On the Use of Linguistic Intensifiers in Schizophrenia, Schizoaffective and Bipolar Disorders

Author One[1*] Author Two[2] Author Three[1]

[1]Organization 1
{auth1, auth3}@org1.edu
[2]Organization 2
auth3@inst2.edu

Abstract

In this paper, we analyse the phenomenon of intensification in spontaneous speech of Hungarian patients suffering from schizophrenia (SZ), schizoaffective (SAD) and bipolar disorders (BD), using NLP tools and methods. Since mental health influences the mode of human communication, the analysis of texts produced by people with mental illness has received increasing attention in recent years. The acquisition and processing of linguistic data (spoken or written) provides an opportunity to reveal interrelations between linguistic factors and psychological aspects. In the present study, we focused on linguistic intensification among the linguistic parameters. To the best of our knowledge, an international study that comparatively scrutinizes intensification in the case of these mental illnesses has not been conducted so far. However, we are aware of that emotion regulation dysfunction is characteristic of psychotic disorders (Chapman et al., 2020; Kring, Elis, et al., 2013) and some research finding also lead us to the conclusion that the use of intensifiers is closely related to emotion regulation (Athanasiadou, 2007; Strous et al., 2009). Relying on these connections between mental illnesses and emotion regulation, within the group of intensifiers, we pay special attention to the so-called negative emotive intensifiers, which are those linguistic intensifiers that have a prior negative emotive semantic content, e.g. *awfully*, *terribly*, etc. One of the main questions of the current study is whether there is a notable and systematic difference in the use of intensifiers among the examined patients and the control group, which may be used as a predictive feature in the future. Our results show that although there is no notable difference in the use of the standard register intensifiers, the specific group of negative emotive intensifiers has strikingly different characteristics among the speakers.

Keywords: Schizophrenia, Schizoaffective disorder, Bipolar disorder, Schizophrenia-bipolar spectrum disorders, Intensifiers, Negative emotive intensifiers, Sentiment, Higher order language functions

1 Introduction

Bipolar disorder (henceforth BD (previously called manic-depressive illness) is a recurrent chronic disorder characterized by episodes of mania, hypomania, and alternating or intertwining episodes of depression (Grunze, 2015). Schizophrenia (henceforth SZ) is a chronic mental health disorder characterized by symptoms of delusions, hallucinations, disorganized speech or behavior, as well as impaired cognitive ability (Patel, Cherian, Gohil, & Atkinson, 2014). In the case of schizoaffective disorder (henceforth SAD), a person has mixed psychotic (hallucinations or delusions) and affective symptoms (mood episodes) (Malhi, Green, Fagiolini, Peselow, & Kumari, 2008; Rose, 2014), hence it occupies an intermediate position between the two diseases in the schizophrenia-bipolar spectrum considered in a dimensional approach (Peralta & Cuesta, 2008). Cognitive impairment is a hallmark symptom of psychotic disorders including SZ and SAD, spanning verbal and non-verbal abilities (Little et al., 2019; Van Rheenen et al., 2016). Recent research findings indicate that patients with BD also have significant impairments in cognitive functioning (Van Rheenen et al., 2016). However, this evidence may not be always quantitatively comparable to those seen in psychotic illnesses (Burdick, Ketter, Goldberg, & Calabrese, 2015; Van Rheenen & Rossell, 2014). Accurate diagnosis of BD is also difficult in clinical practice because in the depressive

*Author One was partially supported by Grant XXX

episode it resembles unipolar depression (Grunze, 2015). Moreover, no diagnostic biomarkers of BD have yet been validated for clinical use (Göteson et al., 2022).

Susceptibility to psychosis spectrum disorders, including schizophrenia-bipolar spectrum, is genetically determined. These diseases usually manifest themselves during the reproductive phase of life. Among them, schizophrenia in particular leads to a significant decrease in fertility (53% for women, 77% for men) (Power et al., 2013). This reproductive disadvantage should lead to the rapid elimination of the given genes from the human genome. Conversely, cross-cultural constancy indicates that co-inheritance with some species-significant genetic variant may be the counterbalancing advantage for the entire population (Crow, 1993, 1995, 1997). The most biologically determined characteristic of our species is language (Crow, 1996), and language separates modern humans from earlier hominids (Bickerton, 1990; Chomsky, 1986; Dronkers, Pinke, & Damasio, 2000; Ganger & Wold, 1998; Trask, 1999). During hominid evolution, the development of brain structural asymmetries is responsible for the development of human-specific components of language (Corballis, 2009; Geschwind & Galaburda, 1985), and these asymmetries are impaired in schizophrenia (Crow, 1998). According to a rather parsimonious conclusion, psychosis and language are related to genetic variations linked to the origin of the species (Berlim, Mattevi, Belmonte-de Abreu, & Crow, 2003). This is why our research group has been extensively investigating the relationship between schizophrenia-spectrum diseases and language behavior.

In our current research project we chose to analyse and compare the speech of Hungarian SZ, SAD and BD patients. For the recent analysis we collected and transcribed 526 recorded texts all together among the subject groups and controls. Text were produced during three directed spontaneous speech tasks when the subjects were asked to talk about given topics. Our written corpus contains 158,386 tokens all together. By examining the transcripts of spoken texts in our research with NLP methods and tools, we tried to ascertain whether there is any linguistic difference in SZ, SAD and BD patients and the control group.

Here, among the linguistic features, we focus on a single phenomenon, that of linguistic intensification, combining our examination with a sentiment analysis. Our choice to focus on these features is justified by the following points: Schizophrenic patients show deficits in emotion perception and expression, and this brain activity related to emotion is not the same as in control subjects (Kring et al., 2013). Moreover, abnormally elevated mood in BD disorder is associated with specific neurocognitive deficits consistent with neuropathology in neural networks that are critical for emotion regulation (Green, Cahill, & Malhi, 2007). Therefore, we argue that sentiment analysis may reveal relevant results in this corpus. As for linguistic intensification, recent research findings also lead us to conclude that the use of intensifiers is closely related to emotion regulation (Athanasiadou, 2007; Strous et al., 2009). What is more, according to Athanasiadou (2007) intensifiers are linguistic markers of speaker subjectivity, and they are seen as having the primary function of signifying the speaker's point of view and attitude. Because of the links between mental illness and emotion regulation, within the group of intensifiers, it is worth focusing on the so-called *negative emotive intensifiers*, whose prior semantic content is related to a negative emotion, but which can function as linguistic intensifiers. It is also worth mentioning that there is evidence in the literature that the general use of intensifiers is different in, for instance, schizophrenia (Strous et al., 2009).

In the current paper, for the sake of insight and clarity, we will refer to *negative emotive intensifiers* as NEIs. What is more, based on Kochetova and Ilyinova (2022), we refer to other, non-negative emotive intensifiers as *standard register intensifiers* called SRIs in their case. Furthermore, sometimes we will refer to the two groups with the notion of *intensifier*, applying it as a general term.

Here, we attempt to discover whether there is a notable and systematic difference in the use of intensifiers among the examined patients and the control group, which may be used as a predictive feature. Our main research questions are the following: How does each patient group use intensifiers in contrast with each other and the controls? Are there any notable qualitative or quantitative differences in the use of these intensifiers? What are the most frequently used SRIs or NEIs in each group? Is there any systematic difference in relation to the parts-of-speech of the word the analysed words modify? And lastly, is there any noteworthy difference in relation to sentiments the SRIs, as well as the NEIs occur together with, among the examined speakers groups?

The paper is structured as follows. In the first part of the next section we will introduce the current literature on linguistic intensification, with special attention to NEIs, while in the second part we will briefly summarize previous results of studies measuring the relationship between language production and mental disorders using NLP methods and tools. In section 3, we describe the methods used in the current study, including the metadata of our corpus, the steps of automatic data processing, and the means and principles of manual annotation. Next, the results of the corpus analysis are presented in

Section 4, where we give the results on SRIs and NEIs separately. In section 5, we discuss the results in greater detail, and offer some linguistic examples and possible explanations for the results. In the final part we give a summary, draw some conclusions and make some suggestions for future research.

2 Literature review

2.1 The phenomenon of linguistic intensification

Intensifiers are used to scale the degree of the semantic content of the word they modify (Chang & Bin, 2020; Méndez-Naya, 2008). Therefore, intensifiers appear with other lexical items, such as adverbs, adjectives, verbs, infrequently with nouns. The term *intensifier* and its definition are not consistently used among the authors. First of all, the term *intensifier* is often used interchangeably with the terms like *degree adverb* or *degree modifier*, see e.g. Bolinger (2013); Méndez-Naya (2008); Paradis (1998); Su (2016). For a recent review, we will use the term *intensifier* and rely on the following definition: "any device that scales a quality, whether up or down or somewhere between the two" (Bolinger, 2013).

Since intensifiers - as already mentioned above - scale up or down the a quality expressed by the word they modify, they can be classified with respect to their position on an intensity scale (Fuchs & Gut, 2016). Authors usually distinguish two separate groups called amplifiers (e.g. *very, notably*) and downtoners (e.g. *fairly, slightly*) (Strohm & Klinger, 2018). *Amplifiers* scale upwards (`[1a]`), while *downtoners* scale downwards from an assumed norm related to the semantic content of the modified head (`[1b]`).

