STUDIA UBB PHILOLOGIA, LXX, 4, 2025, p. 219 - 246
(RECOMMENDED CITATION)
DOI:10.24193/subbphilo.2025.4.12

POLARITY IN THE VERBAL DOMAIN. PRELIMINARY RESULTS AND CONSIDERATIONS[1]

Hajnalka DIMÉNY[2]

Article history: Received 7 May 2025; Revised 20 October 2025; Accepted 20 November 2025; Available online 12 December 2025; Available print 30 December 2025

ABSTRACT. ***Polarity in the Verbal Domain. Preliminary Results and Considerations*****.** The paper presents preliminary results of an extensive investigation of polarity in the verbal domain of Hungarian. Polarity has manifold and wide-ranging connections to different aspects and subdomains of the language system. The present study is concerned with the distribution of polarity in the lexicon and its possible motivations.

In language studies, it has long been hypothesized that negative polarity is richer in its representation than positive polarity on a lexical, constructional and idiomatic level as well. The preliminary results of the investigation show that, in the verbal domain, this is indeed the case. This is not surprising, but the nature of the phenomenon and its ramifications deserve further linguistic exploration.

Other important considerations presented in the study regard the definition of polarity. From a quantitative approach, it seems obvious that affective and evaluative meaning cannot be separated from socio-cultural knowledge and subjective expectations. This last category does not differentiate itself distinctly

[1] I would like to thank the anonymous reviewers for their valuable comments and suggestions, which have greatly helped in improving the quality and clarity of this paper.
The work was funded by the Domus program of the Hungarian Academy of Sciences.

[2] **Hajnalka DIMÉNY** is a university assistant at Babeș-Bolyai University Faculty of Letters, Cluj-Napoca. She has defended her PhD in 2021 with the title *Main features of verbal meanings and lexicalization patterns of Hungarian verbs. Criteria for a semantic typology of Hungarian verbs*. She is interested in semantic phenomena and works in the paradigm of cognitive linguistics, following her supervisor, Sándor SZILÁGYI N. She is also a proponent of science popularization, and works a lot with students interested in linguistic problems, as the head of the Márton Gyula Linguistic College. Contact: hajnalka.dimeny@ubbcluj.ro.

from other, more robust cases of polarity. The connection is also strengthened by structural properties of verbs and by the negativity bias. Because of this, I believe that, besides affective content and evaluative meaning, although strongly connected to these, negative polarity verbs also encode non-preferred alterations to events, while positive polarity verbs encode preferred alterations or outstanding forms of events. For an event to be outstanding in some respect, is relevant in fewer contexts, than excelling in non-preferred ways. This explains the higher elaboration of the negative domain, but makes polarity as a verbal phenomenon even harder to grasp and define.

Keywords: *polarity, verbal meaning, distribution across the lexicon, socio-cultural entrenchment*

REZUMAT. ***Polaritatea în domeniul verbal. Rezultate și considerații preliminare.*** Lucrarea prezintă o cercetare preliminară în cadrul unei investigații extinse asupra polarității în domeniul verbal al limbii maghiare. Polaritatea are conexiuni multiple și de amploare cu diverse aspecte și subdomenii ale sistemului lingvistic. Studiul de față se concentrează asupra distribuției polarității și a posibilelor sale motivații.

În studiile lingvistice, există de multă vreme ipoteza că polaritatea negativă are o reprezentare mai bogată decât polaritatea pozitivă, atât la nivel lexical, cât și la nivel construcțional și idiomatic. Rezultatele preliminare ale cercetării arată că, în domeniul verbal, această ipoteză se confirmă. Acest lucru nu este surprinzător, însă natura fenomenului și implicațiile sale merită o explorare lingvistică mai profundă.

Alte considerații importante prezentate în lucrare privesc definirea conceptului de polaritate. Dintr-o perspectivă cantitativă pare evident că sensurile afective și evaluative nu pot fi separate de factorii și așteptările socio-culturale. Această ultimă categorie nu se diferențiază clar de alte cazuri mai robuste de polaritate. Legătura este consolidată și de proprietățile structurale ale verbelor și de către tendința spre negativitate. Din acest motiv, consider că, pe lângă conținutul afectiv și semnificația evaluativă – deși strâns legate de acestea – verbele cu polaritate negativă codifică și modificări ale evenimentelor considerate ca fiind nedorite sau nepreferate, în timp ce verbele cu polaritate pozitivă codifică forme remarcabile ale evenimentelor. Un eveniment remarcabil într-un anumit sens este relevant în mai puține cazuri, decât manifestarea modurilor nedorite sau nepreferate. Acest lucru explică reprezentarea lexicală mai bogată al domeniului negativ, dar face ca polaritatea, ca fenomen verbal, să fie și mai dificil de înțeles și de definit.

Cuvinte-cheie: *polaritate, sens verbal, distribuție în lexicon, factori socio-culturale*

1. Introduction

In language studies, it has long been hypothesized that negative polarity has a higher level of lexical sophistication (or lexical diversity) than positive polarity on a lexical, constructional and idiomatic level as well. However, substantial research on polarity in lexical elements, especially in verbal lexemes, has not attracted significant attention to date. The paper presents preliminary work for an extensive investigation which aims to contribute to this subject of linguistics by offering a comprehensive analysis of polarity in verbal meanings, on the one hand, and testing the accuracy of automatic annotation tools, on the other hand.

The term *polarity* is used broadly in the paper, referring to positive or negative semantic content in terms of emotive states, social judgements or subjective/moral values expressed by the lexical elements. This procedure is motivated on the one hand by the observation that positive or negative semantic content expressed by verbs can be related to (1) emotions (e.g., *giggle, enjoy, rejoice; fear, fret, shock*), but also to (2) (dis)satisfaction with some features of experience (e.g., *beaver away, brighten, encourage; gorge, misbehave, ramble*), and (3) moral/social values (e.g., *appease, compliment, revive; demoralize, embarrass, reprimand*). On the other hand, extensive approaches to positive and negative content also embrace a wider stance (Martin and White 2005; see chapter 3 for details). Starting from these, and because the aim was to have a comprehensive grasp of polarity in the verbal system, a wider approach to the phenomenon was adopted.

Polarity phenomena in natural languages are the entailment of the basic evaluative and affective character of language. Consequently, they are pervasive, quite diverse, and most likely to some extent universally realized. Besides space and time, polarity is one of the main organizational patterns in the intrafield relational structure (Faber and Mairal Usón 1997b). Thus, an extensive analysis of polarity goes beyond formulating statements about the investigated linguistic system proper. Polarity phenomena can be correlated with meaning extension and analogical changes in the language system, with grammatical and constructional phenomena through semantic prosody, they are tied inseparably to usage-based properties and affective-cognitive foundations of language, to the linguistic expression and reinforcement of biases, and so an. Thus, the comprehensive study of polarity in an entire subsystem of a language (the verbal system in our case) can have far-reaching applications.

The preliminary results of the analysis of 29 905 Hungarian verb meanings processed by one annotator, lead to several important considerations and to the formulation of two hypotheses. First, they suggest that in the verbal

domain, indeed negative polarity is lexically more diverse and more richly represented than positive polarity. This is not surprising, but the nature of the phenomenon, its ramifications and motivations deserve further linguistic investigation. This is carried out in chapters 4.1. and 4.2. Second, the data makes it possible to investigate the correspondences of polarity and word length, testing whether negative polarity verbs are longer on average than positive polarity verbs. According to the Pollyanna hypothesis, evaluatively positive terms are used more often and more diversely than evaluatively negative words (Boucher and Osgood 1969). In the same time, Zipf's law states that words used more frequently are shorter in length and carry less information (Zipf 1936, 1949). From these, we can hypothesize that negative polarity words carry more information and hence are longer in average than evaluatively positive words. This hypothesis is tested in chapter 4.3.

Although further developments of the dataset are needed, the case study suggests that there are important and wide-ranging correspondences to be investigated. Moreover, the data lays the foundations for a future comparison of manual and automated polarity labelling based on dictionary entries, so as to be able to test the accuracy of automated labelling.

Polarity is often encoded in lexical elements of the language. At the same time, the lexicon of a language expresses different characteristics of the community using it, even other factors than emotions and evaluation. Thus, polarity phenomena in natural languages, originally most probably the expression of evaluative and emotive content (see the negativity bias as a cognitive-affective pattern, Jing-Schmidt 2007), are deeply intertwined with socio-cultural and pragmatic factors. To be able to grasp polarity as an element extensively present in the lexicon, the study uses a dataset of all verbal entries of an explicatory dictionary. Because these constitute an abstract, expansive representation of the language, the findings are representative of the entirety of the verbal system, with the possibility of extending it to the verbal system of other natural languages.

