A corpus analysis of negative emotive words in Hungarian
(with a special attention to their discursive functions)

Károly Bibok (kbibok@lit.u-szeged.hu)
University of Szeged and
MTA–SZTE–DE Research Group for Theoretical Linguistics and Informatics)
&
Martina Katalin Szabó (szabomartinakatalin@gmail.com)
University of Szeged and
HUN-REN CSS-RECENS

1. Aims of the current study

In our 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: NEWs) – words having a prior lexical-semantic content associated with negative emotion –, namely, the adjective *durva* 'rough' and its adverbial derivation *durván* 'roughly'.[1] They were revealed by a thorough examination of the data of three corpora, representing three different text domains:

(i)   formal news texts published on a web portal between 2002 and 2018 (Szabó et al. 2023: 1,167,398 news, 78,578,470 tokens without punctuation),[2]

(ii)   informal texts of tweets (Szabó and Guba 2022: 40,880 tweets, 730,337 tokens without punctuation), and

(iii)   a spontaneous oral discourse (Vincze et al. 2021: HuTongue Corpus, 1,149,457 tokens without punctuation).

The following amounts of *durva* and *durván* were automatically filtered out from the corpora:

(i)   news texts
        1,563 occurrences of *durva* and 1,436 occurrences of *durván*, altogether 2,999

(ii)   tweets
        77 occurrences of *durva* and 26 occurrences of *durván*, altogether 103

(iii)   oral discourse
        169 occurrences of *durva* and 30 occurrences of *durván*, altogether 199

Second, we attempt to discuss the results of the manual annotation of corpus data based on the categories gained by the systematization of various semantic and pragmatic functions of the two Hungarian NEWs paying attention to domain dependency of NEWs across various types of texts in our corpora. At the same time, we offer some improvements for the future work with NEWs. It will concern the extension of the range of NEWs to be investigated and the stock of corpora as well as the revelation of the domain-specific properties of NEWs and the comparison of corpus data both from qualitative and quantitative perspectives.

---

[1] As to the English equivalent 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 can be several other equivalents even in connection with the main (group of) meanings. Here we want to draw the reader's attention to only one out of possible translations, namely, *rude(ly)*, which is crucial from the point of view of the negative emotive meaning.

[2] The corpus was provided to us by the Institute for Political Science of the Centre for Social Sciences.

2. Novelties of the current study

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

- We examine such a wide spectrum of different meanings and functions of the NEWs under scrutiny that has not been comprehensively discussed yet.
- To investigate them, we use a large amount of data from three electronic corpora of different text domains.
- We offer a new systematization of these meanings and functions extending our previous findings.
- We do not only discuss our results of the preliminary annotation using the categories of the novel system but also enter into details of our dilemmas concerning how to separate NEWs' different functions from each other. In doing so, 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).

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

Considering its traditional lexicographic descriptions (cf. Bárczi and Országh 1959–1962, Pusztai 2003), one can establish three main (groups of) meanings:

(i) 'rough' as literal sense referring to uneven, not smooth surface, but more widely: 'harsh', i.e., 'jarring to the senses',

(ii) 'not well made, made without much detail' referring either to physical objects, or descriptions, ideas, or even calculations and guesses,

(iii) 'rude' meant as 'using too much force and not enough gentleness offending others' sensitivity'.

Let us realize that the third meaning gives way to treat the word *durva* as a NEW. The previous literature mainly discussed NEWs, including *durva*, as intensifiers, not highlighting other pragmatic aspects. The present authors and their colleagues did not only re-interpret the function of intensifiers as polarity losing (NEWs lose their negative polarity) in a framework of lexical pragmatics but also pointed out to several other pragmatic-related functions (Szabó 2018, Szabó and Bibok 2019, Szabó and Otani 2022, Szabó et al. 2023). NEWs expressing positive evaluation (*polarity shifters*: the negative polarity of NEWs turns into a positive one) were considered a specific type of enantiosemy. In addition, NEWs may indicate the striking, strange feature of the given 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 by Szabó and Otani (2022) as interjective.