(1) a. Highly intelligent people are adaptable and flexible. [https://medium.com/]

b. York's actually a really really good University. (Ito & Tagliamonte, 2003)

In the Hungarian language, within the group of intensifiers Balogh (2009) distinguishes those that play an intensifying role according to their primary function (e.g. *nagyon* 'very', *módfelett* 'consumedly' etc.), and those which have some primary meaning other than intensification, and have an intensifier role only as a secondary function (e.g. *határtalanul* 'infinitely', *mélyen* 'deeply', *mértéktelenül* 'immeasurably' etc.). Within the latter group, NEIs deserve special attention. These are those words which have a semantic content that may be associated with a negative emotion, but they may possess an intensifier function, such as *terribly*, *awfully* and *desperately* in constructions like *terribly nice*, *awfully good* and *desperately important* (Chang & Bin, 2020; Szabó & Bibok, 2019). As for their place in the system of intensifiers (see above), NEIs belong the former group, so they are able to scale upwards the semantic content of the constructional head they modify.

In the literature, NEWs are mainly discussed within the group of intensifiers (see e.g. Chang and Bin (2020); Jing-Schmidt (2007)). However, they can have other meanings and functions as well (Szabó & Otani, 2022). Just to mention a few, they may express even a positive evaluation of the speaker, e.g. *brutális alaplap* (lit. 'brutal motherboard' – 'high quality motherboard') or surprise, e.g. *D-23 for prints? That's crazy!*. They may also be used as an interjection, e.g. *Terrific! What a coincidence!*. At the same time, despite their more recent functions, they can still be used according to their prior meaning, e.g. the Hungarian *durva* ('coarse') can still be used according to its prior lexical meaning, e.g. *durva szemcse* ('coarse grain'). What is more, some NEI have a figurative sense as well. For instance, the above-mentioned *durva* can mean 'physically or verbally abusive', e.g. *durva férfi* ('rude man').

It is worth mentioning that these words proceed through a delexicalization process over time, during which they lose their original meaning in a step-by-step fashion (Lorenz, 2002; Nevalainen & Rissanen, 2002; Partington, 1993). Here, the negative content of the word diminishes over time. At the same time, the importance of the pragmatic role increases – where a strong subjectivation tendency can be seen; the newer use expresses the speaker's subjective, evaluative opinion at a certain level (Kugler, 2014). However, this specific emotive semantic-pragmatic content gradually decreases due to the new collocation pattern. Since NEIs are more often collocated with words carrying the opposite, i.e. positive semantic content, the affective value of NEIs is constantly fading, diminishing (Kugler, 2014). In this study we do not deal with NEIs from a semantic changing perspective, but bearing this in mind should help us to better understand some aspects of our findings here.

2.2 Previous Results

Since mental health influences the method of human communication, the acquisition and processing of linguistic data (spoken or written) provides an opportunity to uncover interrelations between linguistic factors and psychological aspects. Hence findings could be useful for identifying any impairment of mental and brain functioning and they help clarify the mental health circumstances. And recent technological advances such as Natural Language Processing (NLP) and machine learning offer new ways of examining the relationship between language production and metacognition in different types of mental disorders (Calvo, Milne, Hussain, & Christensen, 2017; Lundin, Hochheiser, Minor, Hetrick, & Lysaker, 2020). Just to mention a few, several studies have explored the possibility of utilizing acoustic features in neuroticism and depression detection (Akkaralaertsest & Yingthawornsuk, 2015; Resnik, Garron, & Resnik, 2013; Taguchi et al., 2018) or mild cognitive impairment and Alzheimer's disease (Haider, De La Fuente, & Luz, 2019; Tóth et al., 2015) or Asperger syndrome (Chaput et al., 2013).

One of the most important analytical tools for text-based studies are spoken language corpora that can be analysed using NLP methods and tools. In the last decades, several spoken language corpora supporting psycholinguistic studies have been created in several languages (e.g. (Calvo et al., 2017; Corcoran et al., 2018; Little et al., 2019)), and among them we find Hungarian language corpora as well (e.g. (Gosztolya, Bagi, Szalóki, Szendi, & Hoffmann, 2018; Kálmán et al., 2022; Vincze, Szabó, et al., 2021)). For SZ, we should mention, among others, the corpus described in Minor, Willits, Marggraf, Jones, and Lysaker (2019), Willits, Rubin, Jones, Minor, and Lysaker (2018) and Bagi, Gosztolya, Szalóki, Szendi, and Hoffmann (2019).

NLP and machine learning methods have been applied in the study of language usage and speech production in SZ, SAD and BD as well. Of these disorders, SZ has received particular attention in this research field (DeLisi, 2001). Below we mention some research studies relevant to the current comparative analysis, also focusing on linguistic features.

Lott, Guggenbühl, Schneeberger, Pulver, and Stassen (2002) use speech samples of 100 patients suffering from schizophrenia, bipolar illness and major depression. They attempt to uncover linguistic abnormalities in the speech of patients represent diagnosis-specific characteristics or constitute syndrome-like dimensions of these disorders. The majority of the linguistic variables did not prove to be statistically significant. However, when the respective variables were analysed as a multivariate entity, the variety of subtle differences allowed the authors to discriminate among the diagnostic groups. Mota, Copelli, and Ribeiro (2017) analyse connectedness, a structural feature of speech in SZ and compare the results with bipolar and control groups. As their research revealed, Disorganization Index classifies negative symptoms severity and schizophrenia diagnosis 6 months in advance. However, the Bipolar and Control groups could not be differentiated using connectedness attributes or the Disorganization Index. Voleti et al. (2019) analyse interview transcripts collected from a total of 87 clinical subjects and 22 healthy controls. Of the clinical population, 44 had been diagnosed with bipolar I disorder and 43 had been diagnosed with schizophrenia or schizoaffective disorder but they were taken together in this analysis. Their classifier based on selected language features proved to be effective in distinguishing between healthy controls and patients.

Studies focusing on the interrelation between the use of intensifiers and mental illness are scarce, but report unanimous findings. For example, Vioules, Moulahi, Azé, and Bringay (2018) found that suicidal thoughts could be predicted by a higher use of intensifiers. Weerasinghe, Morales, and Greenstadt (2019) report that individuals with depression used intensifiers more frequently. Al-Mosaiwi and Johnstone (2018) investigate the link between the use of absolutist words (and among them intensifiers) and anxiety, depression and suicidal ideation. They found that absolutist words were used significantly more frequently by individuals suffering from the above mentioned disorders than control groups. It is worth mentioning that all three studies were based on written texts collected from Twitter or Internet forums.

As for the Hungarian population, many papers concentrate on Hungarian patients suffering from SZ, BD and SAD (e.g. (Döme et al., 2005; Inczédy-Farkas et al., 2010; Kéri, Kelemen, Benedek, & Janka, 2001; Kocsis-Bogár, Nemes, & Perczel-Forintos, 2016; Réthelyi et al., 2010; Szabó et al., n.d.)). Most studies focus on the temporal aspects of texts (Bagi et al., 2019; Gosztolya et al., 2018), but research has also been conducted on the occurrence of recursion in the case of SAD (Kárpáti et al., 2018). Bagi, Hoffmann, Szendi, and Toth (2016) examines the understanding of metaphor and irony in SZ, while Simon et al. (2011) focuses on the understanding of irony in SZ and BD. Findings show that both groups perform in the tests poorer than the healthy control group, and, interestingly, BD patients showed significantly higher brain activity when processing irony. Baranyai et al. (2020) found that the executive functions of SZ patients show deficits in all dimensions when compared to healthy controls, but in the

case of SAD patients, signal detection and control functions remain intact. Szabó et al. (2023) analyse the usage of intensifiers by SZ patients, and they pay special attention to NEIs within the intensifer group. While the results do not reveal any significant differences between the SZ and the control group, SZ patients seem to use NEIs much less frequently, and more importantly, they invariably use NEIs in a negative context.

3 Analysis

3.1 The corpus of the current study

Texts of the corpus were recorded by the Prevention of Mental Illnesses Interdisciplinary Research Group (University of Szeged, Hungary) led by István Szendi. Data collection was approved by the Ethics Committee of the University of Szeged, and it was conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained from all the participants involved in the research project. We had official written permission to use the recordings in our study. A medical diagnosis for each person was also provided with the speech samples.

The database contains spontaneous speech recordings of people suffering from three types of mental diseases, as well as controls. There were 90 subjects participating in the data collection procedure altogether. A total of 526 separate transcripts were made; 183 belonged to the SZ, 91 to the SAD, 94 to the BD and 158 to the control group.

In the case of spontaneous speech, in contrast to planned speech, speakers do not have time to prepare their speech, which might better reflect their language specificities, for instance, difficulties in word finding (Vincze, Üveges, Szabó, & Takács, 2021). Here, texts were produced in three different directed spontaneous speech tasks. The first exercise (henceforth Narr1) consisted of three parts. The interviewer first asked the subjects to talk about themselves and then to describe their mother and father. In some cases, the respondents might not have wanted to talk about a parent, in which case they could choose another person close to them as the subject of the monologue. In the first part of the following task (henceforth Narr2), the respondents were asked to recall the last years of their studies or the first years of their employment. The interviewer then asked them to describe the same period in life of someone close to them. Lastly, in the third task (henceforth TegnN), the subjects were asked to talk about their previous day. Basic data of the corpus are presented in Table 1.

