

slovenščina 2.0, letnik 10, številka 1

Slo2.0

2022_1

Univerza v Ljubljani
Filozofska fakulteta



Slovenščina 2.0

Letnik/Volume 10, Številka/Issue 1, 2022

ISSN: 2335-2736

Glavna urednika/Editors-in-Chief

Špela Arhar Holdt, Vojko Gorjanc

Uredniški odbor/Editorial Board

Zoran Bosnić, Simon Dobrišek, Tomaž Erjavec, Ina Ferbežar, Darja Fišer, Polona Gantar, Peter Jurgec, Iztok Kosem, Simon Krek, Nina Ledinek, Nikola Ljubešić, Nataša Logar, Karmen Pižorn, Damjan Popič, Marko Robnik Šikonja, Amanda Saksida, Irena Srdanović, Mojca Šorn, Darinka Verdonik, Špela Vintar

Tehnična urednica/Managing Editor

Eva Pori

Prelom/Layout

Žiga Valetič

Založila/Published by

Založba Univerze v Ljubljani

Za založbo/For the Publisher

Gregor Majdič, rektor Univerze v Ljubljani

Izdala/Issued by

Znanstvena založba Filozofske fakultete Univerze v Ljubljani;
Center za jezikovne vire in tehnologije Univerze v Ljubljani

Za izdajatelja/For the issuer

Mojca Schlamberger Brezar, dekanja Filozofske fakultete

Publikacija je brezplačna./Publication is free of charge.

Publikacija je dostopna na/Avaliable at: <https://journals.uni-lj.si/slovenscina2>

Revija izhaja s podporo Javne agencije za raziskovalno dejavnost Republike Slovenije./
This journal is published with the support of the Slovenian Research Agency (ARRS).



To delo je ponujeno pod licenco Creative Commons Priznanje avtorstva-Deljenje pod enakimi pogoji 4.0 Mednarodna licenca (izjema so fotografije). / This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (except photographs).

Kazalo vsebine

RAZPRAVE

EMOTION ANALYSIS IN SOCIALLY UNACCEPTABLE DISCOURSE 1

Jasmin Franza, Bojan Evkoski, Darja Fišer

RABA JEZIKOVNIH VIROV IN KOLOKACIJSKEGA SLOVARJA SODOBNE SLOVENŠČINE MED ŠTUDENTI PREVAJALSTVA..... 23

Nataša Hirci, Martin Anton Grad

FROM VERBAL TO ADJECTIVAL: EVALUATING THE LEXICALIZATION OF PARTICIPLES IN AN ESTONIAN CORPUS 65

Greda Paulsen, Maria Tuulik, Ahti Lohk, Ene Vainik

PANELNE RAZPRAVE

OKROGLA MIZA »JEZIKOVNA IDEOLOGIJA, POKLICNI STEREOTIPI TER PRIMERI DOBRE PRAKSE LEKTORJEV IN JEZIKOVNIH SVETOVALCEV« 98

Matejka Grgič, Milica Antić Gaber, Vojko Gorjanc, Erika Kržišnik, Tina Lengar Verovnik, Eva Vrtačič

POROČILA

33. EVROPSKA POLETNA ŠOLA LOGIKE, JEZIKA IN INFORMATIKE 126

Mojca Brglez

KONFERENCA JEZIKOVNE TEHNOLOGIJE IN DIGITALNA HUMANISTIKA 2022 131

David Bordon

SIMPOZIJ OBDOBJA 41: NA STIČIŠČU SVETOV: SLOVENŠČINA KOT DRUGI IN TUJI JEZIK 136

Ina Poteko

Emotion analysis in socially unacceptable discourse

Jasmin FRANZA

Faculty of Arts, University of Ljubljana

Bojan EVKOSKI

Jožef Stefan International Postgraduate School; Jožef Stefan Institute

Darja FIŠER

Faculty of Arts, University of Ljubljana; Jožef Stefan Institute;
Institute of Contemporary History

Texts often express the writer's emotional state, and it was shown that emotion information has potential for hate speech detection and analysis. In this work, we present a methodology for quantitative analysis of emotion in text. We define a simple, yet effective metric for an overall emotional charge of text based on the NRC Emotion Lexicon and Plutchik's eight basic emotions. Using this methodology, we investigate the emotional charge of content with socially unacceptable discourse (SUD), as a distinct and potentially harmful type of text which is spreading on social media. We experiment with the proposed method on a corpus of Facebook comments, resulting in four datasets in two languages, namely English and Slovene, and two discussion topics, LGBT+ rights, and the European Migrants crisis. We reveal that SUD content is significantly more emotional than non-SUD comments. Moreover, we show differences in the expression of emotions depending on the language, topic, and target of the comments. Finally, to underpin the findings of the quantitative

Franza, J., Evkoski, B., Fišer, D.: Emotion analysis in socially unacceptable discourse. Slovenščina 2.0, 10(1): 1–22.

1.01 Izvirni znanstveni članek / Original Scientific Article

DOI: <https://doi.org/10.4312/slo2.0.2022.1.1-22>

<https://creativecommons.org/licenses/by-sa/4.0/>



investigation of emotions, we perform a qualitative analysis of the corpus, exploring in more detail the most frequent emotional words of each emotion, for all four datasets. The qualitative analysis shows that the source of emotions in SUD texts heavily depends on the topic of discussion, with substantial overlaps between languages.

Keywords: emotions, socially unacceptable discourse (SUD), hate speech, social media, corpora

1 Introduction

Emotions are a key component of human behaviour and communication. Most often, we use language to manifest, transmit and explain emotions. Meanwhile, the continuously increasing popularity of social media produces unprecedented amounts of user-generated content from people all around the world and in all languages. Oftentimes, this content (posts, comments, descriptions etc.) includes words that reveal the scope of emotions the author tries to unveil while evoking specific emotions from the reader as well. The social media era has also introduced very open outbursts of socially unacceptable discourse (SUD), such as hate, discriminatory, offensive or threatening speech. This has given rise to the necessity of analysing SUD communication practices in order to better understand and effectively tackle them. Here we dive into the field of Emotion Recognition (ER), which aims to recognize and categorize verbalized emotions in texts. By doing so, we hope to understand what outlines SUD content through the viewpoint of emotions.

In this paper, we introduce a novel, yet simple method to analyse emotions in text by utilizing the NRC Emotion Lexicon (Mohammad and Turney, 2010). A metric which we name Emotional Charge (E_c), calculates the overall emotion intensity of a comment. We utilize our approach on social media content by answering how emotions depend regarding the language, topic and most importantly, its SUD contribution.

For that purpose, we focus on emotions expressed in socially unacceptable Facebook comments in Slovene and English on the topics of the European migrant crisis (hereinafter referred to as Migrants) and LGBT+ rights (hereinafter referred to as LGBT+) from the FRENK data-

set (Ljubešić et al., 2021), as it is a uniquely carefully annotated multi-lingual dataset on SUD content which covers two topics. We perform a quantitative analysis for both languages and topics, taking into account the degree of emotional charge in each comment and the representation of individual categories of emotions by using the NRC emotion lexicon, which organizes words into one of the eight basic emotions by Plutchik (1980). To complement the quantitative approach, a qualitative analysis of the emotional words in SUD comments is added, enabling a more thorough understanding of the emotional charge findings.

The two main research questions covered in this paper are:

- Does SUD content differ from non-SUD content in the expression of emotions?
- Does the emotional footprint of SUD comments differ depending on the topic and target they address?

The paper is organised as follows. Section 2 gives an overview of the background and related work; Section 3 focuses on the description of the dataset used and describes the methods for calculating the emotional charge of comments. Subsequently, Section 4 presents the analysis on the emotional landscape of SUD, both from a statistical point of view and from a qualitative angle giving a deeper look at the emotional lexicon connected to SUD. Finally, Section 5 concludes the paper with a discussion and ideas for future work.

2 Background and related work

In the past decade, there has been an increase in research in the field of automatic detection of emotions in user-generated content (Alm et al., 2005; Al-Saqqa et al., 2018). However, although SUD has been intensively analysed in various disciplines and methodological frameworks, approaches to SUD via emotion analysis has so far received little attention (Gitari et al., 2015; Martins et al., 2018). This article presents an approach to comprehensively analyse SUD with the help of emotion lexica, as Markov et al. (2021) showed that emotion-based features provide useful cues for its automatic detection. In this section, we present the theoretical underpinnings for the analytical part of our study.

2.1 Emotions

In psychology, there is no general unanimity on the definition of emotions and their number. Research mostly focuses on two approaches to the representation of emotions, namely the category model and the dimensional model (Scherer, 2005). In the category model, emotions are presented as sets of different basic emotional states (e.g., JOY, ANGER) where basic emotions are understood as those that appear in very early childhood development and their expression and recognition are culturally independent. In the dimensional approach, emotions are presented in the space where each emotion occupies its place in an emotion dimension (e.g., value dimension: positive-negative axis, strength axis: high-low; Russell, 1980). The categorical approach is more widespread in computational linguistics than the dimensional one (Aman and Szpakowicz, 2007; Ghazi, 2016) because it is more intuitive and easier to apply, especially in computational models, which is why we adopt it in the study presented in this paper. We use the categorization into 8 basic emotions according to Plutchik (1980), namely JOY, SADNESS, ANGER, FEAR, TRUST, DISGUST, SURPRISE and ANTICIPATION, as they represent the basic and prototypical emotions, with the combination of which we can build more complex ones, e.g., LOVE, AWE, CONTEMPT. This model is also called Plutchik's wheel of emotions, as each fundamental emotion also has its opposite emotion (e.g., JOY - SADNESS, FEAR - ANGER; Plutchik, 2001; see Figure 1). Their organisation is based on the physiological purpose of each.

Martins et al. (2018) show that the most critical emotions to identify hate speech are the negative ones – ANGER, DISGUST, FEAR AND SADNESS as they occur in 2/3 of hate speech texts, while they claim SURPRISE can be interpreted as a neutral emotion in hate speech. On the other hand, ANTICIPATION, JOY and TRUST can be classified in the positive emotions group.

