <https://doi.org/10.1016/j.pragma.2022.04.010>
- Bárczi, G. & Országh, L. (1959–1962): *A magyar nyelv értelmező szótára* [A Defining Dictionary of the Hungarian Language]. Vols. I–VII. Budapest: Akadémia Kiadó.
- Clark, B. (2013): *Relevance Theory*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139034104>
- Farkas, J., Futó, B., Huszics, A., Kleiber, J., Dóla, M. & Alberti, G. (2020): Similarities and differences between two Hungarian particles for also: *szintén* and *is*. *Linguistics Beyond and Within* 6, 74–91. <https://doi.org/10.31743/lingbaw.11832>
- Furkó, B. P. (2020): *Discourse Marker and Beyond: Descriptive and Critical Perspectives on Discourse-Pragmatic Devices across Genres and Languages*. Cham: Palgrave Macmillan. <https://doi.org/10.1007/978-3-030-37763-2>
- Heine, B., Kaltenböck, G. & Kuteva, T. (2013): On the origin of Grammar. In: Lefebvre, C., Comrie, B. & Cohen, H. (eds.): *New Perspectives on the Origins of Language*. Amsterdam & Philadelphia: John Benjamins, 379–405. <https://doi.org/10.1075/slcs.144.15hei>
- Jing-Schmidt, Z. (2007): Negativity bias in language: A cognitive-affective model of emotive intensifiers. *Cognitive Linguistics* 18.3, 417–443. <https://doi.org/10.1515/COG.2007.023>
- Kaltenböck, G., Heine, B. & Kuteva, T. (2011): On thetical grammar. *Studies in Language* 35.4, 852–897. <https://doi.org/10.1075/sl.35.4.03kal>
- Pusztai, F. (eds.) (2003): *Magyar értelmező kéziszótár* [A Concise Dictionary of Definitions of Hungarian]. 2nd, revised edition. Budapest: Akadémiai Kiadó.

- Szabó, M. K. (2018): *A szentimentérték módosulásának a problémája a magyar nyelvű szövegek szentimentelemzésében, különös tekintettel az értékvesztésre és az értékváltásra* [Problems of Sentiment Analysis in Hungarian Texts, with Particular Regard to Polarity Loss and Polarity Shift]. PhD thesis. Szeged: University of Szeged. <https://doktori.bibl.u-szeged.hu/id/eprint/9976/1/VEGL.pdf>
- Szabó, M. K. & Bibok, K. (2019): Értékvesztésre és értékváltásra képes lexémák újabb vizsgálata [A novel analysis of the lexemes that can undergo polarity loss and polarity shift]. *Argumentum* 15, 639–649.
- Szabó, M. K. & Guba, C. (2022): *An Analysis of Negative Emotive Intensifiers in Hungarian Tweets*. Manuscript.
- Szabó, M. K. & Otani, N. (2022): Corpus analysis of Japanese and Hungarian negative emotive words from a discourse interactional perspective. In: *Proceedings of the 25th Annual Meeting of Pragmatic Society of Japan*. Kyoto University, Japan. Paper submitted.
- Szabó, M. K., Dam, B. & Vincze, V. (2023a): *On the Semantic Development of Negative Emotive Intensifiers in Hungarian News*. Manuscript.
- Szabó, M. K., Vincze, V. & Bibok, K. (2023b): “Thank you for the terrific party!” – An analysis of Hungarian negative emotive words. *Corpus Linguistics and Linguistic Theory* 19.3, 451–485. <https://doi.org/10.1515/cllt-2022-0013>
- Vincze, V., Üveges, I., Szabó, M. K. & Takács, K. (2021): A magyar beszélt és írott nyelv különböző korpuszainak morfológiai és szófaji vizsgálata [Morphological and phonological examination of various corpora of the Hungarian spoken and written language]. In: Berend, G., Gosztolya, G. & Vincze, V. (eds.): *XVII. Magyar Számítógépes Nyelvészeti Konferencia* [17th Conference on Hungarian Computational Linguistics]. Szeged: Szegedi Tudományegyetem, Informatikai Intézet, 169–182.
- Wilson, D. & Sperber, D. (1993): Linguistic form and relevance. *Lingua* 90.1–2, 1–25. [https://doi.org/10.1016/0024-3841\(93\)90058-5](https://doi.org/10.1016/0024-3841(93)90058-5)

Dr. Károly Bibok 
 University of Szeged
 Department of Russian Philology
 H-6722 Szeged
 Egyetem u. 2.
 kbibok@lit.u-szeged.hu

Dr. Martina Katalin Szabó 
 University of Szeged
 Institute of Informatics
 H-6720 Szeged
 Árpád tér 2.
 szabomartinakatalin@gmail.com