	Groups				
	Control	**SZ**	**SAD**	**BD**	**all**
Number of participants	27	31	16	16	90
Number of texts	158	183	91	94	526
Age; M(SD)	38.28(10.03)	38.00(9.78)	40.09(9.77)	49.42(8.49)	38.89(10.55)
Education; M(SD)	14.67(2.99)	14.26(2.86)	14.91(2.96)	16.34(4.05)	14.87(3.18)
Sex ratio; f:m	13:14	10:21	10:6	9:7	42:48

Table 1: Basic data of the four subject groups. Education is given in years.
(M = mean, SD = standard deviation)

3.2 Automatic data processing

Since the texts were manually transcribed, there was no need for any data cleaning steps prior to automatic corpus processing. Thus, as a first step, we performed on the transcribed texts an automatic linguistic analysis with magyarlanc, a linguistic processing toolkit for Hungarian (Zsibrita, Vincze, & Farkas, 2013). With this tool, the texts were first split into sentences, then tokenized, the tokens were lemmatized and assigned a proper part-of-speech and morphological tag.[1] Dictionary analyses were also carried out on the corpus, in which sentiment and intensifier dictionaries were applied.[2]

[1] Lemmatization is especially important in the case of morphologically rich languages like Hungarian (Kondratyuk, Gavenčiak, Straka, & Hajič, 2018).

[2] We deliberately use the term *sentiment* instead of *polarity*. In pragmatics, "polarity is, in essence, the relation between semantic opposites — between meanings (or expressions denoting meanings) which are fundamentally inconsistent with each other. As such polarity encompasses [...] the conceptual relations defining contrary pairs like hot-cold, long-short, and good-bad" (Israel, 2004). Since we are interested in the evaluative semantic content, the term *sentiment* met our

group	tNum	tNumNoP	intNum	int%	NEINum	NEIs%	sentNum	sent%
SZ	42335	34670	767	2.2123	20	0.0576	3242	9.3510
SAD	30350	23685	532	2.2461	19	0.0802	1939	8.1866
BD	49820	39752	1107	2.7848	49	0.1232	3733	9.3907
ALLpat	122505	98107	2406	2.4524	88	0.0896	8914	9.0859
Control	75542	60279	1646	2.7306	48	0.0796	5080	8.4274

Table 2: Basic statistical data of the automatically processed corpus.
(tNum = number of tokens , tNumNoP = number of tokens without punctuation , intNum = number of SRIs, int% = proportion of SRIs to the number of tokens, NEINum = number of NEIs, NEIs% = proportion of NEIs to the number of tokens, sentNum = number of sentiment words, sent% = proportion of sentiment words to the number of tokens)

For the dictionary-based analysis, we used a sentiment dictionary, for which we combined two previously published Hungarian sentiment dictionaries (Szabó (2015) and a Hungarian translation of Liu (2012)). For the identification of SRIs, we used a dictionary consisting of a total of 125 words (Szabó et al., 2023) and a NEI dictionary consisting of 225 words (Szabó & Guba, 2022). The sentiment analysis was performed on the lemmas, and this gave us the opportunity to find all the suffixed forms of the dictionary lemmas in the corpus (cf. the word forms of the lemma *szép* 'nice' and *szépet* 'nice-ACC'). At the same time, the SRI and the NEI lexicons were applied on the word forms instead of the lemmas, since in our wider research project we would like to analyse separately intensifiers with different morphological properties (cf. e.g. *borzasztó* 'awful' and *borzasztóan* 'awfully') (Szabó, Vincze, & Bibok, 2022; Szabó et al., 2023).

The basic statistical data of the processed corpus are presented in Table 2.

As for the frequency of SRIs and NEIs detected via our dictionary-based method, it should be remarked that their pure frequency obviously does not reveal how frequently these words are used in an intensifier function. SRIs and NEIs are highly ambiguous words, and there are several occurrences in the corpus where they do not act as an intensifier. Since a dictionary-based analysis method was applied here, we simply matched the tokens of the corpus to the tokens of the lexicons. In order to find the analysed words in an intensifier function, as a first step of further analysis, we decided to carry out a manual annotation of the resulting data.

3.3 Manual annotation of the corpus

Prior to the automatic data processing steps, all the sentences containing any of the SRIs or NEIs were filtered out from the corpus for manual annotation. At this point, we decided to work with two different approaches for the two word groups.

As for SRIs, we decided to automatically reduce the amount of data before any further steps. To achieve this, we relied on the POS-tags of the collocators of the examined words. Since intensifiers mainly modify adjective, adverb or verb (Bolinger, 2013; Chang & Bin, 2020; Méndez-Naya, 2008), by filtering the occurrences based on POS-tags we were able to remove a significant amount of irrelevant data (Szabó, Vincze, Ring, & Guba, 2022). This way, 4052 examples remained in total. Then, we analysed and annotated manually all the examples. During this phase, we simply decided whether the given word was an intensifier in the given utterance or not.

As for NEIs, since the amount of occurrences was much less than that of SRIs, we decided to annotate all the examples here. The main goal here was again to annotate whether the given occurrence was an intensifier in the given construction or not. At the same time, since NEIs are semantically and pragmatically complex and their linguistic features form a peculiar focus of our wider research project, we annotated several different features here as well. For this, we relied on Szabó and Otani (2022) which applied a multivariable analysis on NEIs, and considered several linguistic factors. However, it was necessary to slightly modify this annotation system since there the researchers annotated conversations but here we analysed monologues. Hence, during this phase we annotated the following features:

- Category: SYNT1. Here we annotated whether the NEI in the given context was an interjection or a modifier. NEIs can function as modifiers (such as intensifiers) as well as interjections (e.g. Durva! Nem hiszek a szememnek! 'Wow! I can't believe my eyes!' (Szabó & Otani, 2022))

research objectives here.

group	tokenNum	IntNum	realInt	Int%	realInt%	ratio%
SZ	34670	767	357	2.2123	1.0297	46.54
SAD	23685	532	215	2.2461	0.9077	40.41
BD	39752	1107	540	2.7848	1.3584	48.78
ALLpat	98107	2406	1115	2.4524	1.1365	46.34
Control	60279	1646	689	2.7306	1.1430	41.86

Table 3: The number and frequency of SRIs.
(tokenNum = number of tokens, IntNum = number of SRIs, realInt = number of SRIs annotated manually as intensifier, Int% = proportion of SRIs to the number of tokens, realInt% = proportion of SRIs annotated manually as intensifiers to the number of tokens, ratio = proportion of SRIs annotated manually as intensifier to all automatic search results)

in the Hungarian language. When they are interjections, they do not modify anything; based on Ameka (1992), interjection are words or short phrases "which can constitute an utterance by themselves and do not normally enter into constructions with other word classes". Interjections traditionally include items like discourse markers, continuers, attention signals, hesitators, and expletives (Norrick, 2009; Quirk, 2010).

- Category: SYNT2. We decided to code the syntactic feature of the modified heads because the collocation features of NEIs modify is related to their semantic change. Thus, this part-of-speech feature may tell us about the semantic-pragmatic peculiarities about the given NEIs. NEIs in the Hungarian language can modify adjectives, adverbs, verbs, nouns or even a whole clause or a sentence. We relied on the results of the automatic pos-stagging here as well, but the manual annotation may be more reliable since NEWs modify sometimes other words instead of the closest following word.

- Category: SEM1. Here we annotated the sentiment value of the given utterance the NEIs occurred in. We borrowed the term "sentiment" from the computational linguistic field, where this term refers to "people's opinions, attitudes and emotions toward an entity" (Medhat, Hassan, & Korashy, 2014). During the annotation procedure, we decided whether the given utterance was positive, negative or neutral. The latter was the case, for instance, when the utterance expressed the surprise of the speaker, or did not have an exact positive or negative sentiment value, respectively. It should be added that for sentiment features we will also rely on the results of the automatic sentiment annotation which determines the sentiment value of the first right collocators of NEIs in the corpus. During the manual sentiment annotation we analysed the sentiment of the whole utterance instead of the closest following word.

- Category: SEM2. Here we annotated basic fucntions and meanings of NEIs. Annotated categories were the following: 1) literal meaning. E.g., in the Hungarian language, the word *durva* literally means 'coarse, rough in texture', for example *durva szemcse* 'coarse grain' (see Section 2.1). 2) An intensifier function. 3) An affective-evaluative meaning expressing evaluation, feelings, impressions of the speaker. 4) Other meanings which are irrelevant to us here in this study. E.g. the word *durva* can also mean 'approximately, roughly'.

.

4 Results

Now, we will present the corpus analysis results. We will provide the data on frequency distributions based on the overall number of tokens without punctuation.

Firstly, we will present the results on SRIs. The number and frequency of SRIs are shown in Table 3. The column "ratio" shows the proportion of words annotated manually as intensifier to all automatic search results.