2.2 Emotion Recognition from Text

Emotion recognition from text can be divided into two groups: the earlier approaches, based on lexical datasets (Mohammad and Turney, 2010), and the latter ones, based on annotated training corpora (Aman and Szpakowicz, 2007; Canales et al., 2019). In the corpus approach,

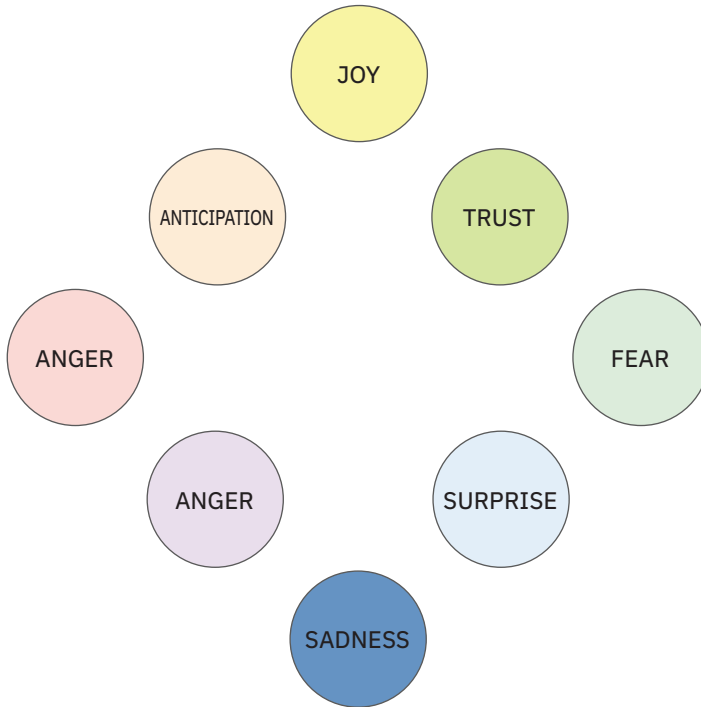


Figure 1: PLUTCHIK'S WHEEL OF EMOTIONS. It shows 8 basic emotions: joy, trust, fear, surprise, sadness, anticipation, anger, and disgust.

machine learning methods based on pre-annotated texts are employed to develop models for annotating new texts, while the lexical approach for identifying emotions in texts uses an external set of vocabulary with emotion tags. Due to the greater universality and adaptability to different domains and genres, we follow the latter paradigm. Additionally, previous research confirms the adequacy of the lexical approach. Mohammad and Yang (2011) have successfully used the NRC Emotion Lexicon to identify predominant emotions in love letters, hate emails, and suicide records. They were mainly interested in the difference between the linguistic expression of emotions in men and women. A similar approach with the help of the SENTIWORDNET lexical database, which contains strings of synonyms with assigned sentiment tags, was successfully used by Denecke (2008) on machine-translated texts to predict sentiment.

2.3 Socially Unacceptable Discourse (SUD)

Hate speech is a widespread phenomenon that attracts many researchers from diverse areas. However, the term is usually used in a very narrow, legally defined sense in the literature, which is why we adopt the term Socially Unacceptable Discourse (SUD), comprising all forms of hateful, discriminatory, offensive, violent or threatening speech (Fišer et al., 2017). A significant part of contemporary SUD research takes place within critical discourse analysis in combination with corpus linguistics (cf. Brindle, 2016; Knoblock, 2017), intending to identify and analyse SUD and its evolution. Assimakopoulos et al. (2017) present several European research projects on SUD, in which the analysis of online content is predominant. They point out that EU legislation alone is not enough to solve the spread of online hate and improve its understanding, as SUD can manifest itself in many subtle ways, such as stereotyping and categorization, patriotism, metaphorical expression, sarcasm, allusions etc., which makes comprehensive linguistic approaches extremely important for better awareness of the issue. A deeper understanding of SUD would mean principally better prevention and identification.

In Slovenia, the most valuable resource of SUD data is the manually annotated FRENK corpus of Facebook comments (Ljubešić et al., 2019; Ljubešić et al., 2021; see Section 3). Vehovar et al. (2020) show that about half of all the comments appearing in the FRENK dataset were identified as SUD: the share is significantly higher for the topic of Migrants (58%) than for the topic of LGBT+ (48%). The dataset was also analysed from the linguistic point of view, revealing SUD has a different lexical footprint (Franza and Fišer, 2019) and showing SUD comments are less standard than non-SUD comments with also a lower frequency of emoticons/emojis and punctuation (Pahor de Maiti et al., 2019).

3 Dataset and methodology

In this section, we describe in detail how the FRENK dataset, which is also used in this paper, was constructed and how it was processed for the purposes of this analysis. Next, we present the NRC Emotion Lexicon and the emotion labels it contains. Based on these, the emotional

charge of each comment in the FRENK dataset is calculated, which is presented in the final subsection.

3.1 FRENK Corpus

The FRENK corpus (Ljubešič et al., 2019; Ljubešič et al., 2021)¹ was collected from Facebook pages of three mainstream news media outlets for each examined language, including Slovene and English. It covers two topics, the EU Migrants crisis and the LGBT+ rights, and was enriched with manual annotations of the comments (Ljubešič et al., 2019). The Slovene part of the corpus contains 30 posts with 6545 comments for Migrants, and 93 posts with 4571 comments for LGBT+. The English part of the corpus consists of 16 posts with 5855 comments for Migrants, and 14 posts with 5906 comments for LGBT+. Additionally, comments were annotated for the type of SUD they produce (acceptable, background-violence, background-offence, other-threat, other-offence and unacceptable), as well as a categorization of the people being the target of the comment (Migrants, members of the LGBT+ community, persons related to Migrants or LGBT+, journalists or media, fellow commenter, other).

The dataset is linguistically processed with the CLASSLA pipeline for Slovene (Ljubešič, 2019, 2020) and Stanza for English (Peng Qi et al., 2020) on the levels of tokenization and sentence splitting, PoS-tagging and lemmatization. Therefore, we were able to annotate the lemmatized English and Slovene datasets with the NRC Emotion Lexicon for the corresponding language, which resulted in a bilingual and comparable emotion-labelled dataset of SUD Facebook comments that we analyse in the remainder of this paper.

3.2 Emotion Annotation

To identify emotions, we used the NRC Emotion Lexicon. The lexicon contains all words from Roget’s Thesaurus that appear more than 120,000 times in Google’s n-gram corpus, resulting in 14,200 entries.

1 The FRENK corpus, besides its Slovene and English parts, was created also for Croatian, French and Dutch, <http://nl.ijs.si/frenk/english>. The Dutch version was created within the Li-LaH project: <https://lilah.eu>.

Each word in the lexicon has a label for its polarity (positive, negative) and for Plutchik’s 8 basic emotions (ANGER, ANTICIPATION, DISGUST, FEAR, JOY, SADNESS, SURPRISE, TRUST). It was annotated manually using the crowdsourcing platform Amazon Mechanical Turk. The lexicon was originally created for English, and was later also automatically translated into 105 languages, including Slovene, with the help of Google Translate (2017). We have performed manual post-editing of the machine-translated lexicon (Daelemans et al., 2020). Examples of the translated lexicon along with the emotion labels can be found in Table 1.

Table 1: Examples of emotion annotation in the NRC Emotion Lexicon

English	abandoned	happiness	wise	ghost	refugee
Slovene translation	opuščen, zapuščen, prekinjen, zavržen	sreča, veselje	moder	duh, prikazen	begunec
ANGER	Yes	No	No	No	No
ANTICIPATION	No	Yes	No	No	No
DISGUST	No	No	No	No	No
FEAR	Yes	No	No	Yes	No
JOY	No	Yes	No	No	No
SADNESS	Yes	No	No	No	Yes
SURPRISE	No	No	No	No	No
TRUST	No	Yes	No	No	No

Note. For each English entry, there is a manually post-edited machine translation in Slovene and annotations for each of the 8 basic emotions.

3.3 Lexicon Limitations

The lexical approach is an efficient method to tackle emotion recognition from text (cf. 2.2). It is essential to work with datasets that are carefully prepared and verified to have reliable results. Our approach in this paper tests this method and achieves interesting outcomes. Nonetheless, it is important to also state the limitations of this specific emotion lexicon, the NRC emotion lexicon (Mohammad and Turney, 2010). We identified two main issues, namely the presence of biases and questionable emotion labelling.

Our work focuses on SUD, and it is important to point out that the lexicon has non-neutral annotations for the two topics we are dealing with, which can be linked to the lack of control and documentation about who the annotators were in the first place as the lexicon is the result of crowdsourcing. For example, *immigrant* is annotated with FEAR, *fugitive* with FEAR, ANGER, DISGUST, SADNESS and TRUST, *lesbian* with DISGUST and SADNESS. It is possible to note that there are some prejudices in these labels and it could be problematic as our work aims to fight against biases. Moreover, some labels appear to be ambiguous. For example, *nurture* is annotated with ANGER, ANTICIPATION, DISGUST, FEAR, JOY and TRUST, which suggests contradictory emotions together and does not give an insightful perspective of the word.

There have been many attempts to create an emotion lexicon, but the NRC emotion lexicon attracted the most attention due to its availability, size, and its choice of Plutchik's expressive eight-class emotion model (Zad et al., 2021). This is also the reason why we decided to use it, but we will take into account the potentially problematic labels in our interpretation of the results and we will complement the analysis with a qualitative study to check for potentially problematic consequences of using the lexicon. There have been also several attempts to improve the NRC emotion lexicon (cf. Zad et al., 2021), which should be further explored in the future.

3.4 Lexicon Coverage

Table 2 shows statistics regarding the NRC lexicon coverage of our dataset, for each of the subsets. Lexicon coverage has been calculated as the percentage of unique emotionally eligible words found in the lexicon, which means not all of them are labelled with emotion tags. The English language subsets contain around 5000 to 9000 unique words, with the NRC lexicon coverage of 20%. Meanwhile, the Slovene subsets, although of similar size, contain more unique words, with around 6000 to 12,000, depending on the dataset. Expectedly, since the Slovene NRC lexicon is the result of the machine-translated English lexicon and has a generally higher number of unique words, the NRC lexicon coverage of the Slovene dataset is a bit lower (around 16%), with small differences depending on the subset. Manual examinations have shown that there is a small

number of false positives and a higher number of false negatives, implying that the lexicon should be further improved. A random (subjective) sample evaluation of 100 English and 100 Slovene comments on the performance of the lexicon revealed the following:

- English NRC lexicon: Precision – 0.96; Recall – 0.65
- Slovene NRC lexicon: Precision – 0.91; Recall – 0.64

The low recall for both English and Slovene shows that the lexicon fails to recognize a large portion of the emotional words present in the comments, which is expected as we focus on a very specific kind of discourse on social media with specific characteristics on a very narrow topic. It is possible to find an explanation for the low recall also in the false negative emotionally eligible words (emotional, but not covered by the lexicon), as for example *shootings*, *frightened*. Moreover, SUD comments exhibit a peculiar tendency towards nonstandard features (Pahor de Maiti et al., 2019), which compromises emotional words recognition, for example *strelat* instead of *streljati* (eng. *to shoot*). Additionally, the evaluation indicates a lower precision of the Slovene lexicon, which could be possibly explained because of more false positives (not emotional, but included in the lexicon), which are mainly due to polysemy and non-canonically spelled words. For example, the Slovene lexicon contains the adjective *sam* (eng. *alone*), but in the comments it is used as an adverb (meaning *just*), which should not be an emotional word.

Table 2: Statistics regarding the NRC Lexicon coverage of our dataset, divided per topic, language and SUD/non-SUD comments

Language	Topic	SUD/ non-SUD	Comments	Unique words	Emotionally eligible words (nouns, verbs, adjectives and adverbs)	Lexicon coverage
English	Migrants	Non-SUD	2964	8401	5291	1046 (20%)
English	Migrants	SUD	2867	9323	6818	1323 (19%)
English	LGBT	Non-SUD	1777	8514	5374	1124 (21%)
English	LGBT	SUD	4080	5622	4297	977 (23%)
Slovene	Migrants	Non-SUD	2646	8401	5889	863 (15%)
Slovene	Migrants	SUD	3795	12486	10020	1325 (13%)
Slovene	LGBT	Non-SUD	1855	6199	4745	878 (19%)
Slovene	LGBT	SUD	2606	10108	8392	1329 (16%)

3.5 Calculating Emotional Charge

The final stage is to use the lemmatized comments and the lexicon to calculate a metric that defines the overall emotion intensity. We introduce this metric in order to be able to compare comments not just on the level of a specific emotion, but also have a universal comparison which includes all, answering the questions posed in the Introduction.

