Corpus Analysis of Japanese and Hungarian Negative Emotive Words from a Discourse Interactional Perspective

Martina Katalin Szabó[1,2,3], Naoki Otani[1]

[1]Graduate School of Global Studies, Tokyo University of Foreign Studies, Tokyo, Japan
[2]Institute of Informatics, University of Szeged, Szeged, Hungary
[3]Computational Social Science – Research Center for Educational and Network Studies (CSS-RECENS), Centre for Social Sciences, Budapest, Hungary

Szabo.Martina@tk.hu
otani@tufs.ac.jp

<Abstract>
The main goal of this paper is to uncover the wide range of different uses and pragmatic functions of the so-called *Negative Emotive Words* (henceforth NEWs) based on the corpus data of two languages belonging to different cultures, namely Hungarian and Japanese. *NEWs* refer to words having a prior negative semantic content, but they can lose it partly or totally in certain use cases, e.g. En. *terribly nice* or *the party was terrific!* (Szabo and Bibok 2019). To that end, two NEWs have been selected: the Hungarian *durva* (lit.) 'harsh' and the Japanese やばい *(yabai)* (lit.) 'dangerous'. We analyze word-form variations of the two words, including *durván* of *durva* and *yabe* and *yabaku* of *yabai* (Putri 2021). Our investigation demonstrated similarities and differences between the two words, identifying several features about the Hungarian and Japanese NEWs that had not been known so far.

Keywords: Hungarian, Japanese, Negative emotive words, Politeness Theory, Intensifiers, Interjections

1. Introduction

This paper discusses Negative Emotive Words (henceforth NEWs) that have a prior negative semantic content (on their own, i.e. without context). It is known that these words often lose the negative semantic content partly or entirely in certain use cases, e.g. En. *terribly nice* or *the party was terrific!* (Szabo and Bibok 2019). In the literature NEWs are mainly discussed within the group of intensifiers. However, they can have several other meanings and functions as well (Szabó et al. 2022). For instance, they may express a positive evaluation or the surprise of the speaker, and they

may also be used as an interjection. In this paper we attempt to discover the range of different uses and pragmatic functions of NEWs in Hungarian and Japanese, on the basis of a corpus analysis. For the investigation, two NEWs have been selected: the Hungarian *durva* (lit.) 'harsh' and the Japanese やばい *(yabai)* (lit.) 'dangerous'. We also analyze word-form variations of the NEWs in question, such as *durván* of *durva* and *yabe* and *yabaku* of *yabai* (Putri 2021).

The paper is structured as follows: First, we briefly discuss the relevant literature on the examination of NEWs, with special regard to studies dealing with NEWs in the Hungarian and Japanese languages. Second, we present the corpus of the analysis along with the methodology of the corpus processing and annotation procedure. Third, we show the basic data of the annotation, and the results of the manual analyses. Fourth, we discuss the results concerning the Hungarian and the Japanese data, and then, from a comparative perspective. Throughout the text, some observations are made with corpus examples. Lastly, we summarize the main findings of the research work, and then state our intentions on plans for future study.

2. Literature review

Despite the fact that more and more authors have been investigating NEWs (e.g. Dragut and Fellbaum 2014; Jing-Schmidt 2007; Laczkó 2007; Paradis 2001, 2008; Szabó and Bibok 2019; Szabó et al. 2022; Wierzbicka 2002, among others), the authors who focus on functions of NEWs other that intensification is just a handful. For instance, as far as we are aware, only Andor (2011), Szabó (2018), Szabó and Bibok (2019) and Szabó et al. (2022) discuss the phenomenon when a NEW expresses a positive evaluation of the speaker towards the given information e.g. *brutális alaplap* (lit. 'brutal motherboard'–'high quality motherboard') (authors call this usecase *polarity shift*). In addition to the intensifier and polarity shifting functions, NEWs may have some other pragmatic function, e.g. in a certain context they express the surprise of the speaker or are used as an interjection (see above, Section 1). However, we are not aware of any paper that systematically scrutinizes all these functions and uses of NEWs based on a large language dataset.

There are some studies that analyze the phenomenon in the Hungarian language. They mainly discuss the issue from the viewpoint of semantic development of NEWs over time (e.g. Andor 2017; Varga 2019). Regarding specifically the word *durva* in the Hungarian language, to the best of our knowledge, there is only one paper (Andor 2011) that dealt specifically with this NEW, and another that examined this NEW together with other NEWs (Szabó et al. 2022).