Linguistic analyses have so far focused mainly on negative polarity items (see Giannakidou 2011 for an overview), on the language of basic emotions (e.g., Wierzbicka 1999), on negative emotive intensifiers (Jing-Schmidt 2007; Kugler 2014; Szabó, Vincze and Bibok 2023), on the usage frequency of positive and negative polarity words (Garcia, Garas and Schweitzer 2012), and on sentiment analysis (e.g., Cambria and Hussain 2015; Hua et al. 2024). However, only a few studies have been carried out regarding the distribution of positive and negative lexical items in the entire lexicon of natural languages, and its ramifications to other subdomains of the language. The most notable is that of Boucher and Osgood (1969), they, however, examined only a small portion of the vocabulary, though on a large language set.

In response, the present study brings new insights regarding the distribution of polarity verbs in the lexicon of the Hungarian language, but also makes quantitative claims regarding the possible socio-pragmatic motivations of it. The findings constitute important data for studies of polarity, of the verbal semantic network, and also corroborate usage-based approaches to language which state that "language is an embodied and social human behavior", and that "usage has an effect on linguistic structure" (Bybee and Beckner 2009, 827). Nonetheless, the results presented are only preliminary, and are to be treated as such. Because of the size of the dataset, automated labelling is planned in the future.

2. The data of the study

As mentioned earlier, the study is based on a dataset of 29 905 Hungarian verb meanings and their relevant semantic features. The backbone of the data was compiled by Sándor Szilágyi N., and it has been complemented and adjusted for the present study by the author.[3] This database comprises the entries of the second largest Hungarian explicatory dictionary (Pusztai 2006), treating each separate meaning and its dictionary interpretation as separate data. A powerful aspect of the present study is (1) that it is based not on an arbitrarily chosen, small fragment of the lexicon, but on a large, extensive sample of the language comprised in an explicatory dictionary, and (2) that it deals with all word meanings separately. Its limitations are also twofold, and are related (1) to the use of one single annotator and (2) to the fact that dictionary data never represents the language use of any speaker. The first one is to be addressed in the future.

The coding for event type was done by the author and followed the lexical domains of the functional lexicological model introduced by Martín Mingorance (1990). The term lexical domain refers here to "the set of lexemes which together lexicalize all or part of a conceptual domain" (Faber and Mairal Usón 1999: 59). This model was also used by Faber and Mairal Usón in their work for a typology of predicate schemata of English (1997a, 1997b) and for constructing a lexicon of English verbs (1999). The main reason for this choice is the practical procedures used by the authors in determining the lexical domains, which make the model easily applicable to the present data. The premise of this method is that dictionary entries contain the keywords for identifying the different relations between word meanings (e.g., causation, part–whole relation,

[3] Szilágyi N. and Dimény (2021). The data underlying this article are not yet available online because of copy rights, but will be gladly shared on reasonable request to the authors.

hyponymy, and hypernymy). Specifically, the use of one verb in the definition of another suggests that the two meanings are related, while the phrasing of the dictionary entry elaborates this relation. Therefore, starting from entirely specific events, in a few steps, I was able to categorize each verb meaning into a broad category of abstract event type and several subtypes.

The lexical domains in the Functional Lexicological Model are the following: EXISTENCE, POSITION, CHANGE, POSSESSION, PERCEPTION, EMOTION, COGNITION, SPEECH, GENERAL ACTION, MOVEMENT, and SENSE PERCEPTION (Faber and Mairal Usón 1999: 88, annotations follow the original). This event typology is not entirely adopted from Faber and Mairal Usón, but adapted to the study data and cross-checked with other verb ontologies, specifically WordNet (Princeton University 2010) and the Rich Event Ontology[4] (Windisch Brown et al. 2017). The motivations for using a somewhat different approach in the verbal typology are manifold. (1) The nature of the semantic content of Hungarian verbs was taken into consideration. In these, the encoding of non-dynamic events seems to be more prevalent than among English verbs (for a few examples see *örül* 'be happy', *szomorkodik* 'be sad'; *zöldell* 'be green', *rozsdállik* 'be rusty'). This resulted in the differentiation of STATES (like being happy) and ATTRIBUTES (such as something being green). (2) Verbs expressing BEHAVIOR have a specific form in Hungarian, besides their specific meaning structure (many of them end in *-kodik/-kedik/-ködik*)[5], and because if this they were considered to be a class of their own. (3) A greater focus has been placed on the nature of the primary figure (or subject argument) to the verb, because in the dictionary the different meanings of verbs are mainly separated based on collocational properties, with strong emphasis on the nature of the primary (and in many cases also the secondary) figure (human being, animals, inanimate things, or natural phenomena etc.). Related to this, (4) the intentionality of the primary figure also received a role. This is also a main feature in separating different verbal meanings in explicatory dictionaries (e.g., verbs referring to human activity in their primary meaning and to inanimate or complex things in their figurative meanings, like *megy* 'go' and *fordul* 'turn', or the English verbs *work* and *run*). This resulted in the more numerous class of verbs encoding ACTION/ACTIVITY, including all intentional activity, even speech and sense perception, beside the classes of HAPPENING, EXISTENCE and PERCEPTION, which encode non-intentional events. Finally, (5) taking into consideration previous literature on verb ontologies

[4] The Rich Verb Ontology comprises 161 verb classes and a multiple nodes deep hierarchy, with, for example, the following classes on the mid-level: COGNITIVE EVENT, LIFE EVENT, INTENTIONALLY ACT, MOTION (annotations follow the the authors). This was too detailed for the present purposes, was provided important insights for the ontology.

[5] Although, this suffix is not exclusive to this class of verbs, it can also refer to other activity types (e.g., *fésülködik* 'to brush one's hair', *tanárkodik* 'to be working as a teacher').

(specifically WordNet, the Rich Event Ontology and Leonard Talmy's semantic model [2000]), MOVEMENT has been distinguished from the more general class of ACTIVITY. Not incidentally, verbs encoding movement are considered to be prototypical verbs by several authors (e.g., Langacker 2008, Tolcsvai Nagy 2016).

These considerations resulted in the following basic event types: HAPPENING, ACTION/ACTIVITY, MOVEMENT, BEHAVIOR, EXISTENCE, PERCEPTION, STATE, and ATTRIBUTE/CHARACTERISTIC. These are expressed most generally in the dictionary entries by verbs such as *történik* 'happen'; *tesz, csinál* 'do'; *megy* 'go', *mozog* 'move'; *viselkedik* 'behave'; *létezik* 'exist', *él* 'live, be alive'; *észlel* 'perceive'; *van* 'be', *mutatkozik* 'appear as', and the adjective *képes* 'be able'. Within some of these basic event types, we can speak of subtypes, most prominently about subtypes of activity verbs. At this moment, these were identified only partially based on the dictionary entries and on the semantic literature available. For example, we can speak of verbs of speaking, verbs of eating and drinking (or digestion), verbs of sound, verbs of transfer, verbs of possession and change of possession, etc. These are only used as examples in this study. Naturally, it could be argued that a different event typology would lead to different results. Remarks regarding this are made in the Discussion part.

As it can be surmised, a typology of few classes was aimed for, with a high emphasis on the nature and characteristics of the primary figure of the event structure. The main reasons for this were (1) the nature of the dataset and (2) the aims of the study at the present moment. For more detailed analyses a more granular typology will be needed in the future.

The coding for polarity is also based on the phrasing of the dictionary entries, because these can be considered a representation of the shared knowledge of linguistic encoding. In the next chapter, I elaborate on the exact method I have used and on the challenges of coding for polarity on this type of data and in general. Here, I will only mention that 456 meanings have multiple labelling, because the dictionary entries sometimes group similar usages which may have different polarities. Verbs referring to being in an emotional state are a typical case of this, because they can express both positive and negative polarity depending on the argument they take. See the entries in (1) for an example.

(1) elragad
'carry away'

<Érzelem,	indulat>	hatalm-á-ba	kerít	és	heves szav-ak-ra,
feeling	emotion	power-GEN.3SG-ILL	get.hold.of.3SG	and	fierce word-PL-SUB
cselekedet-ek-re		késztet	vki-t.		
act-PL-SUB		lead.3SG	someone.3SG-ACC		

'<Feeling, emotion, temper> takes over someone, and leads him/her to some impulsive words or acts.'