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

1. 'rough; harsh' – *durva szivacs* 'rough sponge', *a sziréna durva hangja* 'harsh sound of the siren'

2. 'not well made' – *durva hamisítvány* 'cheap forgery', *durva vázlat* 'first draft', *durva becslés* 'rough estimate'

3. 'rude' in connection with an emotive-evaluative semantic content

A word with such a content is called *sentiment word* in the field of computational linguistics. It may express the speaker's negative or positive evaluation (with or without an emotive

component) towards several pieces of information indicating that the given information is desirable or undesirable from the viewpoint of the speaker. As for the negative and positive polarity, consider, e.g., *durva ember* 'rude (tough) 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"). What is more, there occurs a third, not negative and not positive but, so to say, a neutral, i.e. in-between, usage of a NEW as sentiment word when it indicates that the given 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* used in their 2022 unpublished paper)
The present authors join the trend that uses the term *discourse marker* as an umbrella term covering 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 discourse marker use and other types of non-discourse marker use of the given items. In addition, the term *discourse marker* is an umbrella term in another sense, namely it involves words that belong to different parts of speech, first of all, connectives and particles, if one considers their original grammatical/semantic functions.

One of the main consequences of heterogeneity of discourse markers' functions is that the discourse markers cannot be characterized in a unique way and the characteristic features are not applicable to each discourse marker to the same extent (cf. Furkó 2020: 6–14). Many researchers take the non-propositionality, or non-truth-conditionality, as a sine qua non for discourse marker status. In other words, discourse markers are optional from the point of view of truth-conditional utterance meaning. However, one should take into consideration another peculiarity, following relevance theorists (Wilson and 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 the propositionality, or truth-functionality, can be tested in 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 belong to the class of discourse markers. On the one hand, the expressions *then* and *after that* seem to be propositional, and on the other, *frankly* is conceptual, although 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 delineation of Thetical Grammar from 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 mentioned above when the term itself was introduced.

4. Corpus analysis

Let us continue with the discussion of the corpus analysis results. First of all, some meanings could be easily separated from other meanings, but some could not. The latter cases need to be further investigated. Based on our experiences, we will present the unambiguous categories first, then discuss the problematic categories and our dilemmas.

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

As a first step of the investigation, all the occurrences containing *durva* or *durván* were automatically filtered out from each corpus (cf. Section 1 above). The number of occurrences of the meaning types that could be easily 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 schema in Section 3:

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

tag	meaning of the tag	number of occurrences
1	'rough; harsh'	11
2	'not well made', 'first draft', 'rough estimate'	686
4	intensifier	1518

As we can see, the intensifier function is proved to be the most frequent out 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):

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

	category	**news**	total number of occurrences	**tweets**	total number of occurrences	**oral discourse**	total number of occurrences
durva	**1**	10	**1,563**	–	**77**	1	**169**
	2	33		–		–	
	4	788		10		–	
durván	**1**	–	**1,436**	–	**26**	–	**30**
	2	646		4		3	
	4	686		21		13	

Although the number 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 can already notice some important differences among the examined three domains. 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 broken 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 expressions, namely for *durva* and *durván*, is quite similar. Consider a pair of corpus examples.

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

4

‘At the same time, the longer trend also shows a rough (= ‘notable, significant’) decline [...]’

b.   […] ha a termékárak nem esnek **durván** vissza a hét hátralévő részében.
‘[…] if product prices don’t fall back roughly (= ‘notably, significantly’) for the rest of the week.’

Then, we can also realize that in the case of *durva* the intensifier function was far less frequent in the other two domains: in tweets it had 10 occurrences from the total of 77, and in oral discourse *durva* did not even occur in this function. 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 could be poured into it, which feels so rough (= ‘intensive’) even on the second floor.’

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

As for categories 1 and 2, they were generally underrepresented in our corpora, with one noteworthy exception: category 2 in the case of *durván* was common in news texts (646 occurrences out of 1,436). Based on the manual analysis of the corpus examples, *durván* had the meaning ‘approximately’ (< ‘not well made’) in these cases. Consider, e.g., (3).

(3)   Ez a megkötés **durván** másfél négyzetméterrel növeli a lakások alapterületét.
‘This restriction increases the floor area 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 to categories 3 and 5. The SG–TG dichotomy of Discourse Grammar will be used as a starting point of our discussion.[3] As for the sentiment function, cases of negative sentiment (in other words, *durva* and *durván* with a negatively evaluative meaning) may be relatively easily distinguished. Consider the following examples from news texts (4a–e) and tweets (5):

(4)   a.   **durva** diákcsíny ‘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 tough (= ‘terrible/awful’).’