As for Japanese, most of the papers discussing NEWs focus on their historical development, analyze their acceptability by Japanese speaker groups, or just take note of them as an existing phenomenon in the Japanese

language, which is especially popular in the language of the younger generation. Some of these papers focus specifically on the NEW *yabai* (e.g. Sano 2005; Konno 2015; Putri 2021).

3. The corpus and the methodology of the analysis

For the Hungarian cases, we analyzed the data of HuTongue Corpus, consisting of spontaneous speech texts (Vincze et al. 2021). The corpus consists of 1,149,457 tokens (without punctuation). For the Japanese cases, the Corpus of Everyday Japanese Conversation (CEJC corpus, Koiso et al., 2018) was applied, consisting of approximately 2,400,000 tokens. All occurrences of the NEWs in question were filtered out, and we annotated different syntactic, semantic and pragmatic features of NEWs.

As for syntactic features, we decided whether the NEW in the given example was an adjective, an adverb or an interjection, i.e. they constituted an utterance by themselves (Ameka 1992; Norrick 2009). Regarding semantics, we annotated evaluative semantic content, i.e. how the speaker evaluated what he or she said. There were three options in this case, namely positive, negative or neutral. We shall refer to this semantic feature here by the term *sentiment* from the research field of computational linguistics instead of a more general linguistic term polarity.[1] Besides sentiment, we annotated other semantic features, such as the exact meaning and function of the given NEW. We annotated three types: literal or figurative sense, intensifier function and emotive usage. In the latter case, the original negative semantic content is already bleached and the NEW expresses feelings or impressions of the speaker about the given piece of information, e.g. *A parti őrület volt!* 'The party was crazy!'. Then, we scrutinized all the occurrences from a pragmatics point of view. A study that systematically analyzes NEWs from a pragmatic viewpoint was not been conducted prior to the current analysis, so we did not have any pragmatic systematization that we could rely on. Thus, we referred to some pragmatic papers on interaction, with special regard to the Framework of Politeness Theory[2] (Quirk et al. 1985; Östman 1995; Kockelman 2003; Kádár 2017). This part of our analysis was basically exploratory; we investigated the range of pragmatic functions such as agreement and disagreement, calling someone's attention, a compliment, an evidential marker, an attention marker, etc. (Fraser 2009).

4. Results of the analysis

Firstly, let us review the Hungarian results. The raw frequencies of each form were the following: *durva*: 213 (0.0185%) and *durván*: 44 (0.0038%). Surprisingly, *durva* and *durván* have completely different features in the Hungarian corpus. The main function of *durva* is interjection (42.77%) and it does not have an intensifier function at all. Then, both *durva* and *durván* mostly have a positive or neutral meaning. In the case of *durva*, it accounts for 69.64% all together, while *durván* accounts for 62.92% of the total.

Durva has some specific pragmatic function in almost half the cases (47%). For instance, it occurs in an agreeing utterance, expressing agreement and support toward the speaker (1a), or functions as an attention marker indicating that the listener can understand and follow what was said (1b) (Fraser 2009):

(1) a. –Nem tudtam jobbra fordítani a fejemet. Az nagyon szar volt. 'I couldn't turn my head to the right. That was really shit.'
–**Durva**. '**Crazy**.'

b. –Ki a „hedonista"? 'Who is the „hedonist"?'
–[…]
–Habzsolja az életet. 'The one who really enjoys life.'
–Habzsolja az életet? **Durva**. 'Really enjoys life? **Ok / I see**.'

In contrast to *durva*, *durván* has an intensifier function in most cases (48.14%) and the category "emotive" is less represented here. *Durván* basically has no special pragmatic function, with two exceptions. When they do, they occur in an agreeing, supporting reply or they function as an evidential marker. For the latter, see (2) below.