In (1), the polarity of the verb (and consequently of the sentence, or vice versa) depends on the argument the verb takes. If someone is carried away by a positive feeling or by an event in a positive sense, the verb is associated with positive polarity. But in the case of an argument that expresses a negative feeling or state, negative polarity is associated. In this case, being carried away itself becomes negative.

Multiple labelling is used to suit the aims of the study; by using the database, to be able to account for the encoding of polarity in the language in general.

3. Theoretical prerequisites of polarity

In linguistics, the term *polarity* is most commonly used to refer to the distinction between positive and negative semantic content. This distinction seems quite intuitive, and thus, polarity is often left undefined. On the one hand, the term *polarity* can refer to a grammatical feature of an item associated with affirmation or negation. In this case, we speak about *polarity items*, i.e., linguistic items expressing affirmation or negation (see Giannakidou 2011 for an overview). On the other hand, on a lexical-semantic level, *polarity* refers to the affective evaluation of the speaker expressed by linguistic elements, used mainly in sentiment analysis. In this case, polarity can be defined as an evaluative stance, or as a positive or negative attitude, encoded in the meaning of a lexical item. However, this evaluative stance can be defined in several ways. Martin and White (2005, 39) offer a summary of different approaches to evaluative language analysis and its terminology.

Table 1. A summary of approaches to evaluative language analysis and its terminology as compiled by Martin and White (2005)

Approaches to evaluation	**'entity focused'**	**'proposition focused'**
Chafe & Nichols 1986		evidentiality
Ochs & Schiefflen 1989	affect specifiers	affect intensifiers
Biber & Finnegan 1989	affect	evidentiality
Wierzbicka 1990b	emotion	
Bybee & Fleischman 1995	evaluation	modality
Niemeier & Dirven 1997	emotion	
Conrad & Biber 2000	attitudinal stance	epistemic stance
Hunston & Thompson 2000	opinions about entities	opinions about propositions
Hunston 2000	'status' and 'value' on the 'autonomous plane'	'status' and 'value' on the 'interactive plane'

The authors themselves use the term *appraisal* to refer to the more general evaluative meaning in language. It is "one of the three major discourse semantic resources construing interpersonal meaning (alongside involvement and negotiation). Appraisal itself is regionalised as three interacting domains – 'attitude', 'engagement' and 'graduation'" (Martin and White 2005, 34–35). *Affect* refers to the expression of emotional experiences, *judgement* is connected with behavior, and *appreciation* is concerned with the value of things, including natural phenomena and semiosis.

The more recent and more popular computational linguistic analyses include all these in their investigation of evaluative language, and use the term *sentiment* to refer to them. In computational linguistics, a distinction between entity-level, propositional-level and aspect-based sentiment analysis is being held. E.g., in their systematic review of 727 primary studies on sentiment analysis, Hua et al. (2024, 296) refer to the subject of their investigation as "views and feelings (i.e. sentiment"), but delimit their interest to how these are expressed in user reviews, social media posts, and open-ended survey question responses.

The present paper follows this line of wide approach to evaluative meaning, and uses the terms *polarity* to refer to the emotive states, social judgements and subjective/moral values pertaining to the meaning of words (cf. affect, judgement, and appreciation in Martin and White [2005] mentioned above). This decision is motivated on the one hand by the observation that positive or negative semantic content expressed by verbs can be related to (1) emotions (e.g., *giggle, enjoy, rejoice; fear, fret, shock*), but also to (2) (dis)satisfaction with some features of experience (e.g., *beaver away, brighten, encourage; gorge, misbehave, ramble*), and (3) moral/social values (e.g., *appease, compliment, revive; demoralize, embarrass, reprimand*).

Starting from these, and because the aim was to have a comprehensive grasp of polarity in the verbal system, a wider approach to the phenomenon was adopted. The study only investigates polarity expressed on a lexical level, in verbs, but in the process of coding for polarity a sentence level analysis was used, given the nature of the data.[6] Thus, a sentiment polarity assigned to the sentence explaining a certain meaning, is assigned to the meaning in question. Sentence level sentiment analysis determines whether a positive, negative or neutral opinion is expressed in a sentence (Liu 2012).[7] It takes content words appearing in a sentence, and assigns a value based on the polarity expressed by them.

---

[6] Preverbal forms of verbs that appear as independent entries in the dictionary, are not treated separately, nor they are structurally analysed. Although there are cases in which the preverbal prefix is the morpheme carrying polarity meaning (cf. Evellei 2012), in this investigation they are not treated differently, and polarity is assigned to the whole of the lexical form.

[7] Liu (2012: 7) uses the terms *sentiment analysis* and *opinion mining* interchangeably to refer to "the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes,

In the present study, the coding for polarity was performed by only one person at the moment, the author. However, some measure of accuracy is ensured by the fact that all labelling is based on the dictionary entries, which are the result of the collective work of lexicographers and are the representation of a shared knowledge of the language. Naturally, there are plans for expanding the team of annotators for testing the accuracy of the present annotation.

Three classes make up the polarity coding in the database. Entries containing negative words were coded as "negative", and entries containing positive words were coded as "positive". Those not containing neither negative, nor positive words were coded as "neutral". By the terms *positive words* and *negative words,* I mean the following:

> 1) they refer to an event, state or experience that is in itself widely considered, subjectively, to be positive (good, acceptable) or negative (bad, not acceptable), e.g., *győz* 'to win', *öröm* 'joy', *jó, jól* 'good, well', *hálás* 'grateful', etc.; and *rossz* 'bad', *aggodalmaskodik* 'to worry', *szégyen* 'shame', *megaláztatás* 'humiliation', *erőszakosan* 'aggressively', etc.; here the motivation can vary from emotional experiences (like happiness, worrying) to evaluation of things, experiences or behavior.
> 2) words with positive or negative connotation (rarely also noted in the dictionary by information regarding style/register, like *vulgar*, *euphemistic*), e.g., *tisztel* 'to respect', *kedves* 'nice', *érték* 'value', etc.; and *megbán* 'to regret', *sajnálat* 'pity', *le-* 'down' (in some cases), etc.

Although based on keywords expressing evaluation, the coding takes into account the whole of the dictionary entry (which is practically a sentence). This means that in case words with different polarities appear in an interpretation (which is sometimes the case), the coding is adjusted to the meaning in question. See example (2) for a demonstration. The [+] and [-] signs in subscript indicate the polarity of the lexical item in question.

(2) félrevezet$_{[-]}$
'mislead'

Vki-t	a	helyes$_{[+]}$	útirány-tól	eltérít.
someone.3SG-ACC	ART	right	direction-ABL	divert.3SG

'Lead someone in the wrong direction.'

and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes." That is, although a difference can be made between all these relations to the texts produced by humans, the term *sentiment* is nowadays used very broadly. This way, practically sentence level sentiment analysis was performed when detecting the polarity of a dictionary entry, which was then assigned to a verb meaning.

In an automatic analysis, the presence of the word *helyes* 'right' would lead to a "positive" polarity label, because other words in the entry carry neutral polarity (*valaki* 'someone', *eltérít* 'divert', *útirány* 'way, direction'). However, the presence of the verb *eltérít* 'divert' changes the polarity suggested by the word *helyes* 'right', the verb expressing that the right way is bypassed. Not taking context into consideration would lead to an erroneous classification of the meaning, labelling it with positive polarity instead of negative polarity. The human annotator eliminates this problem, as well as that of idiomatic expressions, questions, conditionals and grammatical elements such as negation. These were also taken into account, and idiomatic expressions were treated as a whole. Idioms are often listed as separate meanings of a verb in the dictionary. In an automatic classification, these would most likely cause problems and inaccurate categorization. However, the relatively detailed dictionary entries to the meanings and the human annotator, deal with them easily and, most importantly, accurately. All ambiguous cases were cross-checked against the larger, though much less recent Hungarian explicatory dictionary (Bárczi and Országh eds. 1959–1962).

For another case of polarity assignment, see example (3) and (4) below. The lexical elements encoding polarity, which determine the polarity assignment of the verbal meaning in question, are marked with bold. The content words which are not marked this way, do not encode polarity.

(3) adományoz$_{[+]}$
'donate'

<Nagyobb	összeg-et,	**érték-**et$_{[+]}$>	**jótékony**$_{[+]}$	vagy
bigger	sum-ACC	value-ACC	charitable	or
közösségi	**cél-**ra$_{[+]}$	ad.		
community	purpose-SUBL	give.3SG		

'Donate a large sum of money or valuables to charity or for community purposes.'