We define Emotional Charge (E_C) as follows: let W be the list of all nouns, verbs, adjectives and adverbs in one comment (as the emotionally eligible word functions). Then, let W_E be the list of all words in W which are labelled as emotional by the emotion lexicon. We define emotional charge E_C of a comment as follows:

$$E_C = \frac{|W_E|}{|W|}$$

To put it simply, Emotional Charge (E_C) calculates the portion of emotional words labelled by the lexicon in the total number of emotionally eligible words. Note that W and W_E are defined as lists and not as sets, thus an emotional word being present twice in a comment is also counted twice in the total score.

Using the emotional charge of each comment, we were able to get a sampling distribution of emotional charge for the desired group of comments (e.g., SUD vs. non-SUD, Slovene vs. English, Migrants vs. LGBT+). Figure 2 shows an example of the procedure for calculating the emotional charge.

Taking into account only word types that can contain emotion (nouns, verbs, adjectives and adverbs) as well as using the emotional charge formula that normalizes comment length makes the emotional charge scores more robust. Yet, this way of calculating emotional charge introduces many “non-emotional” and “highly” emotional short comments, where the emotional charge is 0 or 1 respectively, based on only a few words. Thus, we made a pragmatic decision of excluding comments with less than three words from the rest of our analysis.

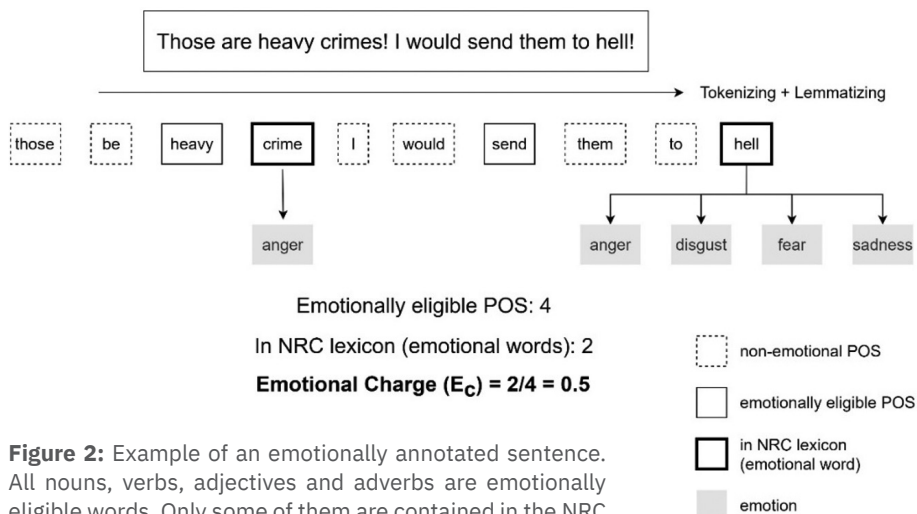


Figure 2: Example of an emotionally annotated sentence. All nouns, verbs, adjectives and adverbs are emotionally eligible words. Only some of them are contained in the NRC Emotion Lexicon. The ones in the Lexicon are counted in the emotional charge of the sentence.

4 Results

In this section, we present our research findings using the emotional charge of SUD comments for both Slovene and English language.

4.1 Emotional Charge Analysis

SUD is more emotional than non-SUD content. Here, we check whether emotion annotation is informative for differentiating between SUD and non-SUD comments by comparing their emotional charge. Once we calculated the distribution of emotional charge for each of the groups, we applied the Kolmogorov-Smirnov two-sample test (Pratt and Gibbons, 1981), which showed a statistical difference between SUD and non-SUD across all four combinations of language and topic (Migrants English $p=3 \times 10^{-7}$; $d=0.18$, LGBT+ English $p=3 \times 10^{-8}$; $d=0.23$, Migrants Slovene $p=1 \times 10^{-10}$; $d=0.18$ and LGBT+ Slovene $p=3 \times 10^{-4}$; $d=0.12$). The effect size d according to Cohen's formula (Cohen, 1988) is considered small to medium (depending on the combination). Figure 3 shows the distribution mean and deviation of all four combinations. Thus, we conclude that a specific analysis on SUD content could indeed be informative as the data showed that these comments are significantly more emotionally charged than non-SUD.

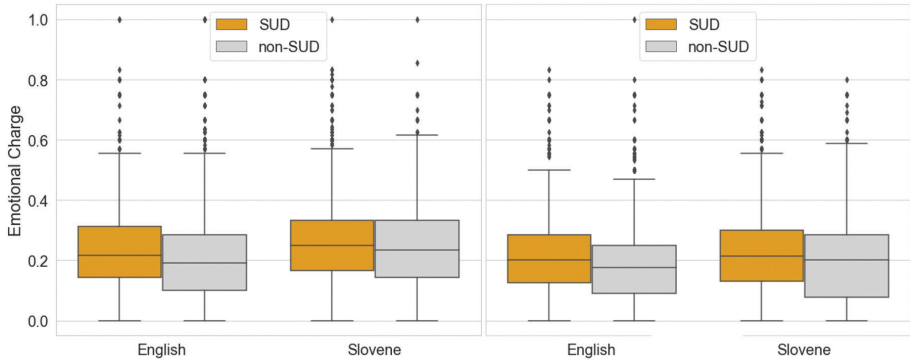


Figure 3: Comparison of emotional charge between languages and topics of SUD comments and non-SUD comments. The figure shows distributions (rectangles) and variance (lines), SUD comments are significantly more emotionally charged than non-SUD ones.

Topics differ in emotional charge – LGBT+ evokes more emotions than the Migrants topic. After confirming a higher emotional charge in SUD comments, we analysed whether one topic attracts more emotional charge than the other. Figure 4 shows, side-by-side, the distributions of the sets we compare (LGBT+ vs. Migrants). The Kolmogorov-Smirnov test suggests that there is a statistical difference in the emotional charge between Migrants and LGBT+, as both in English ($p=3\times 10^{-9}$) and Slovene ($p=2\times 10^{-11}$) comments, the LGBT+ topic carries a higher emotional charge. According to Cohen’s coefficient, in English, the effect size is medium ($d=0.209$) while in Slovene it is small ($d=0.183$).

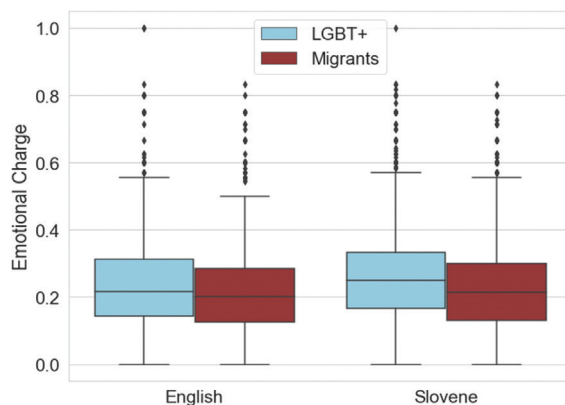


Figure 4: Difference of SUD Emotional charge between the LGBT+ and migrant topics in English and Slovene. The figure shows distributions (rectangles) and variance (lines), resulting in LGBT+ comments being more emotionally charged.

Comments are more emotional when targeted at Migrants/LGBT+.

One of the metadata information of the FRENK dataset is the target of the comment, or in other words, who the comment is directed at. We compared “*the commenter*” and “*the target – migrant/LGBT+ person*” which are the two most frequent targets in the FRENK dataset (see Section 3.1). The comments targeted at *migrants/LGBT+ persons* are explicitly aimed at migrants or members of the LGBT+ community, while the others are targeted at another *commenter* in the discussion thread. As shown in Figure 5, for both topics and languages, comments targeted at *migrants* or *LGBT+* are generally more emotional than comments targeted at *interlocutors* (fellow-commenters in the discussion thread).

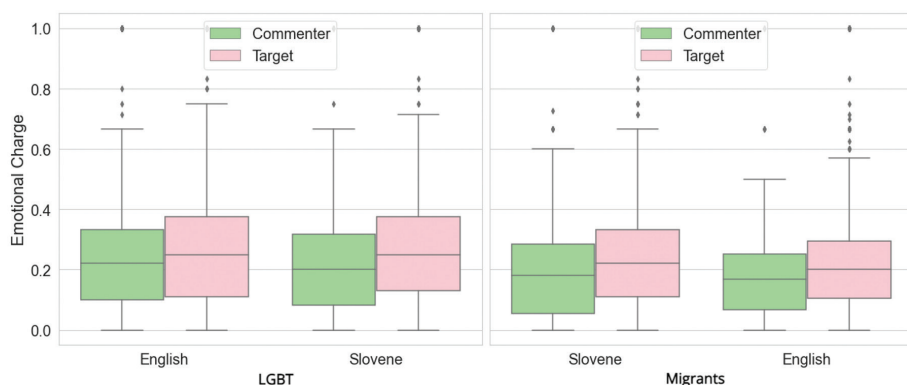


Figure 5: Comparison of emotional charge for different targets of the SUD comment, namely Commenter or Target (LGBT+ persons/migrants), between Slovene and English. The figure shows distributions (rectangles) and variance (lines).

Different topics provoke different emotions. In order to extract the data for a specific emotion, we calculated how much a specific emotion contributes to the total emotional charge. Then, by having the emotional charge distribution of each particular emotion, we were able to compare their manifestation for the two different topics: Migrants and LGBT+. On average, we observed that in English comments users manifest more *DISGUST* and *JOY* for the topic of LGBT+, and more *SADNESS*, *FEAR* and *SURPRISE* for Migrants (Figure 6). In Slovene, users manifest significantly more *ANTICIPATION* and *JOY* for LGBT+, while for the topic of Migrants, they manifest more *ANGER* and *FEAR* (Figure 6). It is interesting to observe that in both languages the LGBT+ topic invokes more

JOY, while the Migrants topic invokes more FEAR. It is also quite evident that emotions are not homogenous for all four subset combinations, with TRUST and FEAR being the most dominant emotions with more than 15% of the total emotional spectrum, while SURPRISE is the least present, taking less than 5% of the emotional spectrum.