(2) Hát igen, ez lenne az alap, csak ez Magyarországon már luxusnak számít, **durván**. 'Well yes, it should be a basic thing but in Hungary it is treated as some kind of luxury, **really**'

As for Japanese, the raw frequencies of each form were the following: *yabai*: 609 (0.025%), *yabe*: 37 (0.0015%) and *yabaku*: 57 (0.0024%). The details of the corpus investigation are shown in the table below:

	adjective	**adverb**	**interjection**	**noun**	**total**
yabai	224	0	377	8	**609**
yabaku	2	32	23	0	**57**
yabe	0	0	32	0	**32**
total	**226**	**32**	**432**	**8**	**698**

The three forms *yabai, yabaku* and *yabe* appear 698 times in total (*yabai* 609 times, *yabaku* 57 times and *yabe* 32 times). They show interesting distribution as to their parts of speech: though *yabai* and *yabe* is etymologically

adjectives and *yabaku* is an adverbial form of yabai, 62% of the three forms are used as interjections. That is, they do not modify anything. Rather, they show subjective feelings by the speakers.

Likewise, we found the interesting behaviors as to the sentiment features. Though *yabai* is originally related to negative values, reversal of polarity is observed: 15% of the three forms expresses positive evaluation. However, though the polarity of *yabai* is somehow weakened in some examples, Japanese *yabai* seems to hold either positive or negative evaluation at least to some extent. So, even if it intensifies the state illustrated by the modified nouns or verbs, *yabai* is not a pure intensifier; rather, it always includes the speakers' subjective evaluation to the modified objects.

5. Discussion

As we have seen, our findings identified several features on the frequency distribution of each wordform and their functions. In addition, differences in the frequency distribution of these meanings and functions of both the Hungarian and Japanese NEWs were noted.

Regarding the Hungarian data, we found that the wordform without a suffix, i.e. *durva* is represented in our corpus approximately five times more often than the suffixed form, i.e. *durván*. We also saw that *durva* and *durván* have markedly different features and functions. The main function of *durva* is interjection and this NEW does not have an intensifier function in the corpus at all. Then, in contrast to *durva*, *durván* does not occur as an interjection in the corpus at all. Our findings also showed that both *durva* and *durván* have a positive or neutral meaning in most cases. In the latter case, namely when the given NEW is neutral, it expresses the extreme nature of the topic or the surprise of the speaker. We think that the fact that the analyzed words are positive or neutral in most cases reflects the desemantization of these NEWs, i.e. their prior negative semantic content has already faded a lot.[3] We also learned that *durva* – in contrast with *durván* – usually has some specific pragmatic function and it is often used as a positive politeness marker, i.e. expressing agreement and support towards the speaker, as well as an attention marker. Based on the markedly different semantic features and syntactic patterning, we argue that both wordforms in question are in an advanced state of semantic-pragmatic development. However, this development has not led to the same result, i.e. exactly the same semantic-pragmatic function.

Regarding the Japanese data, our discussion can be summarized as follows: First, in the investigation, Japanese *yabai* is annotated based on the criteria that was used for the analysis of Hungarian *durva*. Though these criteria are effective for classifying various examples of *yabai* to some extent, it is necessary to focus on other features as well in order to capture the characteristics of *yabai* in more depth. Second, basically, the use of *yabai* can be classified into the two types. The first type shows the state of nouns or clauses such as "it is yabai" or "He is yabai." The second shows the subjective emotion of speakers such as "I feel 'yabai' to something". The former can be labeled

as "objective use" and tends to be realized as adjectives and adverbs, while the latter can be labeled as "subjective use" and tends to be realized as interjections. Third, positive evaluation of *yabai* tends to be associated with certain contexts, such as eating something or complimenting a dog. Fourth, the various pragmatic use of *yabai* might emerge from lexicalization of contextual meanings. When *yabai* co-occurs with certain phrase such as *yabai doushiyou* or *yabai yabai yabai*, its contextual meanings are entrenched in the lexical meaning of *yabai*, and the pragmatic/contextual meanings become a part of lexical meaning. Finally, a reduplication of *yabai* tends to have pragmatic functions. For example, *Yabai*, *yabai*, *yabai* has a function to ask for suggestions, meaning "what should I do."