(4) akadékoskodik$_{[-]}$
'demur'

Vki-t	**túlzott**$_{[-]}$,	**kicsinyes**$_{[-]}$	**aggályoskod ás-**sal$_{[-]}$	gátol.
someone.3SG-ACC	exaggerated	petty	worry-INS	hinder

'Someone hinders another person with excessive, petty concerns.'

(5) változik
'change'

Folyamatosan	más,	lesz,	mint	amilyen	volt
continually	different	will.be	like	such.as	was

'It continually becomes different than it used to be.'

The database still needs some improvements and testing for accuracy. At the moment it only makes possible qualitative analyses and the formulation of hypotheses regarding quantitative aspects. The use of a dataset based on dictionary entries has some advantages that deserve to be mentioned. Most prominently, the data is not only an arbitrarily chosen fragment of the lexicon, but the representation of the entire verbal system of the standard variant of Hungarian. Another benefit is that the database is built not on verbs, but on verb meanings, and this makes it possible to account not only for the primary meanings, but for all meanings listed in the dictionary separately. Thus, the data reflects the entire verbal domain of the language.

An analysis performed on the data at this point, leads to several important lessons and observations which will help to improve the dataset. Overall, the database created specifically for the study of verbal semantics is a useful and effective tool for investigating polarity information encoded in verbs and for formulating domain-general hypotheses.

4. Results and discussion

All findings and correlations are to be treated carefully, as preliminary results. Here they are used to formulate hypotheses and to refine future investigations. Although some of the findings are in line with previous studies, because of the use of a single annotator, the correspondences found can only be regarded as guidelines in subsequent analyses, which are of great variety: regarding polarity more generally, motivations of its lexical encoding, correspondences with meaning extension mechanisms, possible role in analogical changes, and so on.

4.1. Distribution of polarity in the study data

A total of 21 061 verb meanings are labeled "neutral" in the database, and 456 meanings are labeled "positive" and/or "negative" and/or "neutral", because some meanings encode more than one polarity. This gives us a total of 70.43% of neutral polarity in all verb meanings. While the language of the Internet (and most

likely everyday language, too) is "emotionally charged", as shown by Garcia, Garas and Schweitzer (2012), in the overall lexicon of a language, neutral words predominate, which is not surprising. Only a total of 1 977 verb meanings are encoded as "positive" in the dataset, which shows a 6.61% positive polarity, along with 6 411 verb meanings encoding "negative" polarity, which constitutes 21.44%. The remaining 1,52% of the verbs (456 entries) show multiple coding for polarity.

Taking these multiple labelling cases into account, we have the following distribution of polarity in the database: 21 334 instances of "neutral" label (70.21%), 2 247 instances of "positive" polarity (7.40%), and 6 804 instances of "negative" polarity (22.39%). Comparing these ratios to the previous ones, we see no significant difference in the values of the two accounts, as shown in *Table 2* below. The cases of multiple coding do not eventuate significant differences.

Table 2. Summary of the distribution of polarity in the data

Category labelling			**Multiple labelling**	**Total**	**Category labelling including cases of multiple labelling**			
"neutral"	**"positive"**	**"negative"**			**"neutral"**	**"positive"**	**"negative"**	**Total**
21 061	1 977	6 411	456	29 905	21 334	2 247	6 804	30 385
70.43%	6.61%	21.44%	1,52%	100%	70.21%	7.40%	22.39%	100%

Overall, the data suggests that in the second largest Hungarian explicatory dictionary, approximately 70% of the verb meanings express neutral events, about 7% of the data encode positive polarity, and approximately 22% encode negative polarity. These ratios seem to support the old claim that negative polarity has a richer lexical representativity than positive polarity, at least in the verbal domain. A similar analysis on the entire lexicon of the language and on the adjectival domain would shed light on the correctness of the supposition of richer representativity of negative polarity, but the findings are in line with the cognitive-affective phenomenon of negativity bias which is "an automatic tendency to pay significantly more attention to unpleasant than pleasant information" (Jing-Schmidt 2012: 417). While positive polarity is more frequent (Boucher and Osgood 1969; Garcia, Garas and Schweitzer 2012, and others), negative polarity is more richly expressed in language.

Moving on to a more detailed look, we can see that approximately 30% of verb meanings encode polarity, positive or negative (or both, in a few cases). Interestingly, but not surprisingly, the vast majority of these verbs refer to some kind of activity performed by a person, while only 3 verbs express perception. *Table 3* summarizes the occurrence of the different (more general) event types among the verbs encoding polarity. Naturally, these proportions can only have full explanatory power if we compare them to the occurrence of event types in general. This is why we can see this in the last columns of Table 3.

Table 3. The proportion of event types among the verb meanings expressing polarity and among all data

Event type	Number of occurrences among the verb meanings expressing polarity	Percentage among the verb meanings expressing polarity	Number of occurrences in the data	Percentage in the data
ACTION/ACTIVITY	5 560	62.87%	20 751	69.39%
HAPPENING	1 710	19.34%	5 248	17.55%
MOVEMENT	578	6.54%	2 243	7.50%
STATE	520	5.88%	723	2.42%
BEHAVIOR	331	3.74%	367	1.23%
ATTRIBUTE/CHARACTERISTIC	111	1.26%	383	1.28%
EXISTENCE	31	0.35%	133	0.44%
PERCEPTION	3	0.03%	57	0.19%

Looking at the third and fifth columns of *Table 2*, we can see that state and behavior are amply represented among the meanings comprising polarity, while perception verbs are highly underrepresented. A total of 71.92% of state verbs and 90.19% of behavior verbs encode polarity. ACTION/ACTIVITY, HAPPENING, MOVEMENT, ATTRIBUTE/CHARACTERISTIC and EXISTENCE are encoded with a polarity of 23-32%. This distribution of polarity in the verbal semantic system reflects the affective foundations of language on the one hand, and its fundamental social functions on the other hand. STATES and BEHAVIORS are two semantic territories that are particularly highly loaded with affective and evaluative components.

It is even more interesting to see how positive and negative polarity are distributed among the event types. ATTRIBUTES/CHARACTERISTICS is the only event type that shows a higher proportion of positive polarity, 18.81% against 10.05%.[8] In all other event types, we see a strong inclination toward the encoding of negative polarity. In behavior verbs, we can see a ratio of almost one to six, while with other event types (excepting perception, where we have very few cases of polarity[9]), we have a ratio of one to three or one to four. These ratios can be tied to how the lexicon is shaped to our affective, social and communicative needs.

[8] The reason for this might lie in the specificity of the event typology. An attribute/characteristic is usually not considered a separate event type. However, Hungarian verbs can express features such as having a particular color, shape or other specific attribute in verbal form, and I considered these to be different from state verbs. They are also stative verbs, but the feature encoded is generally external and many of them cannot be gradual, contrary to most state verbs. If we were to not distinguish between the two event types, but take them as one, negative polarity would show a much higher encoding rate (68.26%) than positive polarity (32.79%) overall. This result suggests that the event typology needs to be refined on this point.

[9] Furthermore, perception verbs are sometimes treated as state verbs, together with verbs referring to attributes/characteristics.

Table 4. The proportion of positive and negative polarity encoded in different event types

Event types	Proportion of positive polarity encoded	Proportion of negative polarity encoded
ACTION/ACTIVITY	5.53%	20.42%
HAPPENING	7.89%	24.89%
MOVEMENT	6.23%	19.66%
STATE	13.98%	58.21%
BEHAVIOR	13.49%	76.72%
ATTRIBUTE/CHARACTERISTIC	18.81%	10.05%
EXISTENCE	5.84%	18.98%
PERCEPTION	1.75%	3.51%

To illustrate the nature of the correspondences between polarity and event type and their possible motivations, in the following chapter I present examples of verbs encoding polarity and argue for their socio-pragmatic grounding.

4.2. Polarity and its socio-pragmatic grounding

There are several verbs in Hungarian expressing negative polarity and encoding eating in general, but in a particular way. A few examples would be: *csámcsog* 'munch', *csömöszöl* 'eat greedily and improperly much, ram', *dőzsöl* 'eat immoderately at a feast', *fal* 'eat greedily', *güzmölődik* 'eat idly', *habzsol* 'eat hastily and greedily, often loudly, devour', *malackodik* 'eat improperly, staining one's clothes', *nyámmog* 'eat slowly, unwillingly', *pofázik* 'eat greedily', *bezabál* 'eat greedily, hastily, and/or immoderately much', *felzabál* 'eat up something greedily, like a gobbler'. As seen from the examples, eating too much and greedily, too fast or too slowly, and loudly or unwillingly are characteristics of eating that (in Hungarian) need to be spoken about specifically and succinctly (though not very often as a quick search in the Hungarian Gigaword Corpus shows).