Based on the results of Table 3, BD patients use intensifiers the most frequently compared to all the other groups (2.7848%), even when just considering the manually annotated data (1.3584%). The other two patient groups use intensifiers less frequently than the control group (SZ: 2.2123%, SAD: 2.2461%, control: 2.7306%). When we compare all patient groups with the control group we notice that the latter

SZ			SAD			BD			Control		
SRI	**num**	**freq%**	**SRI**	**num**	**freq%**	**SRI**	**num**	**freq%**	**SRI**	**num**	**freq%**
nagyon	247	69.19	nagyon	145	67.44	nagyon	354	65.56	nagyon	408	59.22
elég	22	6.16	elég	25	11.63	annyira	32	5.93	elég	65	9.43
annyira	22	6.16	annyira	12	5.58	elég	32	5.93	annyira	51	7.40
igazán	14	3.92	túl	8	3.72	nagyon-nagyon	26	4.81	teljesen	30	4.35
olyan	12	3.36	teljesen	7	3.26	olyan	22	4.07	olyan	30	4.35

Table 4: Number and frequency of the five most common SRI in each group.

group	**pos**	**pos%**	**neg**	**neg%**	**neut**	**neut%**
SZ	190	53.22	52	14.57	115	32.21
SAD	110	51.17	38	17.67	67	31.16
BD	268	49.63	84	15.56	188	34.81
ALLpat	568	51.08	174	15.65	370	33.27
Control	340	49.35	80	11.61	269	39.04

Table 5: Sentiment analysis results on the collocators of SRIs.

uses intensifiers slightly more frequently (ALLpat: 2.4524%, control: 2.7306%), which remains true even in the case of the manually annotated data (ALLpat: 1.1365%, control: 1.1430%).

As for the most frequently occurring SRIs, patients and the control group seem to prefer to use mostly the same words. Table 4 shows the five most common SRIs in each group based on the manually annotated data.

Next, let us look at the sentiment values of the words collocating with SRIs. As mentioned above (see Section 3.2), in the case of SRIs we used a sentiment dictionary for the task so that we could find the sentiment values of the direct collocators of SRIs. The results can be seen in Table 5.

As we can see, there are notable differences in percentage use of different sentiment values. For instance, the most negative sentiment word appears in the subcorpus of SAD speakers, while the most positive is in the SZ-data.

Lastly, let us see the frequency distribution of verbs, adjectives and adverbs modified by SRIs. Table 6 shows the frequency distribution of verbs, adjectives and adverbs modified by SRIs.

We can see that the adjectival modifier is the most frequent in each speaker group followed by verb, and then the adverbial modifiers. However, the the equalisation of the frequency of the three collocator types is different in the speaker groups. For instance, the proportion of verb modifiers in the SZ-group is higher than in any other sub-corpus; in other words, the proportion of verbs and adjectives is more balanced here than in other groups. At the same time, controls use the most adverbial modifiers here, more than any other group of the current study.

Now, let us present the linguistic specificities of NEIs in our corpus. Given that manual annotation was a complex multivariable analysis, we will not systematically present the results in an exhaustive manner. Instead, we shall focus on the most notable peculiarities, relying on the combination of some individual variable.

From the 225 NEIs in the applied NEI dictionary 41 occurred in our dataset, resulting in 136 corpus examples in total. The number and frequency of NEIs broken down into each speaker group are presented in Table 7. (These figures were calculated before further manual word sense disambiguation.)

As we can see, patients use NEI words slightly more frequently than the members of the control group. At the same time, the BD group use the most of these words relative to the total number of tokens, while the members of the SZ group use the least words from the NEI dictionary.

	SZ		SAD		BD		ALLpat		Control	
POS	**num**	**freq%**	**num**	**freq%**	**num**	**freq%**	**num**	**freq%**	**num**	**freq%**
VERB	131	36.69	65	30.23	143	26.48	339	30.49	195	28.30
ADJ	167	46.78	112	52.09	304	56.30	583	52.43	344	49.93
ADV	59	16.53	38	17.67	93	17.22	190	17.09	150	21.78

Table 6: Proportion of the three examined POS-tags of the collocators of SRIs.

group	tokenNum	NEINum	NEI%
SZ	34670	20	0.0576
SAD	23685	19	0.0802
BD	39752	49	0.1232
ALLpat	98107	88	0.0896
Control	60279	48	0.0796

Table 7: The number and frequency of NEIs based on automatic data processing.

group	tN	lit%	int%	evem%	other%	NEIN	lit%	int%	evem%	other%
SZ	42335	0.0236	0.0118	0.0094	0.0023	20	50	25	20	5
SAD	30350	0.0527	0.0032	0.0032	0	19	84.2105	5.2631	5.2631	0
BD	49820	0.0381	0.0421	0.0160	0.0020	49	38.7755	42.8571	16.3265	2.0408
ALLpat	122505	0.0367	0.0220	0.0106	0.0016	88	51.1363	30.6818	14.7727	2.2727
Control	75542	0.0158	0.0357	0.0079	0.0039	48	25	56.25	12.5	6.25

Table 8: The frequency of different NEI functions in proportion to the total number of tokens, as well as to the total number of NEIs (category: SEM2).
(tN = number of tokens, lit = literal meaning, int = intensifier function, evem = emotive-evaluative function, other = other meaning, NEIN = number of NEIs)

In order to get better acquainted with the linguistic features of NEIs in the corpus, we need to rely on our manual annotation. Below we present the results obtained by annotation on the actual meaning and function of the NEIs (SEM2 in our annotation system, see 3.3). In Table 8, we present the frequency distributions of the four basic functions of NEIs in proportion to the total number of tokens, as well as to the total number of NEIs. In this table, for the sake of brevity, we will not show the figures but just the frequency values.

Based on the data, patients suffering from any of the three mental illnesses use fewer NEIs as an intensifier than the control group. This is definitely true for SAD and SZ patients. In the SAD subcorpus, there is only one occurrence all together (which amounts to only 0.0032% compared to the token number). It should also be added that in the SAD-subcorpus most NEIs occurred in their literal sense (84.2105%). As for SZ-patients, there were only 5 occurrences for the intensifier usage of NEIs . In contrast to SAD and SZ groups, in the BD-corpus the frequency of NEIs is generally higher than in the control-corpus, relative to the total number of tokens (BD: 0.0421, Control: 0.0357). What is more, most of the NEIs are used as an intensifier here, and in the BD-data there are notable differences compared to the two other patient groups in this respect.

Next, let us see the sentiment values of the words collocating with NEIs in an intensifier function. As mentioned above (see Section 3.3), here we relied on our manual sentiment annotation of the whole utterances the given NEI occurred in. The figures can be seen in Table 9.

As we observe, there are notable differences in proportions of different sentiment values in the speakers' groups. For instance, positive, neutral and negative sentiments are quite balanced in the BD-subcorpus and this feature is similar to that of the control group. However, there is no positive utterance containing a NEI in an intensifier function in the two other subcorpora at all.

Lastly, let us look at the frequency distribution of the parts-of-speech of the words modified by the NEI words in an intensifier function. Table 13 lists the frequency of each part-of-speech of the NEIs' collocators calculated using the overall number of NEIs in an intensifier function.

As we can see, NEIs in an intensifier function modifies adjectives, verbs, adverbs and nouns in our corpus. However, the latter is observable only in the patients' group; it does not occur in the control

group	pos	pos%	neg	neg%	neut	neut%
SZ	-	0	3	60	2	40
SAD	-	0	-	0	1	100
BD	10	45.45	5	22.73	7	31.82
ALLpat	10	38.46	6	23.08	10	38.46
Control	10	35.71	4	14.29	14	50

Table 9: Sentiment analysis results for the NEIs in an intensifier function.

group	pos	pos%	neg	neg%	neut	neut%
SZ	2299	6.63	943	2.72	31428	90.65
SAD	1315	5.55	624	2.63	21746	91.81
BD	2559	6.43	1174	2.95	36019	90.61
ALLpat	6173	6.29	2741	2.79	89193	90.90
Control	3609	5.99	1471	2.44	55199	91.57

Table 10: Sentiment analysis results on each sub-corpus.

	SZ		SAD		BD		ALLpat		Control	
POS	num	freq%	num	freq%	num	freq%	num	freq%	num	freq%
VERB	7123	20,5451	4645	19,6115	7506	18,8820	19274	19.6458	11018	18,2783
ADJ	2532	7,3031	1538	6,4935	3042	7,6524	7112	7.2492	4278	7,0969
ADV	6082	17,5425	4463	18,8431	7010	17,6343	17555	17.8937	11328	18,7926

Table 11: The proportion of adjectives, verbs and adverbs in the entire corpus.

data at all.

5 Discussion

Here as a first step, we will examine the results on SRIs.

As we saw in Table 3, controls use more SRIs than patients in general, and this was true for the results of the simple word-to-word matching as well as the manually disambiguated data. However, interestingly, BD patients use SRIs the most frequently compared to all other groups, and not just generally but also in the manually annotated data.