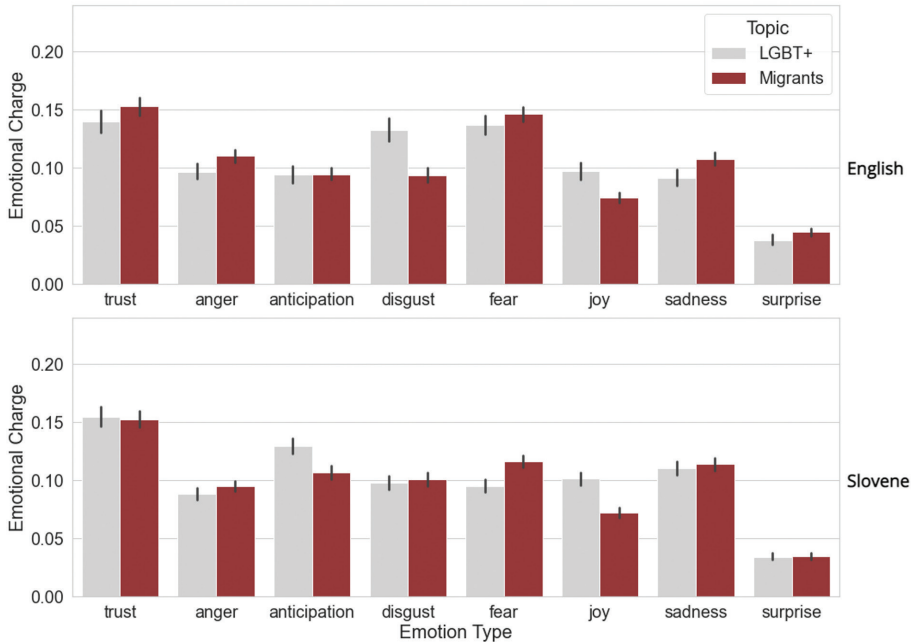


Figure 6: Distribution of emotions in English and Slovene SUD comments for the topics of LGBT+ and Migrants. The figure shows averages (bars) and their confidence intervals (95%) for each emotion present in the NRC Lexicon.

4.2 Emotional Words Analysis

In order to better understand the above quantitative analysis and have a closer look at the investigated data, we performed a qualitative analysis of the emotional words in the corpus.

Table 3 shows the three most frequent emotional words for each language, topic and emotion with the purpose of understanding which words are most commonly connected to which emotions. Some differences and similarities among topics and languages are observed. For the LGBT+ topic, the English commenters seem to be more religion-oriented, frequently using words such as *God*, *disgusting* and *sin*.

Meanwhile, the Slovene take the discussion to a more family-oriented field, using words such as *mother*, *child* and *nurture*. This could be due to the referendum in Slovenia for legalising same-sex marriage that took place in the same period as the data was harvested and has heavily influenced the discussions under Facebook posts by the Slovene media on this topic at the time, where people against framed their arguments around the notion of the traditional family unit, expressing concern with children's rights and same-sex couples' adoptions. Interesting enough, the roles are reversed for the Migrants topic, as now it is the Slovene commenters who are more concerned with religion, using words such as *religion* and *God*, possibly showing fear of a different religion. On the other hand, the English commenters seem to feel more physically threatened by the migrants, using words such as *fight*, *kill* and *idiot*. This could be due to the unprecedented migrant wave through the Balkan route that took place in the same period as the data was harvested and has heavily influenced the discussions under Facebook posts by the Slovene media on this topic at the time as Slovenia has never before experienced anywhere near this rate of the migrant influx, while the topic has been present in the UK political and public debates for many decades.

Individual emotions across languages exhibit mostly similar concepts, yet there are some differences. For example, for the Migrants topic, both English and Slovene commenters use similar words to express emotions, in particular, both groups express DISGUST with the word *terrorist* and show FEAR with *immigrant/fugitive*. On the other hand, English commenters express ANGER for the LGBT+ topic in different terms than the Slovene ones. The English commenters use words such as *disgusting*, *hate* and *sin*, taking the attitude that being LGBT+ is sinful and repulsive, whereas the Slovene ones show hostility towards the target and, once again, concern regarding children with expressions such as *violence*, *nurture* and *against*.

Expectedly, in both languages, word usage varies depending on the topic. For example, English commenters express SADNESS with the word *problem* for both the Migrants and the LGBT+ topic, suggesting they perceive both topics as an issue. Yet, Slovene commenters show DISGUST differently for the two topics, exposing what bothers them most

for each. As expected, for the Migrants topic words such as *fugitive*, *terrorist* and *back* occur frequently, while for the LGBT+ topic *gay*, *garbage* and *nurture* are recurring.

Table 3: Top three most frequent emotional words in SUD comments for each emotion, divided per language and topic (Slovene words have their translation after the dash)

	ENG Migrants	ENG LGBT+	SLO Migrants	SLO LGBT+
ANGER	fight (4.45%) hate (3.12%) money (2.96%)	disgusting (4.22%) hate (3.95%) sin (2.98%)	begunec – fugitive (11.84%) proti – versus (3.73%) terrorist – terrorist (2.73%)	proti – against (5.29%) nasilje – violence (2.82%) vzgjajati – nurture (2.79%)
ANTICIP.	child (7.7%) good (4.71%) time (4.57%)	God (15.78%), marriage (6.56%) sex (6.25%)	otrok – child (7.16%) vera – religion (6.08%) svet – world (4.60%)	otrok – child (25.28%) zakon – marriage (6.06%) svet – world (3.14%)
DISGUST	hate (4.32%) idiot (3.92%), terrorist (3.27%)	disgusting (4.45%) sick (4.23%) hate (4.16%)	begunec – fugitive (14.31%) nazaj – back (7.20%) terrorist – terrorist (3.30%)	peder – gay (5.11%) smeti – garbage (3.52%) vzgjajati – nurture (3.31%)
FEAR	problem (4.80%), immigrant (4.56%) war (3.99%)	God (12.24%) disgusting (2.99%) hate (2.80%)	begunec – fugitive (9.75%) vojna – war (4.07%) bog – God (3.00%)	nasilje – violence (2.64%) vzgjajati – nurture (2.61%) bog – God (2.30%)
JOY	child (8.72%) good (5.34%) money (3.82%)	God (15.15%) love (8.85%) marriage (6.30%)	otrok – child (9.01%) vera – religion (7.65%) bog – God (4.50%)	otrok – child (28.10%) zakon – marriage (7.00%) mama – mother (3.17%)
SADNESS	problem (6.60%) kill (4.63%) leave (4.20%)	sick (4.36%) hate (4.29%) problem (3.23%)	sam – alone (11.57%) begunec – fugitive (10.76%) brez – without (3.47%)	sam – alone (7.73%) peder – gay (3.88%) mama – mother (3.30%)
SURPRISE	good (8.68%) leave (8.67%) money (6.37%)	good (9.6%) Trump (6.72%) marry (6.23%)	dober – good (8.38%) terrorist – terrorist (6.51%) lep – beautiful (5.76%)	dober – good (6.59%) dobro – cool (4.73%) lep – beautiful (4.65%)
TRUST	good (3.16%) show (3.13%) religion (3.11%)	God (12.14%) marriage (5.05%) sex (4.81%)	begunec – fugitive (8.20%) vera – religion (4.27%) svet – council (3.23%)	zakon – marriage (5.60%) pravica – right (4.29%) svet – world (2.90%)

Note. Percentages show the absolute frequency of the word with respect to all the words of the specific emotion. E.g., the word fight covers 4.45% of all anger words for the ENG Migrants dataset.

5 Conclusions

In this paper, we have presented a quantitative analysis of emotions in SUD comments in order to obtain an insight into the emotional footprint of this type of discourse. Applying the NRC Emotion Lexicon, we developed a novel metric named Emotional Charge of the comments to analyse SUD. We implemented this simple, yet effective methodology on the most relevant SUD multilingual dataset which also contains

Slovene data, namely the FRENK dataset, which comprises Facebook comments to posts related to the LGBT+ and the Migrants topic. We showed that SUD comments are more emotional than non-SUD. We also presented how emotions differ depending on the topic. For example, according to the emotion lexicon, the LGBT+ topic invokes more JOY, while the Migrants topic invokes more FEAR. When comparing the emotional charge of SUD comments depending on its target, we observed that comments are more emotional when a user targets the group (LGBT+ or Migrants) compared to a fellow commenter they are having an argument with. Furthermore, we also performed a qualitative analysis of the emotional words, which showed some trends in their usage depending on the topic and language. Slovene commenters to LGBT+ posts are very much concerned with children's wellbeing, while the English ones tend to manifest their opposition and disgust. For the Migrants topic, there is a common tendency in both languages of expressing the same emotion with similar words (e.g., DISGUST – *terrorist*; FEAR – *fugitive/immigrant*).

An original contribution of this study is its demonstration of the methodological potential of the lexical approach for identifying emotions in SUD, which has not been used in the Slovene context yet. The research presented in this paper complements international literature in this domain with the use of richly annotated corpora, emotion lexica and quantitative measures, while also adding a qualitative analysis.

The metric of measuring emotional intensity we have proposed in this paper has proved to be useful and insightful in our research, yet its simplicity could potentially oversimplify the highly complex problem of expressing emotions on social media which transcends linguistic expression and is not only highly context-dependent but is also very culturally nuanced, a common shortcoming of lexicon-based approaches. This is why we propose to experiment with context-aware models and metrics in future work that will better be able to take into account the complexity of this type of communication. We also stress the need for in-depth qualitative sociolinguistic analysis to always complement quantitative and automated approaches that will not only critically evaluate the quantitative approaches of such complex and sensitive phenomena but will also ensure that all relevant aspects of the com-

munication reality are considered before interpreting the results, drawing conclusions and making policy recommendations.

Acknowledgments

The work described in this paper was funded by the Slovenian Research Agency within the national research project »Resources, methods, and tools for the understanding, identification, and classification of various forms of socially unacceptable discourse in the information society« (J7-8280, 2017–2019), the Slovenian-Flemish bilateral basic research project »Linguistic landscape of hate speech on social media« (N06-0099, 2019–2023), the national research programme »Slovene Language – Basic, Contrastive, and Applied Studies« (P6-0215) and the national research programme »Digital Humanities: Resources, Tools and Methods« (P6-0436).

References

- Alm, C., Roth, D., & Sproat, R. (2005). Emotions from Text: Machine Learning for Text-based Emotion Prediction. *Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, October 2005, Vancouver, Canada* (pp. 579–586). Association for Computational Linguistics. doi:10.3115/1220575.1220648
- Al-Saqqa, S., Abdel-Nabi, H., & Awajan, A. (2018). A survey of textual emotion detection. *8th International Conference on Computer Science and Information Technology (CSIT), July 2018* (pp. 136–142). doi: 10.1109/CSIT.2018.8486405
- Aman, S., & Szpakowicz, S. (2007). Identifying Expressions of Emotion in Text. In V. Matoušek & P. Mautner (Eds.), *Text, Speech and Dialogue, SD 2007. Lecture Notes in Computer Science (Vol. 4629)* (pp. 196–205). Berlin, Heidelberg: Springer.
- Assimakopoulos, S., Baider, F. H., & Millar, S. (2017). *Online Hate Speech in the European Union. A Discourse-Analytic Perspective*. Cham: Springer International Publishing.
- Brindle, A. (2016). *The Language of Hate. A Corpus Linguistic Analysis of White Supremacist Language*. London and New York: Routledge.
- Canales, L., Daelemans, W., Boldrini, E., & Martinez-Barco, P. (2019). EmoLabel: Semi-Automatic Methodology for Emotion Annotation of Social Media