By contrast, there are only a few eating verbs that encode positive polarity: *eddegél* 'eat slowly, leisurely', *eszik-iszik* 'eat and drink to one's liking', *falatozik* 'eat calmly and in good taste', *jóllakik* 'eat until satiated' and *lakmározik* 'eat copiously and in good taste'. As shown in the explanations, these meanings refer to situations when eating is calm, copious and generally in good taste. We should note nonetheless that there is no verb to express only eating in good taste. "Positive" eating (in general) can only be expressed with adverbial phrases of the verb *eszik* 'eat'. I believe this to be a kind of linguistic efficiency, where the expected form of the activity is expressed by the more general and the most frequently used verb (*eszik* 'eat'), expressing the default form of eating, without added components, and all important/relevant alterations are encoded in separate lexemes to make possible their rapid, brief phrasing. Faber and Mairal Usón (1997b: 25) also observe that "in language we tend to

find words for what draws our attention, and what surprises us is generally what differs from our expectations or deviated from the socially-excepted norm". This type of markedness (and unmarkedness) fits the more general linguistic view of semantic markedness. This is related to human perception and the conceptual processing of linguistic expressions, and the degree of markedness depends on the ease of their mental processing (Ladányi 2017: 520). Although not enumerated among the cases of semantic markedness, the default form of events as a case of semantic unmarkedness easily fits the list, and it might have both universal linguistic and typological motivations. Semantic markedness of events does not necessarily motivate the morphological structure of linguistic expressions (cf. Ladányi 2017: 521), but are basically expressed in entrenchment and lexical richness. Semantically unmarked forms of events are more entrenched and lexically less rich, while semantically marked forms of events are less entrenched and have a richer lexical representation. However, there are cases where semantic markedness of events in lexical elements is expressed in the morphological structure also. See the example of verbs with a negative suffix expressing scarcity below.

In this view of semantically marked forms of events, besides affective content and evaluative meaning, although strongly connected to these, negative polarity verbs encode non-preferred alterations to the event, while positive polarity verbs encode preferred alterations or outstanding forms of the event. For an event to be outstanding in some respect, is relevant in fewer contexts, than excelling in non-preferred ways. Therefore, the latter cases are more often lexicalized than the former cases, and the expected form is expressed by a verb that is basic category-like. Thus, the richer lexical representation of negative polarity in the verbal domain is motivated and, in some cases, can be explained not by affective or evaluative content per se, but by socio-cultural factors, especially expectations regarding the typical and preferred features of events. This is also the case with behavior verbs, but also speaking verbs, verbs expressing movement, mental activity and others.

In the view of the negativity bias, the richer lexical representation of the negative experiences might be motivated on the neural level of attention allocation (Jing-Schmidt 2012: 423). The negativity bias refers to "an automatic tendency to pay significantly more attention to unpleasant than pleasant information" (Jing-Schmidt 2012: 417). It has been observed on many areas of life that "negative events have a greater impact on people's behaviour than positive events" (Jing-Schmidt 2012: 418). This goes in line with our result regarding the richer lexical representation of the negative experiences, but we must notice that not all negative polarity words refer to experiences that would necessarily frighten us. Just think about eating improperly or greedily. It does not actually refer to an activity which one would avoid at all costs. I believe, this is where a view of negative polarity as alteration to the more common or

preferred way of events can help complete the picture. Not incidentally, this view can be easily reconciled with the negativity bias. If negative is more important in terms of behavior, because "bad signals the need for change" (Jing-Schmidt 2012: 419), it might also be useful to have a richer lexical representation for it, and thus to be able to briefly and succinctly talk about it if needed. Moreover, Jing-Schmidt (2012: 429) also notes that in the psychology of fear, novelty or strangeness is considered one of the natural clues of fear. In parallel to this, familiar and common experiences are not related to strong emotions, but in turn, they are "more entrenched and more expected in verbal communication" (Jing-Schmidt 2012: 434). Naturally, I do not think that an unusual way of eating is something to be feared, but in a broader view of language use, the example of eating verbs falls in line with emotion words both in terms of their relation to the familiarity or novelty/strangeness of experiences and in terms of their morphological features. This latter strengthens the fact that polarity should be regarded as a wider affective-cognitive-social category with some linguistic expressions showing stronger and others showing weaker polarity, and/or stronger or weaker intensity, with some units not even related to emotions, but to expected/not expected forms of events.

The morphological structure of some Hungarian verbs also makes a case for polarity as an alteration to the default form of events. These are derived verbs in the dataset, most of which express negative polarity with a positive or neutral polarity stem and a negative polarity suffix. These verbs (and many others) refer to some kind of inadequate or unfit social behavior, while having a negative suffix expressing scarcity: *-talan/-telen, -(a)tlan/-(e)tlen*. That is, they have the following structure: positive/neutral stem + negative suffix + verbal suffix (e.g., *bizalm-**atlan**-kodik* 'distrust', from the stem *bizalom* meaning 'trust', *enged-**etlen**-kedik* 'disobey', from the stem *enged* meaning 'allow', *ill-**etlen**-kedik* 'behave inappropriately', from the stem *illő* meaning 'appropriate', *szemérm-**etlen**-kedik* 'be/act shameful', from the stem *szemérem* meaning 'modesty', *udvari-**atlan**-kodik* 'be/act impolite', from the stem *udvarias* meaning 'polite').

Moreover, there are no verbs in Hungarian to express the positive equivalent of these actions/behaviors. Instead, besides expressing scarcity, verbs can also refer to the exaggeration of an action, again with a negative connotation, even without a negative suffix. This can be illustrated by *bizalm-as-kodik* 'take liberties with someone' (from the stem *bizalom* 'trust') and *szemérm-es-kedik* 'being/acting too shy' (from the stem *szemérem* 'modesty') among others. Here, neither the stem, nor the suffix encode negative polarity. What is more, both verbs also have a form with a positive stem + negative suffix + verbal suffix structure (see *bizalm-atlan-kodik* and *szemérm-etlen-kedik* above). So that there is a verb to express the lack of trust/shame, and there is a verb to express the

abundance of these, but there is no verb to express the adequate amount of trust/shame in someone's behavior.[10]

Some other examples to support the above claim, would be verbs which have polarity words in their dictionary entry, but would not be generally considered encoding affective or evaluative content: *normalizálódik* 'the situation settles down, gets back to **normal**$_{[+]}$**'** (positive polarity), *stagnál* 'to stagnate, **not to evolve**$_{[-]}$' (negative polarity), *különcködik* 'to behave **eccentrically**$_{[-]}$' (negative polarity), *támogat* '**to support**$_{[+]}$ financially, morally' (positive polarity), *szövegel* 'to talk a lot or **unnecessarily**$_{[-]}$' (negative polarity), *beskatulyáz* 'to **hastily**$_{[-]}$ or **forcefully**$_{[-]}$ label someone/something' (negative polarity), etc. In the view of lexicalizing alterations to the default form of experiences, the polarity of these verbs becomes transparent and its motivations clear. Naturally, polarity is a scale, and these verbs do not show strong polarity or high intensity. Nonetheless, they do fit the case of polarity as an alteration to the default form of events. Behavior that adheres to societal norms, which does not surprise, is not something that needs to be remarked upon. While uncommon, unexpected manifestations, demeanors, manners and attitudes are experiences that attract attention, and at times need to be spoken about, preferably with precision, but succinctly.

We can see a similar case with sound symbolic-verbs. They can be considered marked expressions in terms of structural iconicity (Ladányi 2017: 524–526), and they refer to unfamiliar, uncommon, even odd forms of activities (see for example *boggle, bumble, lumber, totter, trudge*).