As regards the most frequently occurring SRIs, patient groups and the control group seem to prefer using mostly the same words. Table 4 shows the five most common SRI in each group based on the manually annotated data. The five most common intensifiers in each sub-corpora are from the eight most frequent intensifiers of the entire corpus: *nagyon* ('very'), *elég* ('quite'), *annyira* ('so'), *igazán* ('truly'), *olyan* ('such'), *túl* ('too'), *teljesen* ('totally') and *nagyon-nagyon* ('very very'). The most frequently occurring intensifier by far (more than 50% of all instances) is *nagyon* ('very') in all groups. However, we notice that it is used more frequently by patients groups; that is, the use of SRIs by the control group is slightly more balanced, suggesting that their word usage is more diverse.

Let us now discuss the results on sentiment analysis. As we saw (Section 4), there are notable differences in proportions of different sentiment values. To understand the results, let us first look at the number and frequency of sentiment values in the whole corpus (with no regard to intensifiers). The figures are given in Table 10 below.

Based on the data of Table 10, we can see that the most positive sentiment word is used by SZ patients, followed by BD patients. The only patient group that uses fewer positive sentiment words than the controls are SAD patients, whose corpus contains the fewest sentiment word in general. BD patients use the most sentiment words altogether, while controls use the fewest negative words. Let us now compare these general sentiment analysis results to the collocators' sentiments, which can be seen in Table 5. The most positive sentiment words are still used by SZ patients, followed by SAD patients, while controls use the most SRIs with neutral words. Generally, SAD speakers use the most non-neutral sentiment words with SRIs (and the most negative as well), which is interesting, since their data are the most neutral in general (see Table 10 once again).

Lastly, let us discuss the frequency distribution of the three examined parts-of-speech of the collocators. First of all, let us see the overall frequency of the three parts-of speech in the entire corpus. The figures are presented in Table 11. "Frequency" shows the frequencies as a fraction of the total number of all part-of-speech.

We can see that patients use verbs and adjectives more frequently than controls, but adverbs appear more frequently in the corpus of the controls compared to the corpus of all patients. More precisely, verbs are used the most by SZ patients (20.55%), adjectives by BD-s (7.65%), and adverbs by SAD-s (18.84%). Comparing the general data to results restricted to the intensifier function of the examined words, the frequency of verbs and adjectives remains similar to the general frequencies. However, the contrast is greater between the groups: differences between the lowest and highest frequencies are 10%,

while they were no more than 3% in the whole corpus. As for adverbs, they are modified by SRIs the most often in the corpus of the control group.

Let us continue the discussion with the results on NEIs now.

As we saw, BD group uses the most NEIs in the corpus in general. Based on this, one might assume that the corpus of BD patients contains more non-neutral words in terms of sentiment values in general again, with no relation to NEIs. This assumption is confirmed by the figures presented in Table 9 as well. In order to get a more objective picture of these results, let us scrutinise the overall sentiment scores of the whole corpus in Table 10 again (see above). As we have already seen, the BD group uses slightly fewer neutral words than the other groups, but the difference is not notable. So the higher number of NEIs cannot be simply explained by the bigger number of sentiment words in this case.

As for the SZ group, we found only 5 occurrences of NEIs in the entire SZ corpus (0.0118%). What is more, SZ patients never use them in a positive context; they use NEI mainly in negative utterances, e.g. (2) (Szabó et al., 2023).

(2) [...] *hányingerem* *volt,* rettenetes *hányingerem* *volt,* *és* *öhm* *undorral*
nausea-POSS.1SG be.PST.3SG terrible nausea-POSS.1SG be.PST.3SG and um disgust-INS
mentem *be* *dolgozni.*
go-PST.1SG in work-INF

'I was nauseous, terribly nauseous, and I went to work feeling disgusted.'

Here, it is advisable to refer back to Table 10 again. As we saw, SZ-patients generally use more positive words than healthy controls. Thus, the fact that SZ patients never use NEIs either in a positive or in a neutral context is not related or follows from the sentiment properties of the entire corpus (for further discussion on this feature, see below). It is also worth mentioning that the frequency of negative sentiment words was the highest in this sub-corpus, namely the data of SZ-patients. Hence, we may conclude that the biggest number of negative words do not lead to the highest frequency of NEIs, either.

Although the amount of data does not allow us to draw unlimited conclusions, our findings are in accord with other results found in other studies on figurative language use in the case of SZ (see, among others, Pantano, Fu I, Curatolo, Martins, and Elkis (2016); Rossetti, Brambilla, and Papagno (2018)). We therefore hope that our findings will serve as an addition to these research studies.

The results given on the corpus of the BD group are notably different from SZ and SAD results (just remember, the latter group used an NEI just once as an intensifier in their entire sub-corpus). One may conclude that they use NEIs quite "creatively" compared to others. They uses 11 different NEIs in an intensifier function and 47.62% of all the cases occurred in a positive (`[3a]`) or a neutral context (`[3b]`), e.g.

(3) a. [...] *és* rettentő *jó* *szíve* *volt.*
and terrifying good heart-POSS.3SG be.PST.3SG

'[...] and he had a very kind heart.'

b. [...] *egy* hihetetlen *aktív* *nénikéről* *van* *szó,* *föntvan* *a*
an incredible active lady-DIM-DEL be.PRS.3SG word up be.PRS.3SG the
Facebookon *például* [...]
Facebook-SUPESS for_example

'[...] she is an incredibly active lady, she is on Facebook, for example [...]'

In the latter case, the NEI is used for expressing that the given information is surprising, incredible, or very interesting to the speaker. Hence, the given utterance does not have an exact positive or negative sentiment value but it expresses the surprise or amusement of the speaker. This neutral use of NEIs occurred in the BD and the Control sub-corpora exclusively.

Our research results are especially interesting when we consider the characteristics of NEIs regarding their semantic changing processes over time (that is, their prior negative content fades; see Section 2.1 above). The results show that, in the case of the SZ group, the prior negative semantic component of NEIs obviously determines their semantic prosody. Based on J. Sinclair (1991); J. M. Sinclair (1987), many words and phrases show a tendency to occur in a certain semantic environment. In our SZ-subcorpus, we can see that the prior negative semantic content of NEIs is echoed in their collocation features (see also

	SZ		SAD		BD		ALLpat		Control	
POS	**num**	**freq%**	**num**	**freq%**	**num**	**freq%**	**num**	**freq%**	**num**	**freq%**
VERB	7123	20,5451	4645	19,6115	7506	18,8820	19274	19.6458	11018	18,2783
NOUN	6025	17,3781	3961	16,7236	6510	16,3765	16496	16.8142	8991	14,9156
ADJ	2532	7,3031	1538	6,4935	3042	7,6524	7112	7.2492	4278	7,0969
ADV	6082	17,5425	4463	18,8431	7010	17,6343	17555	17.8937	11328	18,7926

Table 12: The proportion of adjectives, verbs, adverbs and nouns in the entire corpus.

the phenomenon of *congruence*, Dilts (2010)). However, as we have seen, it is not the case in the BD-subcorpus where this does not fundamentally determine collocation preferences of NEIs. The same can been said about the control group where the proportion of the NEIs in positive and negative contexts are completely balanced. If we examine the definite words collocating with the NEIs in the BD-subcorpus, we can also see that the proportion of the three sentiment values of the collocators is balanced: NEIs modify positive (e.g. *önzetlen* 'selfless', *jó* 'good', *tehetséges* 'talented'), neutral (e.g. *nagy* 'large', *sok* 'many', *erős* 'strong'), and negative words (e.g. *idegesítő* 'irritating', *szabályellenes* 'malpractice', *utál* 'hate') with almost equal frequency.[3]

Lastly, let us discuss the frequency distribution of the parts-of-speech of the NEIs.

First of all, we examine the overall frequency of parts-of-speech in the entire corpus again, just as we did in the case of SRIs before. Some data on parts-of-speech were previously given in Table 11 above. Here, we repeat the main data supplied by the data on noun collocators in Table 12 again.

As we observe, patients use more verbs and nouns, as well as slightly more adjectives in the corpus. At the same time, they use adverbs with less frequency.

Now let us compare these results with the data restricted to NEIs in an intensifier function. Given the small number of NEIs as an intensifier, we shall no longer divide the frequency of individual parts-of-speech into groups, but only discuss the figures in contrast to patients and controls.

As can be seen, in the data of the healthy controls NEI intensifies proportionally more adjectives

[3] In this study we will not discuss functions on NEIs other than intensifier in detail. However, we shall briefly mention the main findings on these functions here.

As for the an affective-evaluative meaning (see function "evaluative-emotive" in SEM2 in Section 3.3 above), we detected 19 cases in the corpus altogether. Here, as discussed in Section 2.1, the original literal meaning had faded and the NEI is used in an evaluative-emotive sense (without any intensifier semantic component). The frequency of this function is higher in the BD-group than that in the two other patient groups and even in the control group data again. E.g.

(4) a. [...] *ha nyitott ablaknál alszunk, akkor az öntödének* iszonyat *hangja van*
if open window-ADESS sleep-PRS.1PL then the foundry-DAT terror sound-POSS.3SG be.PRS.3SG
az éjszaki műszakkal [...]
the night-ADJ shift-INS

'[...] if we sleep with an open window, the foundry is terribly noisy with the night shift [...].'

b. [...] *rossz választás lett volna, mer nem vagyok* durva *lelkületű ember* [...]
bad choice become.PST.3SG be-COND.3SG because not be-PRS.1SG harsh spirit-ADJ person

'[...] It would have been a bad choice, because I'm not a harsh person [...]'

