

A corpus analysis of negative emotive words in Hungarian

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)

Tokyo University of Foreign Studies,

University of Szeged and

ELKH CSS-RECENS

Aims of the current study

- aims are twofold:
 1. to systematize various semantic and pragmatic functions of two Hungarian negative emotive words
 - words having a prior lexical-semantic content associated with negative emotion
 - the adjective *durva* ‘rough, rude’
 - its adverbial derivation *durván* ‘roughly, rudely’

Those functions 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),
- (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).

Second aim

2. to discuss the results of the manual annotation of corpus data according to the categories gained by the systematization of various semantic and pragmatic functions of the two Hungarian negative emotive words (henceforth: NEWs)

in addition, paying attention to domain dependency of NEWs across various types

at the same time, we offer some improvements for the future work with NEWs

2. Novelties of the current study

- a wide spectrum of different meanings and functions of the NEWs under scrutiny that has not been comprehensively discussed yet
- a large amount of data from three electronic corpora of different text domains
- a new systematization of these meanings and functions extending our previous findings

- results of the preliminary annotation using the categories of the novel system
- dilemmas concerning how to separate NEWs' different functions from each other
- distinction between conceptual and procedural meanings in Relevance Theory (cf. Clark 2013)
- the distinction between Sentence Grammar (SG) and Thetical Grammar (TG) in Discourse Grammar (cf. Kaltenböck et al. 2011)

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

traditional lexicographic descriptions (cf. Bárczi and Országh 1959–1962, Pusztai 2003):

- (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’

- 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
- we re-interpret the function of intensifiers as polarity losing (NEWs lose their negative polarity) in a framework of lexical pragmatics
- several other pragmatic-related functions (Szabó 2018, Szabó and Bibok 2019, Szabó and Otani 2022, Szabó et al. 2022)

- NEWs expressing positive evaluation
polarity shifters: the negative polarity of NEWs turns into a positive one → enantiosemy
- NEWs may indicate the striking, strange feature of the given information or the surprise of the speaker
- markers of positive politeness, evidentiality and attention → interjection (Szabó and Otani 2022)

- Based on the experiences of our current analysis, 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

- *sentiment word* in the field of computational linguistics
- the speaker’s negative or positive evaluation (with or without an emotive component) towards the given information indicating that the given information is desirable or undesirable from the viewpoint of the speaker
 - durva ember* ‘rude man/woman’, *durva modor* ‘rude manners’ (*durva* as a NEW)
 - durva motor* ‘motorcycle of high quality’, *a buli durva volt* ‘the party was terrific = very good’ in corresponding contexts (cf. “polarity shifters”)
- not negative and not positive → neutral usage of a NEW: the given information is surprising, unexpected, worthy of attention, etc. for the speaker (Szabó 2018)

4. function of intensifiers – *durva fejlődés* ‘very fast development’, *durva tempó* ‘very high pace’ (cf. “polarity losers”)

5. function of discourse markers (instead of Szabó and Otani's term *interjective function* used in their 2022 unpublished paper)

- the term *discourse marker*: umbrella term covering both textual (discourse) functions and (inter)subjective (pragmatic) functions (cf. Furkó 2020: 2)
- discourse markers have multiple functions both in the total entirety of a functional class and as individual lexical items

- 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) 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)
- we will focus on the distinction between discourse marker use and other types of non-discourse marker use of the given items

A consequence of heterogeneity of discourse markers' functions

- discourse markers cannot be characterized in a unique way
- characteristic features are not applicable to each discourse marker to the same extent
- non-propositionality or non-truth-conditionality as a sine qua non for discourse marker status (cf. Furkó 2020: 6)
- not one but two distinctions
 - conceptual vs. procedural meaning
 - truth-conditional vs. non-truth-conditional meaning
- none of them is crucial or conclusive for the definition of the notion of discourse marker
- delineation of Thetical Grammar from Sentence Grammar (cf. conceptual theticals in Kaltenböck et al. 2011)

Corpus analysis results

- Some meanings could be easily separated
- But some could not → need to be further considered
- Here we will discuss:
 - unambiguous categories (1, 2 and 4)
 - problematic categories (3 and 5)
 - and our dilemmas