Some adjectives also underpin the claim that negative polarity is related to alterations to the default (familiar/expected/preferred) form of events. These do not refer to negative affect or attitude, but their structure and meaning are analogical to that of negative emotive words. Adjectives such as *unusual, unchanging, unmarried, unexpected* are the case here. We see them in Boucher and Osgood (1969), who analyze positive and negative evaluative words in their study on semantic differential. However, these adjectives have an unclear evaluation in their work (1969, 7). This is not surprising, since *unusual, unchanging* and *unmarried* are most likely considered to be neutral in terms of affective content, which is the subject of interest of their study. Nonetheless, the structure of the adjectives shows the effect of the Pollyanna hypothesis, in that negative polarity affixes are added to positive or neutral polarity stems, this being characteristic of negative polarity elements when derived from positive or neutral polarity items. I believe that the results of Boucher and Osgood

[10] Naturally, there are also verbs with a negative stem and a negative suffix expressing the presence of a negative emotion or situation. E.g., *erőszakoskodik* 'be violent' with the stem *erőszak* 'violence', *finnyáskodik* 'be fastidious' with the stem *finnyás* 'fastidious', *garázdálkodik* 'be rowdy' with the stem *garázda* 'rowdy' etc. The detailed and extensive analysis of all verbs with the *-kodik/-kedik/-ködik* suffix is not carried out here.

(Osgood 1964; Boucher and Osgood 1969) also point to the fact that socio-cultural expectations are also reflected in the structure and distribution of words encoding polarity.

Boucher and Osgood present evidence based on data from 13 languages and a large scale of positive–negative oppositions that "there is a universal human tendency to use evaluatively positive (E+) words more frequently, more diversely and facilely, than evaluatively negative (E–) words" (Boucher and Osgood 1969, 1). This is in line with the findings of the more recent study of Garcia, Garas and Schweitzer (2012). Their comprehensive analysis of 1 034 English, 2 902 German and 1 034 Spanish affective words as used in internet texts, "together with their emotional scores obtained from extensive human ratings" (2012, 3), show that words with a positive emotional content are used more often in the languages studied, and they carry less information in their semantic content in contrast to negative emotion words. Although somewhat different considerations are used in defining the terms *positive* and *negative polarity*, analyses on different subsets of linguistic elements show similar results and a higher frequency of usage of positive polarity words (see also Jing-Schmidt [2007] regarding the positivity bias).

Communicative functions of language leave their "imprint" (Diessel 2017) not only on linguistic structure in the sense of constructional patterns, but also on the overall semantic network of the lexicon. This is because experience not only determines how we talk, but also what is there to be talked about and what is worth talking about. It is not only linguistic structure that makes use of statistical information for speaking, but the occurrence of specific experiences or experience types too. The impact of these can be grasped, among others, by looking at the distribution of lexical items encoding polarity.

Perhaps our human trait that we feel comfortable in familiar situations, also pertains to the problem, in that unusual, unchanging and so on, are among the conspicuous, unexpected, perhaps unwanted situations, and so, among the experiences relevant to be talked about. In terms of the human psyche, what is frequent is good, because is known, and thus one knows what to expect. In this way, words expressing experiences known to us (most likely because they are frequent, common in our everyday experiences), fall in the same category as words expressing positive experiences. (Unless, of course, if they are really dangerous or frightening. Those could not be considered positive, even after becoming familiar with them.)

In the light of the above enumerated examples, I believe we can conclude that alterations to the default mode of experiences are an important organizational principle in the lexicon in general, and it should be considered a part of linguistic polarity phenomena along with affective-emotive content. This is substantiated by structural and semantic features of the linguistic units analysed in the paper, which all fit with the theory of semantic markedness (cf.

Ladányi 2017) and the negativity bias pattern (cf. Jing-Schmidt 2007, Kugler 2014). In this light, the definition of negative polarity should be refined by adding non-preferred or not expected alterations to the default forms of events, and positive polarity by adding preferred or outstanding alterations to its definition. By *default*, here we mean that it is familiar, considered appropriate, suitable, or fitting the socio-cultural norms and subjective expectations of the speaker. This is mainly expressed by the most frequent basic level-like verb for a certain event.[11]

Thus, in verbal meaning (but most certainly in other categories too) 'positive' and 'negative' should be understood not only in terms of positive and negative emotive content, approved or disapproved behavior, positive or negative moral or subjective evaluation of things and experiences (cf. Martin and White 2005), but more broadly as compliance to the preferred and non-preferred form of experiences.

After having a wide-spread look at polarity in verbal meanings, in the next chapter I touch upon the question whether the semantic load of polarity verbs shows correlations with word length, as surmised by the Pollyanna hypothesis and Zipf's law.

4.3. Polarity and word length

As a sequel to the finding that in the verbal system negative polarity is lexically more richly represented in the lexicon than positive polarity, I also used the dataset to test whether negative polarity verbs are longer on average than positive polarity verbs. This is a follow-up to the Pollyanna hypothesis and Zipf's law, the first stating that evaluatively positive terms are used more often and more diversely than evaluatively negative words, and the second stating that words used more frequently are shorter in length and carry less information (Boucher and Osgood 1969; Zipf 1936, 1949). We can infer from these two that evaluatively negative words carry more information and hence are longer in average than evaluatively positive words.

I have used the dataset to make a preliminary assessment of the possible correlations by counting word length in two ways: based on the constituent sounds and based on the constituent syllables.[12] However, no significant difference showed up between the two methods. I only present them both below for the sake of accuracy.

---

[11] Cf. Faber and Mairal Usón: "Although semantic parameters codifying sociocultural norms are different…, it is relatively easy to deduce what such norms are, because they are precisely what is not lexicalized" (1997b: 25, highlight in the original).

[12] Consonants denoted with letter combinations in Hungarian (*cs, dz, gy, ly, ny, sz, ty, zs* etc) were treated as one sound.

The results show no substantial difference in word length between positive and negative polarity verbs, but the ANOVA tests[13] suggest that the correspondences of word length and polarity is not random in the data (F(2, 29902) = 102,2402, p = 0,0000 for the correspondences of polarity and word length counted with constituent sounds, and F(2, 29902) = 83,7886, p = 0,0000 for the correspondences of polarity and word length counted with constituent syllables, respectively).

There is a slight variation with regard to polarity, with evaluatively negative verbs being on average longer than evaluatively neutral verbs, but they are also shorter than evaluatively positive verbs. The table below presents the exact numbers, and the figures show the distribution of the datapoints in a means plot.

Table 5. Average word length in the data

	Average word length counted with constituent sounds	**Average word length counted with constituent syllables**
Evaluatively negative verbs	7.76	3.07
Evaluatively positive verbs	7.83	3.12
Evaluatively neutral verbs	7.36	2.92

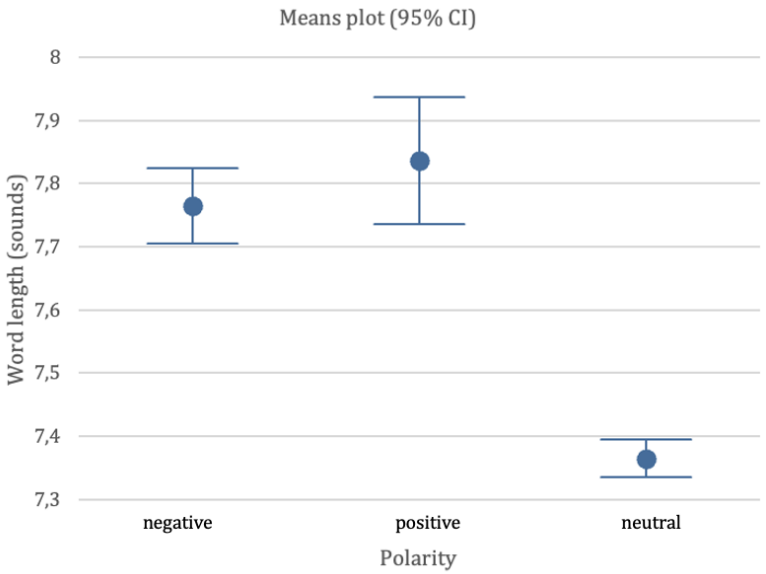

Figure 1. Distribution of polarity and word length (constituent sounds)

[13] One-way ANOVA using Microsoft Excel's StatPlus add-in. I have tested for the correlation between polarity and word length using constituent sounds and syllables too, with similar results.