---

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

We claim with a fairly safe certainty that these examples are mostly unambiguous cases of negative sentiments both from syntactic and semantic points of view. In these cases, *durva* and *durván* are grammatically dependent, and add a negative semantic component to the propositional meaning of utterances. Thus, these cases are related to SG (cf. Kaltenböck et al. 2011, Heine et al. 2013). However, there are also some problems we unavoidably encounter when we examine 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 hardly decided based on the given context. Consider the following example from 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 tough (= 'brutal (–) / impressive (+) / amazing (0)') training scenes.'

As indicated in (6), *durva* may be interpreted three ways from the viewpoint of its sentiment value. First, it may denote the negative evaluation of the author, so the given scenes are physically brutal, brutish. Second, it may hint at his or her positive evaluation. In other words, he or she finds them interesting, worth watching, impressive. Third, it may also be interpreted in a way that the author of the tweet does not want to express neither negative nor positive evaluation but to emphasize that those scenes of the movie are crushing and amazing (without any specific polarity of the evaluation).

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 almost everywhere washed from below by the Vltava, a rather tough (= '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 at stake 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 depicted situation. Consequently, from both syntactic and semantic points of view this case belongs to the realm of SG.

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

(8)   a.   Annyira **durva**, amennyit beleteszel.
                'It's only as tough (= 'cool') as you invest into it.' [tweets]

        b.   Tök **durva**. Akkor a tv2 szeret a mostani zenékben otthon lenni.
                'It's so tough (= 'incredible'). Then tv2 likes to be familiar with current music.' [tweets]

        c.   Nagyon **durva**! Már több mint 200 ezren mondták el a véleményüket a nemzeti konzultációban.
                'It's very tough (= '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 as 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 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.
         '– It's so tough (= 'incredible'). Then tv2 likes to be familiar with current music.'
         – Szerintem egyáltalán nem **durva**. Ez elvárható minden csatornától.
         '– I think it's not tough (= 'incredible') at all. This is to be expected from all channels.'

(10)   – Nagyon **durva**! Már több mint 200 ezren mondták el a véleményüket a nemzeti konzultációban.
         '– It's very tough (= 'incredible')! More than 200,000 people have already expressed their opinion in the national consultation.'
         – Szerintem meg nem **durva**. Mindenki szereti elmondani a véleményét.
         '– I think it's not tough (= '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 and Országh 1959–1962). This gives pace for interpreting *durva* as an intensified sentiment word but for a neutral one expressing an exclamation of surprise unlike the above positive evaluation.

(11)   (Three men have a conversation. One of them is telling about his driving license and he mentions that the day of obtaining the license coincides with another important day.)
         – Sőt elmondom nektek, hogy pont december 11-én lett meg.
         – Eskü?
         – Komolyan. December 11-dike.
         – De **durva**.
         '– Moreover, I'm telling you that it happened exactly on December 11th.
         – Do you swear?
         – Seriously. December 11th.
         – How tough (= 'surprising').'

Finally, there are also occurrences when *durva* or *durván* forms an utterance alone and it tends to be interpreted as a discourse marker, and not as a sentiment word. Here we do not have a room for detailed discussion, but we would like to call attention to the fact that these words not integrated syntactically into the rest of the discourse may have several different pragmatic functions in full accordance with the 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 and solidarity toward the speaker in (12a), or of an attention marker in (12b) and (12c), indicating that the listener can understand and follow what was 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**.
'– I couldn't turn my head to the right. That was really shit.
 – Sorry for you.'

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**.
'– Women motivate me.
 – That's right.
 – The feedback is very motivating, you know.
 – I know. I see.'

c. (Two men are talking. One of them does not know the meaning of the word *hedonist*, so he is asking the other, who is explaining it to him.)
– Ki a „hedonista"?
– Az az, aki öö. Érzéki örömöket hajszolja.
– Hát fogalmam sem volt, hogy ez ezt jelenti. A „hedonista".
– Habzsolja az életet.
– Habzsolja az életet? **Durva**.
'– Who is the „hedonist"?
 – The one who… er…. pursues sensual pleasures.
 – Well, I had no idea that this word meant that. The "hedonist".
 – The one who really enjoys life.
 – Enjoys life? 'I see').'