- Text. *IEEE Transactions on Affective Computing*. Retrieved from <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&number=8758380>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge.
- Daelemans, W., Fišer, D., Franza, J., Kranjčič, D., Lemmens, J., Ljubešič, N., Markov, I., & Popič, D. (2020). *The LiLaH Emotion Lexicon of Croatian, Dutch and Slovene*. Slovenian language resource repository CLARIN.SI. <https://www.clarin.si/repository/xmlui/handle/11356/1318>
- Denecke, K. (2008). Using SentiWordNet for Multilingual Sentiment Analysis. *Proceedings of the 24th International Conference on Data Engineering, 7–12 April 2008, Cancun, Mexico* (pp. 507–512).
- Fišer, D., Ljubešič, N., & Erjavec, T. (2017). Legal framework, dataset and annotation schema for socially unacceptable online discourse practices in Slovene. *Proceedings of the 1st Workshop on Abusive Language Online, ACL 2017, Vancouver, Canada* (pp. 46–51). Association for Computational Linguistics. doi: 10.18653/v1/W17-3007
- Franza, J., & Fišer, D. (2019). The lexical inventory of Slovene socially unacceptable discourse on Facebook. *Proceedings of the 7th Conference on Computer-Mediated Communication (CMC) and Social Media Corpora, CMC-Corpora 2019, Cergy-Pontoise, France*. Retrieved from <https://hal.archives-ouvertes.fr/hal-02292616/document#page=50>
- Ghazi, D. (2016). *Identifying Expressions of Emotions and Their Stimuli in Text*. PhD dissertation. Canada: University of Ottawa.
- Gitari, N. D., Zuping, Z., Hanyurwimfura, D., & Long, J. (2015). A Lexicon-based Approach for Hate Speech Detection. *International Journal of Multimedia and Ubiquitous Engineering (Vol. 10, No.4)* (pp. 215–230).
- Knoblock, N. (2017). Xenophobic Trumpeters: A corpus-assisted discourse study of Donald Trump's Facebook conversations. In A. Musolf (Ed.), *Journal of Language Aggression and Conflict (Vol. 5, No.7)* (pp. 295–322). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Ljubešič, N. (2019). *The CLASSLA-StanfordNLP model for morphosyntactic annotation of standard Slovenian*. Ljubljana: Slovenian language resource repository CLARIN.SI. Retrieved from <http://hdl.handle.net/11356/1251>
- Ljubešič, N. (2020). *The CLASSLA-StanfordNLP model for lemmatisation of standard Slovenian 1.1*, Slovenian language resource repository CLARIN.SI. <http://hdl.handle.net/11356/1286>
- Ljubešič, N., Fišer, D., & Erjavec, T. (2019). *The FRENK datasets of Socially Unacceptable Discourse in Slovene and English*. International Conference on

- Text, Speech, and Dialogue. Springer, Cham. doi: 10.1007/978-3-030-27947-9_9
- Ljubešić, N., Fišer, D., Erjavec, T., & Šulc, A. (2021). *Offensive language dataset of Croatian, English and Slovenian comments FRENK 1.1*. Ljubljana: Slovenian language resource repository CLARIN.SI. Retrieved from <http://hdl.handle.net/11356/1462>
- Markov, I., Ljubešić, N., Fišer, D., & Daelemans, W. (2021). Exploring Stylo-metric and Emotion-Based Features for Multilingual Cross-Domain Hate Speech Detection. *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* (pp. 149–159). Association for Computational Linguistics. Retrieved from <https://aclanthology.org/2021.wassa-1.16/>
- Martins, R., Gomes, M., Almeida, J. J., Novais, P., & Henriques, P. (2018). Hate Speech Classification in Social Media Using Emotional Analysis. *7th Brazilian Conference on Intelligent Systems (BRACIS), 22–25 October 2018, Sao Paulo, Brazil* (pp. 61–66). doi: 10.1109/BRACIS.2018.00019
- Mohammad, S., & Yang T. (2011). Tracking Sentiment in Mail: How Genders Differ on Emotional Axes. *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2.011)* (pp. 70–79). Portland, Oregon: Association for Computational Linguistics.
- Mohammad, S., & Turney, P. D. (2010). Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon. *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, June 2010, Los Angeles, California* (pp. 26–34).
- Pahor de Maiti, K., Fišer, D., & Ljubešić, N. (2019). How haters write: analysis of nonstandard language in online hate speech. *Proceedings of the 7th Conference on Computer-Mediated Communication (CMC) and Social Media Corpora, CMC-Corpora, 9–10 September 2019, Cergy-Pontoise, France*. Retrieved from <https://hal.archives-ouvertes.fr/hal-02292616/document#page=44>
- Peng Q., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020). *Stanza: A Python Natural Language Processing Toolkit for Many Human Languages*. Retrieved from <https://arxiv.org/abs/2003.07082>
- Plutchik, R. (1980). *Emotion: Theory, research and experience*, 1. Academic Press.
- Plutchik, R. (2001). The Nature of Emotions: Human Emotions Have Deep Evolutionary Roots, a Fact That May Explain Their Complexity and Provide Tools for Clinical Practice. *American Scientist* 89(4), 344–350.

- Pratt, J. W., & Gibbons, J. D. (1981). Kolmogorov-Smirnov two-sample tests. *Concepts of nonparametric theory*. Springer, New York, NY. 318–344.
- Russell, J. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161–1178. doi: 10.1037/h0077714
- Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, 44(4), 695–729. doi: 10.1177/0539018405050582
- Vehovar, V., Povž, B., Fišer, D., Ljubešić, N., Šulc, A., & Jontes, D. (2020). Družbeno nesprejemljivi diskurz na Facebookovih straneh novičarskih portalov. *Teorija in Praksa*, 57(2), 622–645.
- Zad, S., Jimenez, J., & Finlayson, M. A. (2021). Hell Hath No Fury? Correcting Bias in the NRC Emotion Lexicon. *Proceedings of the 5th Workshop on On-line Abuse and Harms, 6 August 2021, Bangkok, Thailand* (pp. 102–111). Retrieved from <https://aclanthology.org/2021.woah-1.pdf>

Analiza čustev v družbeno nesprejemljivem diskurzu

Besedila pogosto izražajo avtorjevo čustveno stanje in pokazalo se je, da imajo informacije o čustvih potencial za odkrivanje in analizo sovražnega govora. V prispevku predstavljamo kvantitativno metodologijo analize čustev v besedilu. Na podlagi leksikona čustev NRC Emotion Lexicon in Plutchikovega modela osmih osnovnih čustev smo definirali preprosto, a učinkovito metodo za odkrivanje čustvene zaznamovanosti besedila. Z navedeno metodologijo smo raziskali čustveno zaznamovanost besedil, označenih kot družbeno nesprejemljivi diskurz (DND), ki predstavlja izrazito in potencialno škodljivo vrsto besedila ter se dandanes hitro širi na družbenih omrežjih. Metodo čustvene zaznamovanosti smo aplicirali na korpus komentarjev s Facebooka. Primerjavo in analizo smo izvajali na štirih zbirkah podatkov v dveh jezikih, in sicer v angleščini in slovenščini, ter na dveh temah, pravice LGBT+ skupnosti in evropska migrantska kriza. Ugotovili smo, da je vsebina DND komentarjev bistveno bolj čustvena od tistih, ki ne vsebujejo DND. Poleg tega smo pokazali razlike v izražanju čustev glede na jezik, temo in tarčo komentarjev. Izsledke kvantitativne metodologije analize čustev smo podprli s kvalitativno analizo korpusa, kjer smo preučili najpogostejše čustveno zaznamovane besede, povezane z vsakim čustvom v vseh štirih zbirkah podatkov. Ugotovili smo, da se čustveno zaznamovane besede v DND bistveno razlikujejo glede na temo, medtem ko obstaja med jeziki precejšnje prekrivanje.

Ključne besede: čustva, družbeno nesprejemljivi diskurz (DND), sovražni govor, družbena omrežja, korpusi

Raba jezikovnih virov in *Kolokacijskega slovarja* *sodobne slovenščine* med študenti prevajalstva

Nataša HIRCI

Filozofska fakulteta, Univerza v Ljubljani

Martin Anton GRAD

Filozofska fakulteta, Univerza v Ljubljani

V prispevku se posvečamo študiji, ki predstavlja uporabniško izkušnjo in odnos študentov prevajalstva na Univerzi v Ljubljani do jezikovnih virov in prevajalskih pripomočkov, saj se v času informacijske družbe od bodočih prevajalcev pričakuje, da so v kar največji meri digitalno osveščeni, predvsem pa seznanjeni z relevantnimi viri informacij. Študija se v prvem delu osredotoča na odnos študentov do angleških in slovenskih jezikovnih virov in na strategije reševanja prevajalskih zagat pri prevajanju kolokacij ter rabo različnih relevantnih jezikovnih virov. Raziskava je v drugem delu osredinjena na opazovanje prevajalskega procesa in rabo *Kolokacijskega slovarja sodobne slovenščine* (KSSS) pri prevajanju izbranih kolokacij iz angleščine v slovenščino. Rezultati študije so pokazali, da se študenti zavedajo prednosti reševanja prevodnih težav s pomočjo relevantnih jezikovnih virov, saj se je strategija vključevanja kolokacijskega slovarja v prevajalski proces izkazala za relativno uspešno. Vendar pa je obenem pokazala, da je že v času študija nujno dodatno osveščati študente

Hirci, N., Grad, M. A.: Raba jezikovnih virov in Kolokacijskega slovarja sodobne slovenščine med študenti prevajalstva. Slovenščina 2.0, 10(1): 23–64.

1.01 Izvirni znanstveni članek / Original Scientific Article

DOI: <https://doi.org/10.4312/slo2.0.2022.1.23-64>

<https://creativecommons.org/licenses/by-sa/4.0/>



o naprednih funkcijah tega jezikovnega vira, predvsem pa spodbujati iskanje po relevantnih virih in kritično ovrednotenje najdenih informacij z zavedanjem, da je v prevajalskem procesu nujno poiskati takšne prevodne rešitve, ki so v danem kontekstu najustreznejše.

Ključne besede: uporabniška študija, raba jezikovnih virov in prevajalskih pripomočkov, prevajanje kolokacij, *Kolokacijski slovar sodobne slovenščine*, osveščanje študentov prevajalstva

1 Uvod

V zadnjem času v Sloveniji sicer opažamo porast uporabniških študij, ki vključujejo različne uporabniške skupine, od učiteljev jezika in novinarjev do raziskovalcev in prevajalcev (prim. Arhar Holdt, 2020, 2021; Arhar Holdt idr., 2019; Pori idr., 2020, 2021; Šorli in Ledinek, 2017), vendar pa je področje uporabe slovarskih virov in drugih jezikovnih priročnikov izključno med prevajalci zaenkrat podhranjeno z uporabniškimi raziskavami, saj ima ta skupina dokaj specifične potrebe. Še manj je uporabniških študij med bodočimi prevajalci, torej študenti prevajalstva.

V prispevku zato želimo nasloviti uporabniško izkušnjo študentov Oddelka za prevajalstvo v Ljubljani in njihov odnos do rabe različnih jezikovnih virov, priročnikov in prevajalskih pripomočkov, s posebnim poudarkom na prevodnih rešitvah izbranih kolokacij in rabi *Kolokacijskega slovarja sodobne slovenščine* (KSSS)¹ med prevajalskim procesom. Uporabniška študija je osredinjena na študente prevajalstva, saj se v času informacijske družbe od prevajalcev pričakuje, da so v kar največji meri digitalno osveščeni, predvsem pa seznanjeni z relevantnimi viri informacij. Na Oddelku za prevajalstvo bodoče prevajalce skušamo ustrezno usposobiti za delo na prevajalskem trgu, zato si prizadevamo, da bi bili naši alumni večji uporabe raznovrstnih digitalnih jezikovnih virov in prevajalskih orodij, saj je to bistvenega pomena za njihovo nadaljnje profesionalno prevajalsko delo.