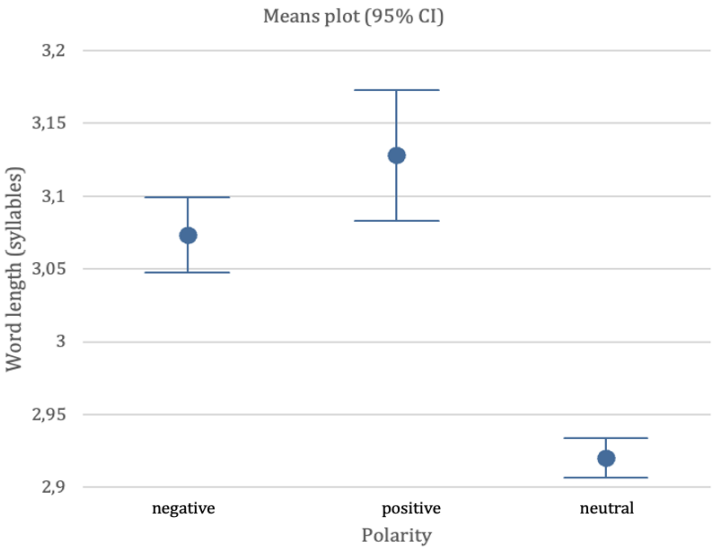

Figure 2. Distribution of polarity and word length (constituent syllables)

We must note first of all, that positive polarity verbs are somewhat longer on average than negative polarity verbs, which challenges the hypothesis that negative polarity verbs are longer. Nonetheless, neutral verbs are shorter than verbs with more semantic load (positive and negative polarity). Thus, we can conclude that in the view of the verbal lexicon of Hungarian, shorter words possibly also carry less information. This statement parallels Zipf's law, according to which words used more frequently are shorter in length and carry less information, although on the basis of the present dataset we cannot make inferences regarding frequency of use, and that is not the aim here.

Naturally, it might also be the case that when looking at the lexicon of a language, there is no difference in word length between positive and negative polarity verbs, because in both cases there is a semantic overload compared to non-polarity verbs. This way, we should see a difference only between polarity verbs and non-polarity verbs – which the present analysis suggests. Taking into consideration that the omega squares are 0,0067 and 0,0055 respectively (suggesting that only 0,06 and 0,05 percent of the variation among the length of the verbs is explained by polarity), it seems obvious that there is more to the relation of word length and polarity, given that polarity is not the only linguistic feature that might be correlated with word length, and so the correspondence between the two factors needs further investigation. This claim is corroborated by post hoc tests, which almost all show correlation between positive and

neutral, and negative and neutral categories. These correlations are not discussed here, because of the nature of the study, but descriptive statistics can be found in Appendix 1 for information. The correspondences of word length and polarity will be studied in more depth in the future.

Naturally, we must also take into consideration that perhaps the data is not representative, and with the use of more annotators the distribution of verbs encoding polarity would be somewhat different. This could lead to different correlations. In all cases further analyses are required to test the correlations between word length and polarity. The preliminary study shows most importantly that correspondences regarding word length and polarity in the view of the lexicon (as opposed to frequency studies) must also be tested on a somewhat different dataset: polarity verbs versus non-polarity verbs. Also, other factors must be considered in the investigation of this problem, which could be correlated with word length.

5. Conclusions

In the paper I presented the preliminary results and the important considerations of a larger study regarding the distribution of polarity in the verbal domain of Hungarian, with possibilities of generalizing the results to other domains of the language and other languages. The study of the problem is motivated by the long-standing hypothesis that negative polarity has a richer lexical representativity than positive polarity on a lexical, constructional and idiomatic level as well. To be able to test this, I have created a database of 29 905 Hungarian verb meanings and coded them for polarity, using one annotator (up to this moment) and a three-category typology: "positive", "negative" and "neutral".

The preliminary results of the investigation show first of all that, in the verbal domain, indeed negative polarity is lexically more diverse than positive polarity. Naturally, this correspondence should also be tested on data from more annotators, and this is planned in the future, but we should note that other scholars have also observed this distribution among verbs in other languages. Faber and Mairal Usón (1997b: 25) write that "in language we tend to find words for what draws our attention, and what surprises us is generally what differs from our expectations or deviated from the socially-accepted norm"; and that "the opposition of *good* and *bad* consistently appears as a structuring device within the semantic domains studied" (1997b: 30, highlights in the original). Moreover, these preliminary results fall in line with other linguistic and affective-cognitive patterns, which are worth mentioning again.

Based on the distribution of polarity among the event types and on the example of Hungarian verbs derived with a negative suffix expressing scarcity

(*-talan/-telen, -(a)tlan/-(e)tlen*) and on adjectives derived with a negative suffix, but usually not considered to have negative polarity (*unusual, unchanging, unmarried, unexpected*), I have argued that besides affective content and evaluative meaning, the definition of polarity should also include the concept of default events, where negative polarity verbs encode non-preferred alterations to the event, while positive polarity verbs encode preferred alterations or outstanding forms of the event. For an event to be outstanding in some respect, is relevant in fewer contexts, than excelling in non-preferred ways. Therefore, the latter cases are more often lexicalized than the former cases, and the expected form is expressed by a verb that is basic category-like. Thus, the richer lexical representation of negative polarity in the verbal domain is motivated and explained not only by affective or evaluative content, but also by socio-cultural factors, especially expectations regarding the typical and preferred features of events. Naturally, the negativity bias has its imprint on the expression of negative polarity in language. But besides the fact that negative experiences have a greater impact on us than positive events, novelty or strangeness are also considered natural clues of fear (Jing-Schmidt 2012: 418, 434). Obviously, from an evolutionary and a neural perspective the negativity bias is the first and strongest motivation for the richer lexical representation of negative polarity in the affective field, but the factor of novelty/strangeness as an important trigger can easily be transferred to social experiences, and this is what might have happened.

I believe that this view is corroborated by the example of eating verbs and adjectives presented in the paper, which fall in line with emotion words both in terms of their relation to the familiarity/novelty of experiences and in terms of their morphological features. Moreover, this approach fits with the view of semantic markedness (Ladányi 2017), in the sense that semantically unmarked forms of events are more entrenched and lexically less rich, while semantically marked forms of events are less entrenched and have a richer lexical representation. Alterations to the default mode of experiences might be an important organizational principle in the lexicon, its ties with polarity in structural and semantic features being only one example of its manifestation.

Another intriguing problem investigated in the paper is the correspondence of polarity and word length. If more frequent words are shorter in length (Zipf 1936, 1949) and positive polarity words are used more frequently (Boucher and Osgood 1969), than negative polarity verbs should be longer on average than positive polarity verbs. The analysis did not confirm this hypothesis, but made clear an important aspect in the investigation of polarity distribution in the entire verbal domain (as opposed to usage data), namely that the dataset should also be able to deal with polarity verbs as one group as opposed to non-polarity verbs. It might be the case that the difference in word

length cannot be found between positive and negative polarity verbs, but among polarity and non-polarity verbs when looking at the lexicon of a language. Naturally, it might also be that this distribution is not characteristic to the lexicon of a language, but can be captured only in usage data with regard to the frequency of positive polarity expressions.

Based on these results, I believe that polarity should be regarded as a wider affective-cognitive-social category with some linguistic expressions showing stronger and others showing weaker polarity, and/or stronger or weaker intensity, with some expressions not even related to emotions, but to expected/not expected forms of events.

The research will continue based on these considerations stemming from important realizations regarding diverse and far-reaching connections of polarity in the lexicon, with a special attention to the structuring property of polarity in the lexicon, of which the details still need to be revealed.

WORKS CITED

Bárczi Géza, and Országh László (eds.). 1959–1962. *A magyar nyelv értelmező szótára [Hungarian Explicatory Dictionary]*. Budapest: Akadémiai Kiadó.

Boucher, Jerry, and Charles E. Osgood. 1969. "The Pollyanna Hypothesis." *Journal of verbal learning and verbal behavior* 8: 1–8.

Bybee, Joan L., and Clay Beckner. 2009. "Usage-based theory." In *The Oxford Handbook of Linguistic Analysis*, edited by Narrog, Heiko, and Bernd Heine, 827–55. Oxford: Oxford University Press. https://www.unm.edu/~jbybee/downloads/BybeeBeckner2010UsageBasedTheory.pdf

Cambria, Erik, and Amir Hussain. 2015. *Sentic Computing: A Common-Sense-Based Framework for Concept-Level Sentiment Analysis*. Heidelberg–New York–Dordrecht–London: Springer. https://sentic.net/sentic-computing.pdf

Diessel, Holger. 2017. „Usage-based linguistics." In *Oxford Research Encyclopedia of Linguistics*, edited by Aronoff, Mark, 1–26. New York: Oxford University Press.

Faber, Pamela B., and Ricardo Mairal Usón. 1997a. "The paradigmatic and syntagmatic structure of the semantic field of existence in the elaboration of a semantic macronet." *Studies in language* 21, no. 1: 129–67.