- Method: all the occurrences were automatically filtered out, then a manual analysis was carried out
- Data:
 - (i) news texts: 1,563 occurrences of *durva*
1,436 occurrences of *durván*,
altogether 2,999 occurrences,
 - (ii) tweets: 77 + 26, altogether 103 occurrences,
 - (iii) speech: 169 + 30, altogether 199 occurrences

Results on categories 1, 2 and 4

1. ‘rough; harsh’;
2. ‘not well made’, ‘first draft’, ‘rough estimate’;
4. intensifier

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

Results broken down by corpora and the two words

category		news	Total	tweets	Total	speech	Total
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	

Category 4

- (e1) a. Ugyanakkor a hosszabb trend is durva visszaesést mutat
 ‘At the same time, the longer trend also shows a rough
 (‘notable, significant’) decline’ [news]
- 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.’ [news]

Almost half the cases were intensifiers (2,999/1474); the frequency of
durva and *durván* are quite similar:

durva – 1,563/788, *durván* – 1,436/686

durva in the other two domains

tweets: 77/10, speech: no occurrence

durván in tweets is an intensifier in most of the cases:
26/21

- (e2) [...] 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.’ [tweets]

Categories 1 and 2

- generally underrepresented
- one noteworthy exception: *durván* in category 2
common in news texts (1,436/646): ‘approximately’

(e3) 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.’

Categories 3 and 5

- negative sentiments may be relatively easily distinguished

- (e4) a. durva diákcsíny ‘rude student prank’
 durva bánásmód ‘rude treatment’
 durva kijelentés ‘rude statement’
 durván viselkedik ‘sy behaves rudely’
 durván beszél ‘sy speaks rudely’ [news texts]

b. [...] 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’).’ [tweets]

- mostly unambiguous cases: both from syntactic and semantic points of view
- *durva* and *durván* are grammatically dependent and add a negative semantic component to the propositional meaning
- → SG (Heine et al. 2013)

Problems

1. sentiment value
2. sentiment word or pragmatic marker

Problem 1

- (e5) (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.' [tweets]
- *durva* may be interpreted in three ways:
negative, i.e. physically brutal, brutish
positive, i.e. impressive, interesting, worth watching
neutral, i.e. amazing, crushing

Problem 2

(e6) a. [tweets]

Durva. Még szerencse, hogy időben észrevettétek [...]

‘Tough! (‘Hell! / Damn!') It's lucky you noticed it in time [...]

(e6) b. [tweets]

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

Both occurrences seem to be syntactically independent.

- Example (6a): the first element of the turn, expletive-type reaction paraphrasable as *Huh!* ‘Hah’, *Azta!* ‘Whatta!?’ , *Basszus!* ‘Damn!’ → discourse marker → TG (Kaltenböck et al. 2011; Ahn and Yap 2022)
- ↔ example (6b): two interpretations: discourse marker like (6a) or negative sentiment word, i.e. negative evaluation of the previous or subsequent utterance
- if one takes into account that the second utterance does not stand syntactically alone
- but it is connected to the first one (Hungarian pronominal subject *ez* ‘this’ can be dropped) or
- to the third utterance is subordinated to the second one (the complementizer *hogy* ‘that’ is missing)

Differentiating non-negative (positive or neutral) cases

(e7) a.

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 sight.' [tweets]

(e7) b.

Annyira durva, amennyit beleteszel.

'It's as much tough ('cool') as you invest into it.' [tweets]

(e7) c.

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]

(e7) d.

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]

Negating replies

(e8) a.

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 don't think it's tough (‘incredible, surprising’) at all. This is to be expected from all channels.’

(e8) b.

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 don't think it's tough (‘incredible, surprising’). Everyone likes to have their say.’

Ambiguous cases of multiword expressions

(e9) (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.
- You swear?
- Seriously. December 11th.
- How tough.’ [speech]

(e10) a.

azta durva mindenit... véletlenül ide tévedtem és mit láttam??? új bejegyzések és ez kirááááály!!!!

‘Wow!/Hah... I wandered here by accident and what did I see??? new posts and it’s cool!!!!’