An other interesting finding is that the neutral type, so when the utterance expresses that the speaker find the given information special or surprising, occurs in the control data exclusively. E.g.

(5) [...] *többen lettünk eggyel öhm ... ami szintén egy egy öhm ilyen* elképesztő *érzés*
many.COMP-ADV become-PST.1PL one-INS um which also a a um such astonishing feeling
[...]

'[...] we became more people by one um ... which is also such an astonishing feeling [...]'

What is more, there is only one example for the interjective usage of NEIs and it is also in the control data (see SYNT1 in section 2.2 above):

(6) *Hát ez* durva*, nem hallottam.*
well this harsh not hear-PST.1SG

'That's crazy. I haven't heard about that.'

group	intNum	a%	v%	av%	n%
ALLpat	27	55.55	25.9259	3.7037	14.8148
Control	27	66.66	14.8148	18.5185	0

Table 13: Proportion of the POS-tags of the collocators to the total number of NEIs in an intensifier function (category: SYNT2).
(a = adjective, v = verb, av = adverb, n = noun

than in patients' data - despite the fact that the controls use proportionally fewer adjectives in general. The frequency distributions of other parts-of-speech here are in line with the general frequencies. At the same time, as already mentioned, there is no NEI intensifying a noun in the control group data.

Lastly, we were curious to know whether the frequency distribution of parts-of-speech of first right collocators could tell us anything about the delexicalization of NEI words, regardless of their actual functions (in contrast to Table 13). NEIs when they occur in their prior sense modify nouns (e.g. *durva szemcse* 'coarse grain', *borzasztó tapasztalat* 'terrible experience') or verbs in their adverbial forms (e.g. *rettenetesen viselkedik* 'he/she behaves terribly', *borzasztóan néz ki* 'he/she looks terrible'). At the same time, in their intensifier function they mainly modify adjectives, adverbs or verbs, and rarely nouns. Consequently, The higher frequency of the latter parts-of-speech may also reflect some divergence from the prior function of NEIs. Figure 1 shows the proportion of all the part-of-speech of collocators occurring together with NEIs in any function.

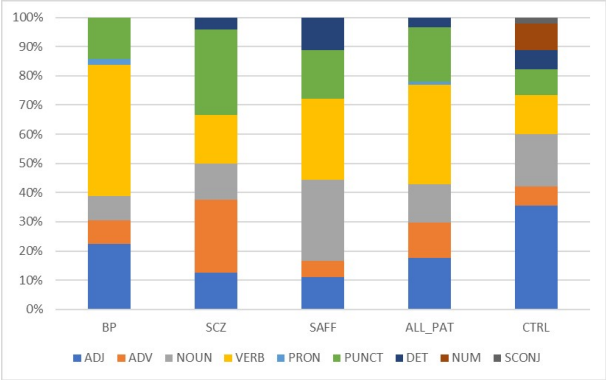

Figure 1: The frequency distribution of parts-of-speech of first right collocators of NEIs in general (disregarding NEIs' actual semantic function)

As we can see in the figure, SAD-patients use the examined words with nouns with notably higher frequency than the other two patient groups, and differs from the control-data as well. What is more, BD-group is the closest to the controls in terms of the frequency of adjective collocators. The results may mean that some disorder types make the delexicalized use of NEIs difficult for the patients. However, we also see a high occurrence of verbial collocates in the BD-group, and for that of the adverbial collocates in SZ-group, which cannot be explained in this phase of the research project. Based on this, we consider it to be necessary to carry out more research work on this.

6 Conclusions and limitations

In this paper, we analysed the phenomenon of intensification in spontaneous speech of Hungarian patients suffering from schizophrenia (SZ), schizoaffective (SAD) and bipolar disorders (BD), using NLP tools and methods. We focused on linguistic intensification among the linguistic parameters. Relying on connections between mental illnesses and emotion regulation, within the group of intensifiers, we paid special attention to the so-called negative emotive intensifiers (we referred to them as NEIs here).

As for the standard register intensifiers (we referred to them as SRIs here), we learned that controls use more SRIs than patients in general, and this was true for the results of the simple word-to-word matching as well as the manually disambiguated data. What is more, the most frequently occurring SRIs was *nagyon* ('very') in all groups, but the use of SRIs by the control group turned out to be slightly more balanced, suggesting that their word usage is more diverse.

Based on the sentiment analysis results for the entire corpus, the most positive sentiment words are used by SZ patients. The only patient group using fewer positive sentiment words than the controls were SAD patients. BD patients generally used the most (non-neutral) sentiment words but the difference was not remarkable. SAD speakers use the most non-neutral sentiment words with SRIs (and the most negative as well), which is interesting, since their data is the most neutral in general (see Table 10 once again).

As for the NEIs, the study revealed that the BD group uses the most NEIs in the corpus in general. At the same time, the bigger number of NEIs cannot be explained by the bigger number of sentiment words since this group did not use non-neutral words with a notewothy higher frequency than the other speakers' groups.

We found only 5 occurrences of NEIs in an intensifier function in the entire SZ-subcorpus and 1 in SAD-subcorpus. It was quite surprising that SZ patients never used them in a positive context; they used them in negative or neutral contexts. We concluded that, in the case of the SZ group, the prior negative semantic component of NEIs probably determines their semantic prosody.

The corpus of the BD group proved to be notably different from SZ and SAD data; BD-patients used NEIs quite "creatively" compared to others: they used 11 different NEIs in an intensifier function almost the half of the cases occurred in a positive or a neutral context. It also means that the frequency of sentiment words in the BD-subcorpus does not really determine the collocation preferences of NEIs. It was also true for the control group subcorpus where the proportion of the NEIs in positive and negative contexts were quite balanced.

When evaluating and using the results presented here, it is important to keep in mind that the amount of data was limited during the current work. However, our results of the examination indicated some noteworthy differences in language use, especially in the case of NEIs. Based on this, we think that the current research work, as well as further investigation of NEIs could be useful in ascertaining the linguistic characteristics of mental illnesses, especially in relation to higher-order language functions. Thus, the research should definitely be continued using larger amounts of linguistic data.

Acknowledgements Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

[anonimity]

References

Akkaralaertsest, T., & Yingthawornsuk, T. (2015). Comparative analysis of vocal characteristics in speakers with depression and high-risk suicide. *International Journal of Computer Theory and Engineering*, *7*(6), 448.

Al-Mosaiwi, M., & Johnstone, T. (2018). In an absolute state: Elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. *Clinical Psychological Science*, *6*(4), 529-542. Retrieved from `https://doi.org/10.1177/2167702617747074` (PMID: 30886766) doi: 10.1177/2167702617747074

Ameka, F. (1992). Interjections: The universal yet neglected part of speech. *Journal of pragmatics*, *18*(2-3), 101–118.

Athanasiadou, A. (2007). On the subjectivity of intensifiers. *Language sciences*, *29*(4), 554–565.

Bagi, A., Gosztolya, G., Szalóki, S., Szendi, I., & Hoffmann, I. (2019). Szkizofrénia azonosítása spontán beszéd temporális paraméterei alapján–egy pilot kutatás eredményei. In V. V. Berend Gábor Gosztolya Gábor (Ed.), *Magyar számítógépes nyelvészeti konferencia 15.* (p. 189-201). Szegedi Tudományegyetem, Informatikai Intézet.

Bagi, A., Hoffmann, I., Szendi, I., & Toth, L. (2016). Metaphor in a different way: The understanding of metaphor and irony in schizophrenia. In (p. 7-18). Palacký University, Olomouc.

Balogh, P. (2009). Gender-markerek a nyelvben. *Wiener Elektronische Beiträge des Instituts für Finno-Ugristik (WEBFU)*, 1–7.

Baranyai, A., Szegedi-Hallgató, E., Szôllôsi, , Marián, M., Pajkossy, P., Bagi, A., ... Racsmány, M. (2020). Végrehajtó funkciók deficitjei szkizofréniában és szkizoaffektív zavarban - egy kutatás pilot eredményei. *Psychiatria Hungarica*, *XXXV. Suppl.1.*, 12–13.

Berlim, M. T., Mattevi, B. S., Belmonte-de Abreu, P., & Crow, T. J. (2003). The etiology of schizophrenia and the origin of language: overview of a theory. *Comprehensive psychiatry*, *44*(1), 7–14.

Bickerton, D. (1990). *Language and species*. University of Chicago Press.

Bolinger, D. (2013). *Degree words*. De Gruyter Mouton.

Burdick, K. E., Ketter, T. A., Goldberg, J. F., & Calabrese, J. R. (2015). Assessing cognitive function in bipolar disorder: challenges and recommendations for clinical trial design. *The Journal of clinical psychiatry*, *76*(3), 1427.

Calvo, R. A., Milne, D. N., Hussain, M. S., & Christensen, H. (2017). Natural language processing in mental health applications using non-clinical texts. *Natural Language Engineering*, *23*(5), 649–685.