Faber, Pamela B., and Ricardo Mairal Usón. 1997b. "Towards a typology of predicate schemata in a Functional-Lexematic Model." In *Towards a Functional Lexicology*, edited by Wotjak, Gerd, 11–37. Frankfurt: Peter Lang.

Faber, Pamela B., and Ricardo Mairal Usón. 1999. *Constructing a Lexicon of English Verbs*. Berlin-New York: Mouton de Gruyter.

Garcia, David, Antonios Garas, and Frank Schweitzer. 2012. "Positive words carry less information than negative words." *EPJ Data Science* 1, no. 3: 1–12. https://epjdatascience.springeropen.com/articles/10.1140/epjds3

Hua, Yan Cathy, Paul Denny, Jörg Wicker, and Katerina Taskova. 2024. "A systematic review of aspect-based sentiment analysis: Domains, methods, and trends." *Artificial Intelligence Review* 57: 296. https://link.springer.com/article/10.1007/s10462-024-10906-z

Giannakidou, Anastasia. 2011. "Negative and positive polarity items: Variation, licensing, and compositionality." In *Semantics: An international handbook of natural language meaning*, edited by Maienborn, Claudia, Klaus von Heusinger, and Paul Portner, 1660–712. Berlin: Mouton de Gruyter.

Jing-Schmidt, Zhuo 2007. "Negativity bias in language: A cognitive-affective model of emotive intensifiers." *Cognitive Linguistics* 18, no. 3: 417-443.

Mingorance, Martín L. 1990. "Functional grammar and lexematics." In *Meaning and Lexicography*, edited by Tomaszczyk, Jerzy, and Barbara Lewandowska-Tomaszczyk, 227–86. Amsterdam/Philadelphia: John Benjamins.

Martin, Jennifer R. and Peter Rupert Robert White. 2005. *The language of evaluation. Appraisal in English*. New York: Palgrave Macmillan.

Kugler Nóra 2014. "A nyelvi polaritás kifejezésének egy mintázata, avagy milyen a *félelmetesen jó*?" *Magyar Nyelvőr* 138, no. 2: 129-139.

Ladányi Mária. 2017. "Alaktan." In *Nyelvtan [Grammar]*, edited by Tolcsvai Nagy Gábor, 503-660, Budapest: Osiris.

Langacker, Ronald W. 2008. *Cognitive grammar. A basic introduction*. New York: Oxford University Press.

Oravecz Csaba, Váradi Tamás, and Sass Bálint. 2014. The Hungarian Gigaword Corpus. In: *Proceedings of LREC 2014*.

Osgood, Charles E. 1964. "Semantic differential technique in the comparative study of cultures." *American Anthropology* 66: 171–200.

Princeton University. 2010. *About WordNet*. WordNet. Princeton University. https://wordnet.princeton.edu/

Pusztai Ferenc. 2006. *Magyar értelmező kéziszótár [Hungarian explicatory dictionary]*. 2nd, extended edn. Budapest: Akadémiai Kiadó.

Szabó Martina, Vincze Veronika, and Bibok Károly 2023. "Thank you for the terrific party!" – An analysis of Hungarian negative emotive words. *Corpus Linguistics and Linguistic Theory* 19, no. 3: 451-485.

Szilágyi N., Sándor, and Dimény Hajnalka. 2021. *Magyar igék szemantikai adatbázisa [Semantic database of Hungarian Verbs]*. Cluj-Napoca: Babeș-Bolyai University.

Talmy, Leonard. 2000. *Toward a cognitive semantics I. Concept structuring systems*. London: MIT Press.

Tolcsvai Nagy Gábor. 2016. "Van: változatok a létezés nyelvi konstruálására [Van 'there is': Variations of the linguistic construal of existence]." *Magyar Nyelvőr* 140, no. 1: 7–22.

Wierzbicka, Anna. 1999. *Emotions across languages and cultures*. New York: Cambridge University Press.

Windisch Brown, Susan, Claire Bonial, Leo Obrst, and Martha Palmer. 2017. "The Rich Event Ontology." In *Proceedings of the Events and Stories in the News Workshop*, edited by Caselli, Tommaso, Ben Miller, Marieke van Erp, Piek Vossen, Martha Palmer, Eduard Hovy, Teruko Mitamura, and David Caswell, 87–97. Vancouver: Association for Computational Linguistics. https://aclanthology.org/W17-2712/

Zipf, George Kingsley. 1936. *The Psychobiology of Language*. London: Routledge.
Zipf, George Kingsley. 1949. *Human Behavior and the Principle of Least Effort*. New York: Addison-Wesley Press.

Appendix 1.

Descriptive statistics of the ANOVA analyses.

Comparisons among groups (Factor 1 - Polarity) **Polarity vs. word length counted with constituent sounds**					
Scheffe					
Group vs. Group (Contrast)	*Difference*	*95% Confidence Interval*		*Test Statistic*	*p-value*
Neg vs Pos	-0,0717	-0,2113	0,0679	1,2568	0,4540
Neg vs Neu	0,4000	0,3200	0,4799	12,2451	0,0000
Pos vs Neu	0,4717	0,3448	0,5985	9,1029	0,0000
Tukey-Kramer					
Groups	*Difference*	*Test Statistic*	*p-value*	*Significant*	
Neg vs Pos	-0,0717	1,7774	0,8527	*No*	
Neg vs Neu	0,4000	17,3172	0,7460	*No*	
Pos vs Neu	0,4717	12,8734	0,7460	*No*	
Bonferroni					
Alpha/N	0,0167				
Group vs. Group (Contrast)	*Difference*	*95% Confidence Interval*		*Test Statistic*	*p-value*
Neg vs Pos	-0,0717	-0,2083	0,0649	1,2568	0,6265
Neg vs Neu	0,4000	0,3218	0,4782	12,2451	0,0000
Pos vs Neu	0,4717	0,3476	0,5957	9,1029	0,0000
Fisher LSD					
Group vs. Group (Contrast)	*Difference*	*Test Statistic*	*p-value*	*Significant*	
Neg vs Pos	-0,0717	1,2568	0,2088	*No*	
Neg vs Neu	0,4000	12,2451	0,0000	*Yes*	
Neg vs Neu	0,4717	9,1029	0,0000	*Yes*	
Dunnett's test (Control Group = 1)					
Group vs. Group (Contrast)	*Difference*	*Test Statistic*	*p-value*	*Significant*	
Pos vs Neg	0,0717	1,2568	0,3534	*No*	
Neu vs Neg	-0,4000	12,2451	-0,0016	*Yes*	

Comparisons among groups (Factor 1 - Polarity) **Polarity vs. word length counted with constituent syllables**					
Scheffe					
Group vs. Group (Contrast)	*Difference*	*95% Confidence Interval*		*Test Statistic*	*p-value*
Neg vs Pos	-0,0547	-0,1169	0,0075	2,1544	0,0982
Neg vs Neu	0,1530	0,1174	0,1887	10,5215	0,0000
Pos vs Neu	0,2078	0,1513	0,2643	9,0047	0,0000
Tukey-Kramer					
Groups	*Difference*	*Test Statistic*	*p-value*	*Significant*	
Neg vs Pos	-0,0547	3,0468	0,7661	*No*	
Neg vs Neu	0,1530	14,8797	0,7460	*No*	
Pos vs Neu	0,2078	12,7345	0,7460	*No*	
Bonferroni					
Alpha/N	0,0167				
Group vs. Group (Contrast)	*Difference*	*95% Confidence Interval*		*Test Statistic*	*p-value*
Neg vs Pos	-0,0547	-0,1156	0,0061	2,1544	0,0936
Neg vs Neu	0,1530	0,1182	0,1879	10,5215	0,0000
Pos vs Neu	0,2078	0,1525	0,2630	9,0047	0,0000
Fisher LSD					
Group vs. Group (Contrast)	*Difference*	*Test Statistic*	*p-value*	*Significant*	
Neg vs Pos	-0,0547	2,1544	0,0312	*No*	
Neg vs Neu	0,1530	10,5215	0,0000	*Yes*	
Neg vs Neu	0,2078	9,0047	0,0000	*Yes*	
Dunnett's test (Control Group = 1)					
Group vs. Group (Contrast)	*Difference*	*Test Statistic*	*p-value*	*Significant*	
Pos vs Neg	0,0547	2,1544	0,0585	*No*	
Neu vs Neg	-0,1530	10,5215	-0,0012	*Yes*	