Ker pa je raziskav, ki so ciljno usmerjene v uporabniško izkušnjo študentov prevajalstva relativno malo, želimo s svojo študijo vsaj delo-

1 <https://viri.cjvt.si/kolokacije/slv/#>; podrobnejši opis slovarja v Kosem idr., 2018a; glej tudi Kosem (2018b, 2019), Arhar Holdt (2021) in Pori idr. (2020, 2021).

ma zapolniti to vrzel in s tem prispevati k spodbujanju k bolj osveščeni rabi jezikovnih virov med to populacijo že v času izobraževalnega procesa. Odgovoriti skušamo na vprašanja o rabi slovarskih virov in drugih jezikovnih priročnikov, ki se jih poslužujejo študenti prevajalstva pri prevajanju kolokacij, in na kakšne načine bodoči prevajalci v svoj prevajalski proces umeščajo rabo KSSS². Slovarji kolokacij so namreč lahko zelo uporaben pripomoček pri ustvarjanju besedila, torej pri jezikovni produkciji, še posebej, ko skušamo besedilo obogatiti in se želimo izogniti pretiranemu ponavljanju. Prav zaradi tega so lahko dobrodošel vir informacij tudi pri prevajalskem delu.

2 Kratek pregled področja

2.1 O rabi jezikovnih virov in drugih prevajalskih pripomočkov

Področje uporabniških raziskav o rabi slovarjev in drugih jezikovnih virov je relativno bogato podprto z raziskavami o uporabi jezikovnih virov med študentsko populacijo pri usvajanju angleščine kot tujega jezika, nekaj se jih osredotoča tudi na bodoče prevajalce (prim. Béjoint, 1981, 2000; Campoy Cubillo, 2001; Mackintosh, 1998; Humblé, 2001; Roberts, 1992, 1997; Starren in Thelen, 1990; Varantola, 1998; Atkins in Varantola 1998; Livbjerg in Mees, 2003), medtem ko je v Sloveniji področje, ki se celovito ukvarja z uporabniškimi izkušnjami pri rabi slovarjev in drugih jezikovnih priročnikov med prevajalsko skupnostjo, relativno slabo raziskano.

Čibej idr. (2015, str. 169) pravijo, da so se dosedanje raziskave, ki proučujejo rabo virov v prevajalski skupnosti, ukvarjale predvsem z rabo slovarjev med profesionalnimi prevajalci (prim. Roberts, 1997; Atkins in Varantola, 1998; Varantola, 1998; Sánchez Ramos, 2005) ali z rabo slovarjev pri poučevanju prevajalcev med samim pedagoškim procesom (Roberts, 1992; Sánchez Ramos, 2005; Hirci, 2013). Avtorji obenem ugotavljajo, da je nujno proučevati tudi uporabnikove potrebe

2 Nekaj rezultatov dela te študije, vezanih na prevajanje kolokacij, je bilo sicer v manjšem obsegu že predstavljenih na konferenci *Jezikovne tehnologije in digitalna humanistika* (prim. Grad in Hirci, 2022), kjer smo naslovili uspešnost študentov pri reševanju dveh izbranih kolokacij, v pričujočem prispevku pa želimo bolj celostno prikazati rezultate uporabniške študije, ki so povezani predvsem z rabo KSSS.

in pričakovanja o rabi prav določenega jezikovnega vira, če naj bi bil ta »karseda široko uporaben in uporabniku prijazen« (Čibej idr., 2015, str. 169). Če to idejo zastavimo malce širše: da bi lahko imeli na voljo ustrezne in prevajalcem prijazne jezikovne vire in priročnike, je nujno poznati tudi potrebe in pričakovanja prevajalske skupnosti (Arhar Holdt, 2015; Arhar Holdt idr., 2016).

Da pri nas primanjkuje uporabniških študij, ki bi osvetljevale potrebe in zmožnosti uporabnikov pri rabi jezikovnih priročnikov, je bilo v preteklosti sicer večkrat opozorjeno (več o tem v Rozman, 2004; Stabej, 2009; Logar Berginc, 2009; Arhar Holdt, 2015). Vendar pa je v zadnjih letih – verjetno tudi v luči izpostavitve podhranjenosti tega raziskovalnega polja – prav to področje postalo raziskovalno dokaj živahno. Še posebej so v porastu uporabniške študije, ki so povezane z odzivnimi jezikovnimi viri, med katere spada tudi KSSS (prim. Pori idr., 2020, 2021; Arhar Holdt, 2020, 2021). Tudi Šorli in Ledinek (2017) poročata o uporabniških raziskavah, vendar ne zgolj med prevajalci, temveč tudi med drugimi profili, ki se primarno ukvarjajo z jeziki, npr. učitelji tujih jezikov, lektorji, novinarji in publicisti. Podobno velja za Mikolič (2015), ki se ukvarja z rabo jezikovnih virov med raziskovalci, oglaševalci in novinarji.

Čibej idr. (2015, str. 171) v prispevku, ki pa se primarno ukvarja s prevajalsko skupnostjo, predstavljajo raziskavo o poklicnih prevajalcih in njihovem iskanju pomoči pri prevajalskih zagatah na Facebook strani stanovske skupnosti *Prevajalci, na pomoč!*. Podobno velja tudi za Šorli (2018), ki se v svoji študiji o uporabi in potrebah prevajalcev in tolmačev naslanja na člane stanovskih združenj (tj. med prevajalskimi in tolmaškimi društvi ter prevajalskim sektorjem pri Generalnem sekretariatu Vlade RS), medtem ko Fišer (2008) in Kocijančič Pokorn (2016) orisujeta spremembe na prevajalskem trgu in s tem povezanim statusom profesionalnih prevajalcev v Sloveniji.

Večinoma se z uporabo jezikovnih virov ter rabo korpusnih virov med bodočimi prevajalci ukvarjajo predvsem raziskovalci in raziskovalke z Oddelka za prevajalstvo (prim. Hirci, 2007, 2009a; Mikolič Južnič, 2009; Pisanski Peterlin, 2003; Vintar, 1999), če najprej omenimo zgolj nekaj starejših. Med novejšimi uporabniškimi raziskavami, ki se osredotočajo na prevajalsko skupnost bodočih prevajalcev, velja omeniti raziskavo Gorjanca (2014), ki v svojem prispevku predstavlja pričakovanja

študentov prevajalstva o enojezičnih slovarjih (prim. tudi Arhar Holdt, 2018, kjer pa so poleg prevajalcev vključeni tudi drugi uporabniki enojezičnih slovarjev). Avtor poudarja, da so študenti prevajalstva sicer relativno dobro osveščeni glede sodobnih tehnoloških možnosti in so zato njihova pričakovanja precej visoka. Zagotovo so dokaj domači v digitalnem okolju, tako da se poslužujejo najrazličnejših možnosti, ki jih ponujajo spletna orodja. O podobnih rezultatih poroča tudi Hirci (2013) v primerjavi stanja o rabi jezikovnih virov med študenti prevajalstva v Ljubljani v dveh časovnih obdobjih, v letih 2005 in 2012. Iz rezultatov je razvidno, da je v slabem desetletju prišlo do precejšnje spremembe pri uporabi slovarjev in drugih jezikovnih virov med študenti prevajalstva, saj se večina študentov vse pogosteje zateka po pomoč k spletnim in elektronskim virom, medtem ko se virov v tiskani obliki skorajda ne poslužujejo več. Bistveno bolj kot tiskani referenčni viri se uporabljajo predvsem prosto dostopni elektronski viri.

2.2 Opis Kolokacijskega slovarja sodobne slovenščine

V času nastajanja pričujočega prispevka je KSSS 1.0 vseboval 35.989 iztočnic, 7.338.801 kolokacij in 34.935.880 zglede. Je prvi tovrstni slovar v Sloveniji, ki lahko pomembno prispeva k jezikovni produkciji v slovenščini in jezikovni ozaveščenosti uporabnikov pri nas.

KSSS sodi med tako imenovane odzivne slovarje (poleg tega imamo v Sloveniji tudi *Slovar sopomenk sodobne slovenščine* (SSSS),³ ki je bil prvi odzivni slovar pri nas). Odzivni kolokacijski slovarji nastajajo z naprednimi računalniškimi metodami za prepoznavo kolokacij: avtomatska priprava podatkov je tako časovno kot finančno bolj ugodna kot ročna, obenem pa taka sestava omogoča izjemno hitro posodabljanje in nadgrajevanje jezikovnega vira, saj uporabnikom omogoča glasovanje o ustreznosti slovarske informacije (za podrobnejšo definicijo odzivnih slovarjev glej Arhar Holdt idr. (2018)). KSSS je sicer podrobneje predstavljen v Kosem idr. (2018a, 2018b, 2019), Arhar Holdt (2021) in Pori idr. (2020, 2021).

3 <https://viri.cjvt.si/sopomenke/slv/> (Arhar Holdt idr., 2018; Arhar in Čibej, 2020)

2.2.1 O kolokacijah v KSSS

V tem razdelku bomo najprej na kratko orisali pojem kolokacij, kot ta koncept razumejo snovalci KSSS. Gantar idr. (2021) kot pogoj za vključitev v KSSS kolokacije opredeljujejo s pomočjo statističnega, skladenjskega in semantičnega vidika.⁴ Statistični vidik povzemajo po Atkins in Rundell (2008, str. 302), ki kolokacije definirata kot »ponavljajoče se kombinacije besed, v katerih kaže določen leksikalni element (jedro) očitno tendenco sopojavljanja z drugim leksikalnim elementom (kolokatorjem), s frekvenco, ki je večja od naključne sopojavitve«. Drugi vidik je skladenjski, saj med kolokacijskima elementoma obstaja hierarhično razmerje, v katerem baza določa kolokator (prim. Hausmann, 1984, str. 401). Kot tretji, najpomembnejši, a hkrati najbolj problematičen vidik Gantar idr. (2021) izpostavijo pomenski vidik. Ta je tesno povezan s statističnim vidikom, ki kolokacije uvršča med pola, ki ju predstavljajo proste besedne zveze na eni strani in popolnoma ustaljene večbesedne enote na drugi, kar posledično vpliva na semantične spremembe in omejitve v izbiri kolokacijskih elementov.

Dodana vrednost tovrstnega jezikovnega vira za prevajalce je v tem, da na enem mestu na pregleden način prikaže širok nabor kolokacij, med katerimi je nato mogoče izbrati tisto, ki je tako s pomenskega kot tudi sobesedilnega vidika najustreznejša v danem kontekstu.

Pri razumevanju kolokacij in njihovi rabi je ključno upoštevanje besedilnega konteksta, zato so vsi primeri kolokacij v KSSS ponazorjeni z zgledi rabe iz dejanskih besedil, na voljo pa je tudi povezava do korpusa *Gigafida*⁵. KSSS vsebuje stopenjski prikaz gesel, predstavitev kolokacijskih nizov v sobesedilu ter različne možnosti filtriranja in razvrščanja kolokacijskih podatkov.