As for the features of NEWs from a comparative linguistic perspective, our study showed that NEWs are able to express a range of speaker's subjective and intersubjective stances in both examined languages, and they should be viewed as noteworthy markers of positive politeness (despite this, it is not widely discussed topic in the literature). We also observed signs of semantic-pragmatic development (bleaching the prior negative semantic content) of NEWs, as well as the differences in the progress and results of this process. In addition, the change of word classes was observed in both datasets. For instance, NEWs can be used as interjections in both languages, indicating the speaker's emotion, but Japanese *yabai* and Hungarian *durván* are mainly used as intensifiers. As far as the main differences are concerned, as we mentioned, *durva* and *durván* are mostly neutral but *yabai* is negative in most cases. Then, *yabai* tends to be used as an expletive while Hungarian *durván* tends to be used as an intensifier.

6. Summary

Here, we analyzed the wide range of different usage and functions of Negative Emotive Words (NEWs) based on corpus data of the Hungarian and Japanese languages. Our findings uncovered several features of the Hungarian and Japanese NEWs (along with the similarities and differences between each examined wordform) that had not encountered so far, as well as similarities and differences between the two languages. Semantic bleaching of meaning, reversal and neutralization of negative semantic content, along with the change of word classes (e.g. interjections and intensifiers) were observed in both languages. Closely connected to all of these, NEWs are able to express a range of speaker's subjective and intersubjective stances. At the same time, identified semantic-pragmatic functions and syntactic patterning highlighted the progress of these NEWs in the semantic-pragmatic development at different levels. What is more, this development has not led to the same result, even within each language.

In the future, we intend to carry out a more detailed analysis of the pragmatic features of NEWs. In the study, we plan to especially focus on the semantic and pragmatic development of NEWs over time, utilizing historical corpora and corpora of other domains as well.

[1] In pragmatics, polarity is, in essence, the relation between semantic opposites, such as *hot-cold* (Israel 2004). Since we are interested in the evaluative semantic content, the term sentiment – referring to the evaluative meaning specifically – is more accurate and appropriate here.

[2] A speech act can threaten the other's "negative face," their wish to be left in peace, or "positive face," their wish to be appreciated (Kádár 2017). Politeness comes into existence with the other's face needs in mind: In order to ameliorate the effect of a face threatening act, the speaker will make use of negative politeness strategies. What is more, positive politeness strategies express friendliness and solidarity (Daly et al. 2004).

[3] NEWs proceed through a desemantization procedure over time, during which they gradually lose their lexical meaning (for more details, see Szabó et al. 2022).

References

Ameka, F. 1992. "Interjections: the Universal yet Neglected Part of Speech." *Journal of Pragmatics* 18, 101-118.

Andor, J. 2011. "De durva ez a tema! – Megfigyelések a melléknévi polaritásváltásról." [This is a Damn Topic! – Observations on Polarity Changing of Adjectives] *Hungarológiai Évkönyv* 12., 33-42.

Andor, J. 2017. "A lexikai fokozás polaritásváltásairól." [On the Polarity Change of Lexical Intensification] In Gecső T. and Szabó M. (eds.) *Az ellentét nyelvi és nonverbális kifejezésének lehetőségei,* 11-17. Székesfehérvár–Budapest: Kodolányi János Főiskola–Tinta Könyvkiadó.

Daly, N., Holmes, J., Newton, J., Stubbe, M. 2004. "Expletives as Solidarity Signals in FTAs on the Factory Floor." *Journal of Pragmatics* 36(5), 945-964.

Dragut, E., Fellbaum, C. 2014. "The Role of Adverbs in Sentiment Analysis." In *Frame Semantics in NLP: A Workshop in Honor of Chuck Fillmore*, 38-41. ACL, Baltimore, MD.

Fraser, B. 2009. "An Account of Discourse Markers. " *International review of Pragmatics*, 1(2), 293-320.

Israel, M. 2004. "The Pragmatics of Polarity. " In Horn, L., Ward, G. (eds.) *The Handbook of Pragmatics*, 701-723. Oxford: Blackwell.

Jing-Schmidt, Z. 2007. "Negativity Bias in Language: A Cognitive-Affective Model of Emotive Intensifiers." *Cognitive Linguistics* 18(3), 417-443.

Fraser, B. 2009. "An Account of Discourse Markers." *International Review of Pragmatics* 1(2), 293-320.

Kádár, D.Z. 2017. Politeness in Pragmatics. In *Oxford Research Encyclopedia of Linguistics*.

Kockelman, P. 2003. "The Meanings of Interjections in Q'eqchi' Maya: from Emotive Reaction to Social and Discursive Action." *Current Anthropology* 44, 467-490.