Let us take another pair of examples from tweets:[4]

(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 ingeyenes [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 paid data traffic is assigned to it, according to the other, if there is money on my balance. Tough ('Hell! / Damn!' or 'Incredible!') They have no idea what's going on.'

Durva in (13a), as the first element of the given 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, can be interpreted as an expletive-type reaction on the information received. Thus, *durva* can be paraphrased with, for instance, the following Hungarian expletive interjections: *Huh!* 'Hah', *Azta!* 'Whatta!?', *Basszus!* 'Damn!', *Baszki!* (vulg.) 'Fuck'. Hence, here *durva* may be considered a discourse marker, and this way it is related to TG; because it is syntactically autonomous and its meaning is basically procedural rather than conceptual

---

[4] In (7b), besides the typo *ingeyenes*, several commas are also missing in comparison with the standard rules of spelling.

(Kaltenböck et al. 2011, Ahn and 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 like (13a) or a negative evaluation of the previous or subsequent utterance. The second case is available if one takes the second utterance as not standing syntactically alone but it is connected to the first one (while the Hungarian pronominal subject *ez* 'this' is dropped) or the third utterance is subordinated to the second one (while the complementizer *hogy* 'that' is missing).

Furthermore, there are occurrences when the word at point appears 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-compositional constructions with a function of interjection, these expressions belong to multiword discourse markers. There could be found similar Hungarian collocations containing *durva*, e.g. *azta durva* 'wow!, hah!' and *azta durva mindenit* 'wow!, hah!'. There are no examples for these multiword discourse markers in our corpora but here we provide some examples from the web:

(14)   a.   Azta **durva**! A kedvezményes jegyekből már csak 100 db a boltokban! [5]
              'Wow! Only 100 of the discount tickets are left in the shops!'

        b.   azta **durva** mindenit... véletlenül ide tévedtem és mit láttam??? új bejegyzések és ez kirááááály!!!! [6]
              'Wow!... I wandered 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 but procedural one.

One must admit that the above cases are really very complex from both a semantic and a pragmatic point of view and that it is sometimes even difficult to decide whether the examples 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 have 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 and it is only a second step to carry out disambiguation if contexts make any disambiguation possible.

5. Conclusions

In the present paper we have proposed a detailed schema of the meaning structure of semantic and pragmatic functions the Hungarian adjective *durva* 'rough' and its adverbial derivation *durván* 'roughly' occur with in the three corpora, representing three different text domains. Then we have presented the results of testing this schema in annotation of those corpora. Besides unproblematic cases, our presentation has also highlighted ambiguous occurrences concerning the possible semantic and pragmatic values of NEWs as sentiment words and the delineation of the discourse marker function from the sentiment one. We have provided insights into our (annotators') dilemmas with the help of illuminating corpus examples. In addition, despite the fact that we have still not revealed the frequency distributions of each and every

---

[5] https://www.facebook.com/ultrapuszta/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: 24. 07. 2024.

[6] ildisusatrip.blogspot.com/2012/06/24-ora-alatt-4-orszag-elso-nap.html – accessed: 24. 07. 2024.

function and meaning of *durva* and *durván* in our corpora, some features seem to be domain-dependent.

The main lesson to learn from the above discussion is twofold. First, meanings and functions of such lexical items as NEWs are far beyond what descriptions of lexical entries contain in dictionaries especially regarding 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, what points out that a carefully developed annotation scheme should be flexible to register ambiguities existing even in real contexts of discourse.

Acknowledgments

References

Ahn, M. and 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. and 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. and 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. and Kuteva, T. (2013): On the origin of Grammar. In: Lefebvre, C., Comrie, B. and Cohen, H. (eds.): *New Perspectives on the Origins of Language*. Amsterdam and Philadelphia: John Benjamins, 379–405. https://doi.org/10.1075/slcs.144.15hei

Kaltenböck, G., Heine, B. and 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]. Budapest: Akadémiai Kiadó, 2nd, revised edition.

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.

Szabó, M. K. and 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. and Guba, C. (2022): *An Analysis of Negative Emotive Intensifiers in Hungarian Tweets.* Manuscript.

Szabó, M. K. and 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. and Vincze, V. (2023a): *On the Semantic Development of Negative Emotive Intensifiers in Hungarian News.* Manuscript.

Szabó, M. K., Vincze, V. and 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. and 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. and 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