[ildisusatrip.blogspot.com/2012/06/24-ora-alatt-4-orzag-elso-nap.html]

(e10) b.

Azta durva! A kedvezményes jegyekből már csak 100 db a boltokban!

‘Wow! Only 100 of the discount tickets are left in the shops!’

[facebook.com]

Tendency of being a discourse marker

- *Durva* or *durván* forms an utterance alone and it tends to be interpreted as a discourse marker.
- several different pragmatic functions, e.g. support and solidarity (11a), attention marker (11b–c) , indicating that the listener can understand and follow what was said (on multifunctionality of DMs, see, among others, Fraser 2009 and Furkó 2020)

Support and solidarity

(11) 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.**’ [speech]

Attention marker

(11) 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.**’ [speech]

Attention marker

(11) c.

(Two men are talking. One of them does not know the meaning of the word *hedonist*, so he asks the other, who explains it to him.)

- Ki a „hedonista”?
- Az az, aki öö. Érzéki örömöket hajszoja.
- Hát fogalmam sem volt, hogy ez ezt jelenti. A „hedonista”.
- Habzsoja az életet.
- Habzsoja 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? **Ok, I see.**’ [speech]

Conclusions

- a detailed schema of the meaning structure of semantic and pragmatic functions the Hungarian adjective *durva* ‘rough, rude’ and its adverbial derivation *durván* ‘roughly, rudely’
- functions occurring in three corpora, representing three different text domains
- results of testing this schema in annotation of the corpora
- besides unproblematic cases, 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
- 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

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 schema 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 schema should be flexible to register ambiguities existing even in real contexts of discourse.

Acknowledgments

- The research of **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 of the Hungarian Academy of Sciences (MTA).
- The research work of **Martina Katalin Szabó** was funded by the International Research Fellowship Program of Japan Society for the Promotion of Science (JSPS, Postdoctoral Fellowships for Research in Japan (Standard)) and by the OTKA Postdoctoral Research Grant, the Hungarian Research Fund of the National Research, Development and Innovation Office of Hungary (NKFIH, grant number: PD 132312). This research project was partly supported by the Centre for Social Sciences, the Eötvös Loránd Research Network, the Hungarian Academy of Sciences Centre of Excellence.

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.
- Bárczi, Géza and Ország, László 1959-1962. A magyar nyelv értelmező szótára [A Defining Dictionary of the Hungarian Language], vols. I–VII. Budapest, Akadémia Kiadó (abbr.: ÉrtSz.).
- Furkó, B. Péter 2020. Discourse Markers and Beyond: Descriptive and Critical Perspectives on Discourse-Pragmatic Devices across Genres and Languages. Palgrave Macmillan, Cham.
- 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*. John Benjamins, Amsterdam, Philadelphia, 379–405.
- Kaltenböck, G., Heine, B. and Kuteva, T. 2011. On thetical grammar. *Studies in Language* 35(4), 852–897.
- Pusztai, Ferenc (ed.) 2003. Magyar értelmező kéziszótár [A Concise Dictionary of Definitions of Hungarian]. Akadémiai Kiadó, Budapest (abbr.: ÉKsz.²).

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. *Proceedings of the 25th Annual Meeting of Pragmatic Society of Japan*, Kyoto University, Japan. Paper submitted.

Szabó, M.K., Vincze, V. and Bibok, K. 2022. “Thank you for the terrific party!” – An analysis of Hungarian negative emotive words. *Corpus Linguistics and Linguistic Theory*. Open access: <https://www.degruyter.com/document/doi/10.1515/cllt-2022-0013/html>

Szabó M.K., Dam B., Vincze V. 2023. *On the Semantic Development of Negative Emotive Intensifiers in Hungarian News*. Manuscript.

Vincze, V., Üveges, I., Szabó, M. K. and Takács, K. 2021. A magyar beszélt és írott nyelv különböző korpuszainak morfológiai és szófaji vizsgálata In: Berend, G., Gosztolya, G. and Vincze, V. (eds.), *XVII. Magyar Számítógépes Nyelvészeti Konferencia*. Szegedi Tudományegyetem, Informatikai Intézet, Szeged. 169–182.

Thank you for your attention!