Koiso, H., Den, Y., Iseki, Y., Kashino, W., Kawabata, Y., Nishikawa, K. Y., ... & Usuda, Y. (2018). Construction of the Corpus of Everyday Japanese Conversation: An Interim Report. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation* (LREC 2018), 4259-4264.

Konno, H. 2015. "The Grammatical Significance of Private Expression and its Implications for the Three-Tier Model of Language Use." *English Linguistics* 32(1), 139-155.

Laczkó M. 2007. "Napjaink tizenéveseinek beszéde szóhasználati jellemzők alapján." [The Speech of Today's Teenagers Based on Word Usage Characteristics] *Magyar Nyelvőr* 131(2), 173-184.

Norrick, N. R. 2009. "Interjections as Pragmatic Markers." *Journal of Pragmatics* 41(5), 866-891.

Östman, J.-O. 1995. "Pragmatic Particles Twenty Years After." In Wårvik, B., Tanskanen, S.-K., Hiltunen, R. (eds.) *Organization in Discourse. Proceedings from the Turku Conference*, 95-108. University of Turku.

Paradis, C. 2001. "Adjectives and Boundedness." *Cognitive Linguistics* 12(1), 47-65. https://doi.org/10.1515/cogl.12.1.47

Paradis, C. 2008. "Configurations, Construals and Change: Expressions of Degree." *English Language and Linguistics* 12(2), 317-343. https://doi.org/10.1017/S1360674308002645

Partington, A. 1993. "Corpus Evidence of Language Change: The Case of the Intensifier." In Baker, M., Francis, G., Tognini-Bonelli, E. (eds.), *Text and Technology: In Honour of John Sinclair*, 177-192. Amsterdam: John Benjamins.

Putri, A.A., Haristiani, N. 2021. "Register Analysis on High School Students' Language in Japanese Manga and Anime." In *Fifth International Conference on Language, Literature, Culture, and Education (ICOLLITE 2021)*, 104-111. Atlantis Press.

Quirk, R., Greenbaum, S., Leech, G., Svartvik, J. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
Sano, S. 2005. "On the Positive Meaning of the Adjective *yabai* in Japanese." *Sophia Linguistica: Working Papers in Linguistics* (53), 109-130.
Shioda, T. 2003. "Attributive Variation in Popular New Expressions and the Media: '-tte-iuka', 'yabaku-nai?' From the Nationwide Survey on Linguistic Change in the Japanese Language in the Recent Years (1)." *The NHK Monthly Report on Broadcast Research* (53-4), 12-33.
Sinclair, J.M. (ed). 1987. *Looking up*. London: Collins COBUILD.
Sinclair, J.M. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
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 on Hungarian Texts, with Particular Regard to Polarity Loss and Polarity Shift] PhD Thesis. University of Szeged, Doctoral School in Linguistics, Ph.D. Programme in Hungarian Linguistics, Szeged, Hungary. June 2018. https:// doi.org/10.14232/phd.9976
Szabó M.K., Bibok K. 2019. "Értékvesztésre és értékváltásra képes lexémák újabb vizsgálata." [A novel Analysis of the Elements that Can Undergo a Loss of Polarity and Polarity Shift] *Argumentum* 15, 639-649.
Szabó M.K., Vincze V., Bibok K. 2022. "'Thank You for the Terrific Party!' – An Analysis of Hungarian Negative Emotive Words". *Corpus Linguistics and Linguistic Theory*. https://doi.org/10.1515/cllt-2022-0013. In press.
Szabó M.K., Dam B., Vincze V. 2023. *On the Semantic Development of Negative Emotive Intensifiers in Hungarian News*. Manuscript.
Varga M. 2019. *Középmagyar kori világi szövegek nyelvtörténeti vizsgálata, különös tekintettel a szövegtani és pragmatikai sajátságokra.* [A Historical Linguistic Investigation of Middle Hungarian Secular Texts: Textual and Pragmatic Aspects] PhD Thesis. ELTE, Nyelvtudományi Doktori Iskola.
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* 169-182. Szegedi Tudományegyetem, Informatikai Intézet, Szeged. ISBN 9789633067819
Wierzbicka, A. 2002. "Australian Cultural Scripts – *Bloody* Revisited." *Journal of Pragmatics* 34(9), 1167-1209.