Chang, Y., & Bin, S. (2020). Literature Review on Negative Emotive Intensifiers in English. *Higher Education of Social Science*, *18*(2), 55–59.

Chapman, H. C., Visser, K. F., Mittal, V. A., Gibb, B. E., Coles, M. E., & Strauss, G. P. (2020). Emotion regulation across the psychosis continuum. *Development and psychopathology*, *32*(1), 219–227.

Chaput, V., Amsellem, F., Urdapilleta, I., Chaste, P., Leboyer, M., Delorme, R., & Goussé, V. (2013). Episodic memory and self-awareness in asperger syndrome: Analysis of memory narratives. *Research in Autism Spectrum Disorders*, *7*(9), 1062–1067.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. Greenwood Publishing Group.

Corballis, M. C. (2009). The evolution of language. *Annals of the New York Academy of Sciences*, *1156*(1), 19–43.

Corcoran, C. M., Carrillo, F., Fernández-Slezak, D., Bedi, G., Klim, C., Javitt, D. C., ... Cecchi, G. A. (2018). Prediction of psychosis across protocols and risk cohorts using automated language analysis. *World Psychiatry*, *17*(1), 67-75. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1002/wps.20491` doi: https://doi.org/10.1002/wps.20491

Crow, T. J. (1993). Sexual selection, machiavellian intelligence, and the origins of psychosis. *The Lancet*, *342*(8871), 594–598.

Crow, T. J. (1995). A darwinian approach to the origins of psychosis. *British Journal of Psychiatry*, *167*(1), 12–25.

Crow, T. J. (1996). Language and psychosis: common evolutionary origins. *Endeavour*, *20*(3), 105–109.

Crow, T. J. (1997). Schizophrenia as failure of hemispheric dominance for language. *Trends in neurosciences*, *20*(8), 339–343.

Crow, T. J. (1998). Why cerebral asymmetry is the key to the origin of homo sapiens: how to find the gene or eliminate the theory. *Cahiers De Psychologie Cognitive-Current Psychology of Cognition*, *17*(6), 1237–1277.

DeLisi, L. E. (2001). Speech disorder in schizophrenia: review of the literature and exploration of its relation to the uniquely human capacity for language. *Schizophrenia bulletin*, *27*(3), 481–496.

Dilts, P. (2010). Good nouns, bad nouns: what the corpus says and what native speakers think. In *Corpus-linguistic applications* (pp. 103–117). Brill Rodopi.

Döme, P., Rihmer, Z., Gonda, X., Pestality, P., Kovács, G., Teleki, Z., & Mandl, P. (2005). Cigarette smoking and psychiatric disorders in Hungary. *International Journal of Psychiatry in Clinical Practice*, *9*(2), 145–148.

Dronkers, N., Pinke, S., & Damasio, A. (2000). Language and the aphasias. *Principles of Neural Science*, *4*, 1169—1173.

Fuchs, R., & Gut, U. (2016). Register variation in intensifier usage across Asian Englishes. *Discourse-pragmatic variation and change: Insights from English*, 185–210.

Ganger, J., & Wold, K. S. (1998). Innateness, evolution, and genetics of language. *Human biology*, 199–213.

Geschwind, N., & Galaburda, A. M. (1985). Cerebral lateralization: Biological mechanisms, associations, and pathology: I. a hypothesis and a program for research. *Archives of neurology*, *42*(5), 428–459.

Gosztolya, G., Bagi, A., Szalóki, S., Szendi, I., & Hoffmann, I. (2018). Identifying schizophrenia based on temporal parameters in spontaneous speech. In *Proc. interspeech 2018* (p. 3408-3412). International Speech Communication Association (ISCA).

Göteson, A., Isgren, A., Sparding, T., Holmén-Larsson, J., Jakobsson, J., Pålsson, E., & Landén, M. (2022). A serum proteomic study of two case-control cohorts identifies novel biomarkers for bipolar disorder. *Translational Psychiatry*, *12*(1), 1–8.

Green, M. J., Cahill, C. M., & Malhi, G. S. (2007). The cognitive and neurophysiological basis of emotion

dysregulation in bipolar disorder. *Journal of affective disorders*, *103*(1-3), 29–42.
Grunze, H. (2015). Bipolar disorder. In *Neurobiology of brain disorders* (pp. 655–673). Elsevier.
Haider, F., De La Fuente, S., & Luz, S. (2019). An assessment of paralinguistic acoustic features for detection of alzheimer's dementia in spontaneous speech. *IEEE Journal of Selected Topics in Signal Processing*, *14*(2), 272–281.
Inczédy-Farkas, G., Benkovits, J., Balogh, N., Álmos, P., Scholtz, B., Zahuczky, G., ... others (2010). Schizobank – the Hungarian national schizophrenia biobank and its role in schizophrenia research. *Orvosi Hetilap*, *151*(35), 1403–1408.
Israel, M. (2004). The pragmatics of polarity. *The handbook of pragmatics*, *16*, 701–723.
Ito, R., & Tagliamonte, S. (2003). Well weird, right dodgy, very strange, really cool: Layering and recycling in English intensifiers. *Language in society*, *32*(2), 257–279.
Jing-Schmidt, Z. (2007). Negativity bias in language: A cognitive-affective model of emotive intensifiers. *Cognitive Linguistics*, *18*(3), 417-443.
Kálmán, J., Devanand, D. P., Gosztolya, G., Balogh, R., Imre, N., Tóth, L., ... Pákáski, M. (2022). Temporal speech parameters detect mild cognitive impairment in different languages: validation and comparison of the speech-gap test® in English and Hungarian. *Current Alzheimer Research*, *19*(5), 373–386.
Kárpáti, E., Bagi, A., Szendi, I., Tóth, L. B., Janacsek, K., & Hoffmann, I. (2018). Rekurzió egy szkizoaffektív zavarral élő személy diskurzusaiban–esettanulmány. *Iskolakultúra: Pedagógusok Szakmai-Tudományos Folyóirata*, *28*(5-6), 40–54.
Kéri, S., Kelemen, O., Benedek, G., & Janka, Z. (2001). Different trait markers for schizophrenia and bipolar disorder: a neurocognitive approach. *Psychological medicine*, *31*(5), 915–922.
Kochetova, L. A., & Ilyinova, E. (2022). Diachronic and dialect variation of english intensifying adverbs in the film dialogue discourse: Corpus-based study. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Serija 2. Jazykoznanije*.
Kocsis-Bogár, K., Nemes, Z., & Perczel-Forintos, D. (2016). Factorial structure of the hungarian version of oxford-liverpool inventory of feelings and experiences and its applicability on the schizophrenia-schizotypy continuum. *Personality and Individual Differences*, *90*, 130–136.
Kondratyuk, D., Gavenčiak, T., Straka, M., & Hajič, J. (2018). Lemmatag: jointly tagging and lemmatizing for morphologically-rich languages with brnns. *arXiv preprint arXiv:1808.03703*.
Kring, A. M., Elis, O., et al. (2013). Emotion deficits in people with schizophrenia. *Annual review of clinical psychology*, *9*(1), 409–433.
Kugler, N. (2014). A nyelvi polaritás kifejezésének egy mintázata, avagy milyen a félelmetesen jó? *Magyar Nyelvőr*, *138*(2), 129–139.
Little, B., Gallagher, P., Zimmerer, V., Varley, R., Douglas, M., Spencer, H., ... others (2019). Language in schizophrenia and aphasia: the relationship with non-verbal cognition and thought disorder. *Cognitive Neuropsychiatry*, *24*(6), 389–405.
Liu, B. (2012). *Sentiment analysis and opinion mining*. Morgan & Claypool Publishers.
Lorenz, G. (2002). Really worthwhile or not really significant? a corpus-based approach to the delexicalization. *New reflections on grammaticalization*, *49*, 143-163.
Lott, P., Guggenbühl, S., Schneeberger, A., Pulver, A., & Stassen, H. (2002). Linguistic analysis of the speech output of schizophrenic, bipolar, and depressive patients. *Psychopathology*, *35*(4), 220–227.
Lundin, N. B., Hochheiser, J., Minor, K. S., Hetrick, W. P., & Lysaker, P. H. (2020). Piecing together fragments: linguistic cohesion mediates the relationship between executive function and metacognition in schizophrenia. *Schizophrenia research*, *215*, 54–60.
Malhi, G. S., Green, M., Fagiolini, A., Peselow, E. D., & Kumari, V. (2008). Schizoaffective disorder: diagnostic issues and future recommendations. *Bipolar Disorders*, *10*(1p2), 215–230.
Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, *5*(4), 1093–1113.
Méndez-Naya, B. (2008). Special issue on English intensifiers. *English Language & Linguistics*, *12*(2), 213–219.
Minor, K. S., Willits, J. A., Marggraf, M. P., Jones, M. N., & Lysaker, P. H. (2019). Measuring disorganized speech in schizophrenia: automated analysis explains variance in cognitive deficits beyond clinician-rated scales. *Psychological Medicine*, *49*(3), 440–448. doi: 10.1017/S0033291718001046
Mota, N. B., Copelli, M., & Ribeiro, S. (2017). Thought disorder measured as random speech structure classifies negative symptoms and schizophrenia diagnosis 6 months in advance. *npj Schizophrenia*, *3*(1), 1–10.
Nevalainen, T., & Rissanen, M. (2002). Fairly pretty or pretty fair? on the development and grammat-