Slika 1 prikazuje rezultat iskanja kolokacij za samostalnik *informacija*. Na levi strani je ob izpisu gesla podan grafični prikaz stopnje prečiščenosti gesla v obliki rdeče piramide – geslo *informacija* je tako na drugi stopnji od petih, kjer so »odstranjene (...) skladenjske strukture z veliko šuma ter kolokatorji, ki se v večini primerov pojavljajo v neustreznih kolokacijah«. ⁶ Na desni strani je prikazanih nekaj primerov

4 Širši pregled področja kolokacij ponuja Gries (2013).

5 <https://viri.cjvt.si/gigafida/> (Krek idr., 2020; Logar Berginc idr., 2020)

6 Za opise ostalih stopenj glej <https://viri.cjvt.si/kolokacije/slv/about#>

The screenshot shows the KSSS website interface. At the top, there is a search bar with the word 'informacija' entered. Below the search bar, the results are displayed in a grid format. On the left side, there is a sidebar with filters for grammatical categories: 's samostalniki', 'z glagoli', 's pridevniki', 's prislovi', and 'Predlogi'. The main content area shows a grid of collocations for the word 'informacija'.

dobiti informacijo	posredovati	pridobiti	zbirati
pisati v informaciji	gledati v	zapisati v	objaviti v
postreči z informacijo	razpolagati z	seznaniti z	oborožiti z
nositi v informacijo	vpogledati v	pretvoriti v	posredovati v
graditi na informaciji	vprašati na	temeljiti na	sloneti na

Slika 1: Primer iztočnice *informacija* v KSSS.

kolokacij, ki so razvrščene glede na kolokatorje (za podrobnejši opis razvrstitev kolokacij v slovarskem vmesniku glej Arhar Holdt, 2021). Ena izmed možnosti naprednega iskanja so filtri na levi strani, kjer za dotični primer lahko izberemo kategorije »s samostalniki«, »z glagoli«, »s pridevniki«, »s prislovi« in »Predlogi«. Pri slednjih lahko s spustnega seznama določimo predlog, za katerega želimo poiskati primere kolokacij. Podobni dodatni možnosti se ponudita tudi pri samostalnikih in pridevniki, kjer lahko rezultate omejimo glede na sklon.

3 Študija in metodologija

3.1 Namen in raziskovalna vprašanja

Študenti Oddelka za prevajalstvo na Filozofski fakulteti v Ljubljani se v prvem letniku dodiplomskega študija Prevajalstvo pri predmetu *Uvod v prevajanje in informacijsko družbo* seznanijo z različnimi jezikovnimi viri in tehnologijami za slovenščino, med drugim tudi s KSSS. Ker je namen našega prispevka osvetliti odnos študentov do jezikovnih in drugih prevajalskih virov in priročnikov ter njihovo mnenje o uporabniški izkušnji s KSSS, bomo v raziskavi skušali odgovoriti na naslednja vprašanja:

1. S katerimi slovarskimi viri in drugimi jezikovnimi pripomočki so seznanjeni študenti Oddelka za prevajalstvo v Ljubljani in katere uporabljajo pri svojem prevajalskem delu?

2. S kakšnimi viri si študenti med prevajanjem iz angleščine v slovenščino pomagajo pri iskanju prevodnih rešitev za kolokacije?
3. Kako dobro so študenti seznanjeni s KSSS in na kakšne načine ga umeščajo v svoj prevajalski proces pri reševanju specifičnih prevodnih težav med prevajanjem kolokacij?

3.2 Zasnova študije

O tem, kako lahko raziskujemo različne potrebe in mnenja uporabnikov različnih jezikovnih virov, lahko najdemo več v Arhar Holdt idr. (2015, str. 197), ki navajajo, da se uporabniške raziskave lahko poslužujejo:

»(I) anket in intervjujev, ki tipično poizvedujejo, kako pogosto vprašani slovar(je) uporabljajo, za katere namene, kako v splošnem ocenjujejo uporabnost obstoječih priročnikov in kakšnih sprememb si želijo; (II) opazovanja slovarske rabe, kjer se z različnimi pristopi ugotavlja, katere informacije slovarski uporabniki iščejo, kako iskanje poteka, ali so bili podatki uspešno pridobljeni ali ne; (III) eksperimentov in testov, npr. za preverjanje, kako dobro testiranci slovarske podatke razumejo; in (IV) evalvacij posameznih slovarjev, npr. intuitivnosti in preglednosti uporabniškega vmesnika in podobno.«

Vsi ti metodološki principi lahko raziskovalcem razkrijejo poznavanje jezikovnih priročnikov in njihovo rabo, obenem pa osvetljujejo, kako uporabniki »sami ocenjujejo svoje potrebe, navade in želje glede uporabe slovarja, kaj jim je pri določenem slovarju všeč in kaj ne, pa tudi, kaj v določenem slovarju iščejo v procesu uporabe« (Arhar Holdt idr., 2015, str. 197). Arhar Holdt (2015, str. 147) prav tako pravi, da je za raziskovanje uporabniških navad in mnenj glede rabe jezikovnih virov sicer »smotrno uporabiti obstoječo metodologijo področja«, kot so na primer raba anket, intervjujev in testiranj, vendar nujno »z upoštevanjem opozoril stroke« (prim. Tarp, 2009; Müller-Spitzer, 2014).

V naši raziskavi smo tako uporabili kombiniran metodološki pristop, kjer smo povezali predvsem prvi in drugi metodološki pristop, ki ju navajajo Arhar Holdt idr. (2015, str. 197). V prvem delu raziskave smo proučevali podatke, pridobljene s pomočjo uporabniškega vpra-

šalnika o rabi jezikovnih virov med študenti prevajalstva, v drugem pa analizirali podatke, pridobljene s pomočjo opazovanja prevajalskega procesa izbrane ciljne skupine študentov prevajalstva pri iskanju prevodnih rešitev izbranih kolokacij, še posebej s pomočjo KSSS, ter preverjali njihovo uporabniško izkušnjo s pomočjo polstrukturiranega intervjuja po prevajalski nalogi. Šorli in Ledinek (2017, str. 388), ki sta se v svoji raziskavi prav tako odločili za uporabo spletnega vprašalnika, se pri tem sklicujeta na Müller-Spitzer idr. (2012, str. 5), ki pravijo, da na tak način lahko pridobimo bolj informirane podatke (prim. tudi Šorli, 2018). Podobno so s kombiniranjem različnih metodoloških pristopov svoje raziskave izvajali tudi Pori idr. (2020) ter Pori idr. (2021) pri evalvaciji uporabniškega vmesnika KSSS, prav tako pa tudi Arhar Holdt idr. (2018) ter Arhar Holdt in Čibej (2020) pri uporabniških raziskavah *Slovarja sopomenk sodobne slovenščine*.

Zgoraj omenjene študije so prispevale k odločitvi, da s prepletom različnih metodoloških pristopov lahko najboljše osvetlimo uporabniško izkušnjo in odnos do rabe različnih jezikovnih virov in priročnikov med bodočimi prevajalci, saj s pomočjo triangulacije podatkov (prim. npr. Alves, 2003; Hirci, 2007) lahko prispevamo k celovitejši osvetlitvi obravnavane tematike.

3.2.1 Prvi del raziskave: Vprašalnik in kvantitativna analiza podatkov

V prvem delu bomo predstavili rezultate kvantitativnih podatkov, ki smo jih pridobili s pomočjo vprašalnika o rabi jezikovnih virov med študenti prevajalstva. Arhar Holdt (2015) v svojem članku sicer izpostavlja opozorila Tarpa (2009, str. 290), da je ob vključevanju študentov v tovrstne raziskave treba upoštevati, da njihove uporabniške navade, želje in preference ne predstavljajo uporabniških izkušenj celotne populacije, vendar pa je pri raziskovanju »rabe slovarjev ali pridobivanja jezikovne kompetence v času študija« izbira te uporabniške skupine »smiselna in smotrna« (Arhar Holdt, 2015, str. 144). Ta uporabniška skupina je bila za našo raziskavo izbrana namenoma, saj vsako izboljšavo izobraževalnega procesa razumemo kot dodano vrednost našemu pedagoškemu udejstvovanju. Na Oddelku za prevajalstvo v Ljubljani namreč stremimo

k temu, da bi diplomante čim bolj opremili z ustreznim poznavanjem in rabo relevantnih jezikovnih virov ter kritičnim ovrednotenjem najdenih informacij, da so čim bolj pripravljeni za nadaljnje profesionalno prevajalsko delo.

Po pregledu različnih slovarjev in drugih jezikovnih virov za jezikovni par slovenščina-angleščina so študenti Oddelka za prevajalstvo Univerze v Ljubljani odgovarjali na specifična vprašanja, vezana na uporabniško izkušnjo s KSSS: kako dobro so seznanjeni s tem jezikovnim virom in kako pogosto ga uporabljajo pri svojem prevajalskem delu, kako uporaben se jim zdi pri iskanju specifičnih kolokacij in podobno.

Vprašalnik, ki sestoji iz 22 vprašanj tako odprtega kot zaprtega tipa, smo sestavili s pomočjo spletnega orodja za anketiranje 1KA. Vprašalnik vsebuje dva dela – vsebinski del vključuje 19 vprašanj, ki se nanašajo na pridobivanje podatkov o jezikovnih virih in uporabniških izkušnjah, zadnja tri vprašanja pa zajemajo demografske podatke o spolu, letniku študija in jezikovni smeri respondentov.

V vsebinskem delu so najprej zastavljena vprašanja o uporabniških izkušnjah z jezikovnimi viri za jezikovni par slovenščina-angleščina, v drugem delu pa so navedena bolj specifična vprašanja, ki se osredotočajo na delo s kolokacijami, s strategijami iskanja ustreznih kolokacij in uporabniško izkušnjo s KSSS. Čeprav je večina vprašanj zaprtega tipa, kjer so možni odgovori da/ne ali izbiranje med več različnimi vnaprej podanimi možnostmi, je pri vsakem vprašanju respondent povabljen, da poda tudi druge informacije ali dodatno opiše svojo uporabniško izkušnjo. Prav te vrste informacij so še posebej dragocene, saj podrobneje osvetljujejo raziskovano tematiko.

K izpolnjevanju spletnega vprašalnika 1KA o uporabniški izkušnji z različnimi jezikovnimi viri, s poudarkom na prevajanju kolokacij in uporabniški izkušnji s KSSS, so bili povabljeni vsi redno vpisani študenti Oddelka za prevajalstvo Univerze v Ljubljani. Izpolnjevanje vprašalnika je bilo na voljo osem tednov, in sicer od 15. januarja do 15. marca 2021. Po nekaj tednih aktivne ankete smo vse morebitne respondente ponovno povabili k sodelovanju pri izpolnjevanju vprašalnika.