icalization of english downtoners. *Language Sciences*, *24*(3-4), 359–380.
Norrick, N. R. (2009). Interjections as pragmatic markers. *Journal of pragmatics*, *41*(5), 866–891.
Pantano, T., Fu I, L., Curatolo, E., Martins, C. B., & Elkis, H. (2016). Thought and language disorders in very early onset schizophrenia, schizoaffective disorder and bipolar disorder. *Archives of Clinical Psychiatry (São Paulo)*, *43*, 67–73.
Paradis, C. (1998). Degree modifiers of adjectives in spoken british english. *Functions of Language*, *4*(2).
Partington, A. (1993). Corpus evidence of language change. *Text and technology. In honour of John Sinclair*, 177–192.
Patel, K. R., Cherian, J., Gohil, K., & Atkinson, D. (2014). Schizophrenia: overview and treatment options. *Pharmacy and Therapeutics*, *39*(9), 638.
Peralta, V., & Cuesta, M. J. (2008). Exploring the borders of the schizoaffective spectrum: a categorical and dimensional approach. *Journal of affective disorders*, *108*(1-2), 71–86.
Power, R. A., Kyaga, S., Uher, R., MacCabe, J. H., Långström, N., Landen, M., ... Svensson, A. C. (2013). Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings. *JAMA psychiatry*, *70*(1), 22–30.
Quirk, R. (2010). *A comprehensive grammar of the english language*. Pearson Education India.
Resnik, P., Garron, A., & Resnik, R. (2013). Using topic modeling to improve prediction of neuroticism and depression in college students. In *Proceedings of the 2013 conference on empirical methods in natural language processing* (pp. 1348–1353).
Réthelyi, J. M., Bakker, S. C., Polgár, P., Czobor, P., Strengman, E., Pásztor, P. I., ... Bitter, I. (2010). Association study of nrg1, dtnbp1, rgs4, g72/g30, and pip5k2a with schizophrenia and symptom severity in a Hungarian sample. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *153*(3), 792–801.
Rose, D. (2014). Schizophrenia/psychosis. In *Encyclopedia of the neurological sciences* (pp. 99–103). Academic press.
Rossetti, I., Brambilla, P., & Papagno, C. (2018). Metaphor comprehension in schizophrenic patients. *Frontiers in Psychology*, *9*, 670-685.
Simon, M., Varga, E., Hajnal, A., Orsi, G., Schnell, Z., Tényi, T., ... Herold, R. (2011). Az iróniaértés eltérései szkizofréniában és bipoláris zavarban – fmri vizsgálat. *PSYCHIATRIA HUNGARICA*, *26. Suppl.*, 112.
Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press, USA.
Sinclair, J. M. (1987). *Looking up: An account of the cobuild project in lexical computing and the development of the collins cobuild english language dictionary*. Collins Elt.
Strohm, F., & Klinger, R. (2018). An empirical analysis of the role of amplifiers, downtoners, and negations in emotion classification in microblogs. In *2018 ieee 5th international conference on data science and advanced analytics (dsaa)* (pp. 673–681).
Strous, R. D., Koppel, M., Fine, J., Nachliel, S., Shaked, G., & Zivotofsky, A. Z. (2009). Automated characterization and identification of schizophrenia in writing. *The Journal of nervous and mental disease*, *197*(8), 585–588.
Su, Y. (2016). Corpus-based comparative study of intensifiers: quite, pretty, rather and fairly. *Journal of world languages*, *3*(3), 224–236.
Szabó, M. K. (2015). Egy magyar nyelvű szentimentlexikon létrehozásának tapasztalatai és dilemmái. In *Segédkönyvek a nyelvészet tanulmányozásához 177* (pp. 278–285). Budapest: Tinta.
Szabó, M. K., & Bibok, K. (2019). Értékvesztésre és értékváltásra képes lexémák újabb vizsgálata. *Argumentum*, *15*, 639–649.
Szabó, M. K., & Guba, Cs. (2022). *Analysis of Negative Emotive Intensifiers in Hungarian Tweets.* (Kézirat)
Szabó, M. K., & Otani, N. (2022). *Corpus Analysis of Japanese and Hungarian Negative Emotive Words from a Discourse Interactional Perspective.* The 25th Annual Conference of the Pragmatics Society of Japan The 25th Annual Meeting of Pragmatic Society of Japan. (Kyoto, Japan)
Szabó, M. K., Vincze, V., & Bibok, K. (2022). "Thank you for the terrific party!" – An Analysis of Hungarian Negative Emotive Words, with Special Regard to The Phenomenon of Polarity Shift. *Corpus Linguistics and Linguistic Theory*. (in press)
Szabó, M. K., Vincze, V., Ring, O., & Guba, C. (2022). Nagyot mondó képviselők? : fokozás a politikai kommunikációban. In *Magyar számítógépes nyelvészeti konferencia 18.* (pp. 325–341).
Szabó, M. K., Vincze, V., Guba, C., Dam, B., Solymos, A., Bagi, A., & Szendi, I. (n.d.). Humendisco: A hungarian speech corpus of schizophrenia, schizoaffective and bipolar disorders. *Language Resources*

and Evaluation. (Status: Paper submitted)

Szabó, M. K., Vincze, V., Guba, C., Dam, B., Solymos, A., Bagi, A., & Szendi, I. (2023). Fokozás szkizofréniában. In *Xix. magyar számítógépes nyelvészeti konferencia* (pp. 17–32).

Taguchi, T., Tachikawa, H., Nemoto, K., Suzuki, M., Nagano, T., Tachibana, R., ... Arai, T. (2018). Major depressive disorder discrimination using vocal acoustic features. *Journal of affective disorders*, *225*, 214–220.

Tóth, L., Gosztolya, G., Vincze, V., Hoffmann, I., Szatlóczki, G., Biró, E., ... Kálmán, J. (2015). Automatic detection of mild cognitive impairment from spontaneous speech using asr. In *Proc. interspeech 2015* (pp. 2694–2698).

Trask, R. (1999). Derek bickerton, language and human behaviour. london: Ucl press, 1996. pp. ix+ 180. *Journal of Linguistics*, *35*(3), 657–670.

Van Rheenen, T. E., Bryce, S., Tan, E. J., Neill, E., Gurvich, C., Louise, S., & Rossell, S. L. (2016). Does cognitive performance map to categorical diagnoses of schizophrenia, schizoaffective disorder and bipolar disorder? a discriminant functions analysis. *Journal of Affective Disorders*, *192*, 109–115.

Van Rheenen, T. E., & Rossell, S. L. (2014). Investigation of the component processes involved in verbal declarative memory function in bipolar disorder: utility of the hopkins verbal learning test-revised. *Journal of the International Neuropsychological Society*, *20*(7), 727–735.

Vincze, V., Szabó, M. K., Hoffmann, I., Tóth, L., Pákáski, M., Kálmán, J., & Gosztolya, G. (2021). Linguistic parameters of spontaneous speech for identifying mild cognitive impairment and alzheimer's disease. *Computational Linguistics*, 1–34.

Vincze, V., Üveges, I., Szabó, M. K., & Takács, K. (2021). A magyar beszélt és írott nyelv különböző korpuszainak morfológiai és szófaji vizsgálata. In G. Berend, G. Gosztolya, & V. Vincze (Eds.), *Xvii. magyar számítógépes nyelvészeti konferencia* (pp. 169–182). Szeged: Szegedi Tudományegyetem.

Vioules, M. J., Moulahi, B., Azé, J., & Bringay, S. (2018). Detection of suicide-related posts in twitter data streams. *IBM Journal of Research and Development*, *62*(1), 7–1.

Voleti, R., Woolridge, S., Liss, J. M., Milanovic, M., Bowie, C. R., & Berisha, V. (2019). Objective assessment of social skills using automated language analysis for identification of schizophrenia and bipolar disorder. *arXiv preprint arXiv:1904.10622*.

Weerasinghe, J., Morales, K., & Greenstadt, R. (2019, 10). "because... i was told... so much": Linguistic indicators of mental health status on twitter. *Proceedings on Privacy Enhancing Technologies*, *2019*, 152-171. doi: 10.2478/popets-2019-0063

Willits, J. A., Rubin, T., Jones, M. N., Minor, K. S., & Lysaker, P. H. (2018). Evidence of disturbances of deep levels of semantic cohesion within personal narratives in schizophrenia. *Schizophrenia Research*, *197*, 365-369. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0920996417307107` doi: https://doi.org/10.1016/j.schres.2017.11.014

Zsibrita, J., Vincze, V., & Farkas, R. (2013, September). magyarlanc: A tool for morphological and dependency parsing of Hungarian. In *Proceedings of the international conference recent advances in natural language processing RANLP 2013* (pp. 763–771). Hissar, Bulgaria: INCOMA Ltd. Shoumen, BULGARIA. Retrieved from `https://aclanthology.org/R13-1099`