3.2.2 Drugi del raziskave: Snemanje zaslona med prevajalsko nalogo in zaključni intervju

Kvalitativne podatke smo pridobili z opazovanjem prevajalskega procesa pri prevajanju izbranih kolokacij. Neposredno opazovanje uporabnikove aktivnosti na zaslonu namreč nudi podatke o tem, kdaj in kako prevajalci tvorijo prevod, tj. v kakšnih časovnih razsežnostih se ga lotevajo, kako pogosto revidirajo določene segmente, kako pogosto iščejo po virih informacij, kako pogosto preverjajo pravilno razumevanje besedila in ustreznost prevodnih rešitev v slovarjih, enciklopedijah in drugih virih informacij (prim. npr. Hansen, 1999, 2003; Hirci, 2007, 2009b; Jakobsen, 2003; Jensen in Jakobsen, 2000; Jääskeläinen, 1999). Podatke o prevajalskem procesu lahko pridobimo s snemanjem oz. beleženjem celotnega prevajalskega procesa (npr. Translog), s snemanjem oz. z zajemom računalniškega zaslona (npr. Camtasia ali CamStudio, OBS Studio ipd.), ali z različnimi tehnologijami za sledenje premikanja oči po zaslonu (angl. *eye-tracking devices*). Programi za zajem računalniškega zaslona posnamejo vse aktivnosti: od premikanja z miško, do odpiranja oken na spletu, pregledovanja različnih dokumentov, iskanja po različnih virih informacij, jezikovnih priročnikih in podobno. Natančno dokumentirajo, kaj se dogaja na zaslonu in v obliki avdio-video zapisa beležijo vse aktivnosti uporabnika. Shranjene posnetke lahko raziskovalec ponovno pregleduje in analizira (podrobneje v Hirci, 2007 in Hirci, 2009b). Ta metoda se je v preteklosti pogosto kombinirala z metodo PGR (tj. protokol glasnega razmišljanja, angl. TAP oz. *think-aloud protocol*). Ker pa je Bernardini (2001, str. 242–263) že pred časom opozarjala, da mora biti izpolnjenih kar nekaj pogojev za objektivizacijo dela z metodo PGR, sicer težave, ki se lahko pojavljajo pri delu s to metodo, postavljajo pod vprašaj veljavnost njene aplikacije na prevajalsko delo, se v svoji raziskavi nismo odločili za uporabo te metode. Smo se pa za pridobitev dodatnih informacij, ki prav tako osvetljujejo prevajalski proces, odločili uporabiti polstrukturirani intervju.

V drugem delu študije, kjer smo želeli pridobiti kvalitativne podatke, smo študente najprej povabili k sodelovanju. S tistimi, ki so se odločili sodelovati, smo se dogovorili za termin, ko bo študija izvedena, in jim posredovali dodatna navodila o celotnem poteku študije. Pred izvedbo uporabniške raziskave smo sodelujoče seznanili tudi s tem, da

bodo podatki, pridobljeni med snemanjem zaslona, uporabljeni izključno v raziskovalne namene, in jih prosili za pisno privolitev za sodelovanje (prim. Pori idr., 2020).

Študenti so dobili v prevod krajše angleško besedilo (glej Prilogo I), med prevajalskim delom pa smo jih snemali s programom za zajem zaslona (v našem primeru smo uporabili Zoom, ki so ga študenti zaradi študija na daljavo v zadnjih dveh letih dodobra spoznali in so vajeni dela s tem programom). Ker je snemanje prevajalskega procesa potekalo na platformi Zoom, so študenti lahko uporabili možnost deljenja celotnega zaslona. Tako smo pri analizi podatkov lahko upoštevali celoten potek prevajanja in načine iskanja po različnih prevajalskih virih. V navodilih so bili sodelujoči pozvani, da naj prevedejo zgolj tiste povedi (in en podnaslov), v katerih se nahajajo označene kolokacije. Poudarili smo, da se pri prevajanju lahko poslužujejo katerih koli jezikovnih virov in orodij, ki jih tudi sicer uporabljajo pri svojem prevajalskem delu. Študenti za dokončanje prevoda niso imeli časovnih omejitev.

Po zaključenem prevodu izbranih delov besedila smo s študenti izvedli še kratek intervju o dejanski uporabniški izkušnji med procesom prevajanja. Ta korak do neke mere omogoča preverbo predpostavk oz. dopolnitev vrzeli med samim snemanjem. Intervjuvanci so med polstrukturiranim intervjujem (glej Prilogo II) odgovarjali na vprašanja o uporabi in koristnosti KSSS, kaj bi dodali ali spremenili v tem jezikovnem viru, o prečiščenosti vira, o preklapljanju med dodatnimi viri in podobno.

3.3 Subjekti raziskave

V prvem delu vprašalnika, namenjenega raziskavi rabe jezikovnih virov, so sodelovali študenti prevajalstva na Oddelku za prevajalstvo Filozofske fakultete Univerze v Ljubljani, ki že poznajo delo s KSSS: s tem virom se namreč seznanijo med pedagoškim procesom v prvem letniku dodiplomskega študija (BA); sodelovali so torej študenti od 2. letnika BA naprej.

Petnajst študentov prostovoljcev, osem iz 2. (označeni z II 1–8) in sedem iz 3. letnika (III 1–7) dodiplomskega študija, je kasneje sodelovalo tudi v drugem delu raziskave, tj. pri opazovanju prevajalskega procesa in iskanju prevodnih rešitev kolokacij.

4 Rezultati in diskusija

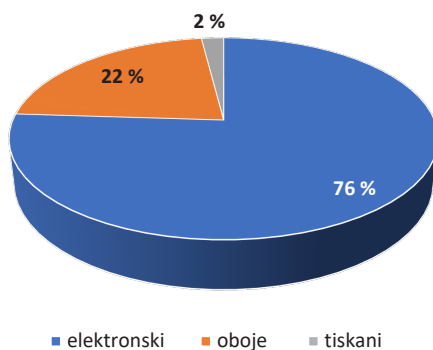
4.1 Rezultati vprašalnika o prevajalskih virih in rabi KSSS

Na spletno anketo je odgovorilo 100 anketirancev. Slaba polovica respondentov je bila s stopnje BA (42 %): osem iz prvega letnika, 21 iz drugega letnika in 13 iz tretjega letnika dodiplomske stopnje. Relativno dober odziv je bil tudi med študenti na magistrski (MA) stopnji, še posebej med študenti prvega letnika MA (18 %), medtem ko je sodelovalo zgolj štiri odstotke študentov MA II. Med respondenti je bilo 84 % študentk in 16 % študentov, kar se relativno dobro sklada tudi s siceršnjo zastopanostjo po spolu na Oddelku za prevajalstvo.

Anketo je v celoti izpolnilo 64 respondentov, medtem ko je bila v 36 primerih v določenih delih vprašalnika izpolnjena le delno. Več kot polovico odgovorov (n = 64) so prispevali študenti z jezikovno kombinacijo slovenščina-angleščina-nemščina, kar je odslilkava dejanske sestave študentov prevajalstva, saj je ta jezikovna kombinacija najpogosteje zastopana na Oddelku za prevajalstvo v Ljubljani.

4.1.1 Tip jezikovnih virov

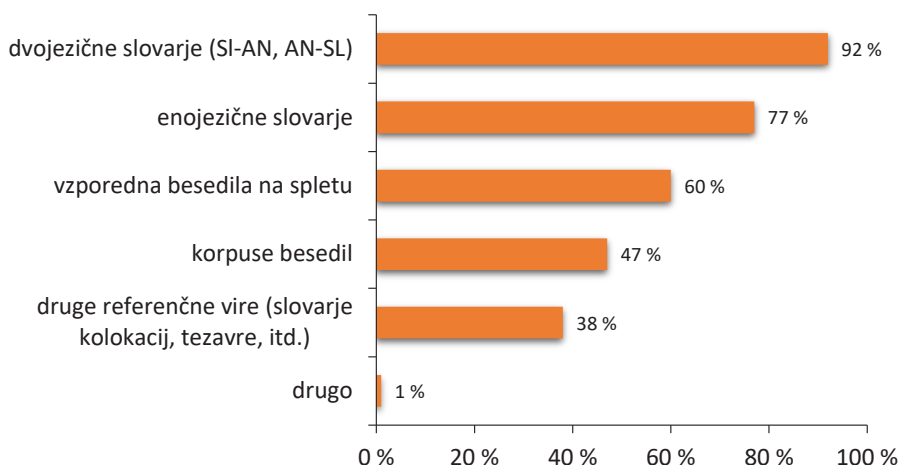
Na vprašanje o tipu jezikovnih virov (glej Grafikon 1) je večina respondentov odgovorila, da najpogosteje uporabljajo jezikovne vire v elektronski obliki (76 %) ali pa kombinacijo tiskanih in elektronskih virov (22 %).



Grafikon 1: V kakšni obliki ponavadi uporabljate jezikovne vire – v tiskani ali elektronski? (n = 100)

4.1.2 Raba jezikovnih virov pri prevajanju iz angleškega v slovenski jezik

Pri prevajanju iz angleščine v slovenščino študenti prevajalstva najpogosteje iščejo prevodne rešitve v eno- (77 %) in dvojezičnih slovarjih (92 %), pogosto tudi s pomočjo vzporednih besedil na spletu (60 %). Skoraj polovica se jih poslužuje korpusev besedil (47 %), pogosta je tudi raba slovarjev kolokacij, medtem ko po ostalih virih informacij respondenti posegajo manj pogosto (glej Grafikon 2).

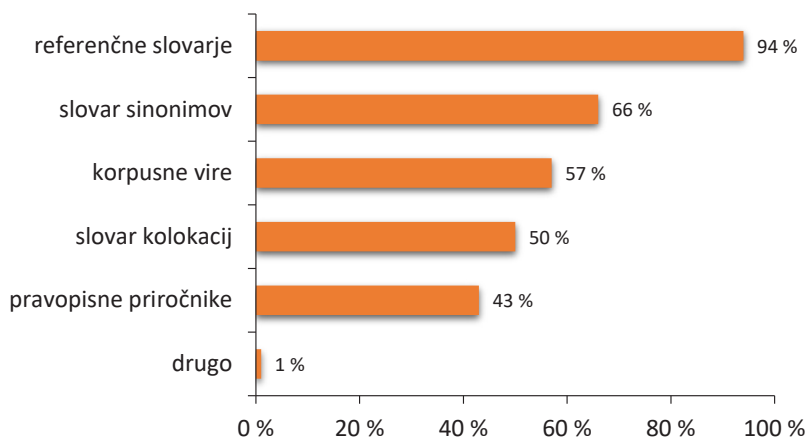


Grafikon 2: Katere jezikovne vire najpogosteje uporabljate pri prevajanju iz angleškega v slovenski jezik? (n = 90)

4.1.3 Referenčni viri za slovenščino

Respondenti so dobro seznanjeni s slovenskimi jezikovnimi in drugimi referenčnimi viri. Iz Grafikona 3 je tako razvidno, da jih največ uporablja enojezične referenčne slovarje (94 %), manj jih posega po pravopisnih priročnikih (43 %). Več kot polovica respondentov se poslužuje tudi korpusev virov (57 %) in slovarja sinonimov (66 %), nekaj manj kot polovica pa med svojim prevajalskim delom posega po KSSS (50 %). Eden od respondentov je v dodatnem komentarju omenil, da se včasih za nasvet obrne na jezikovno svetovalnico (sicer v dodatnem komentarju ni specifično označeno, a verjetno gre za svetovalnico ISJFR⁷).

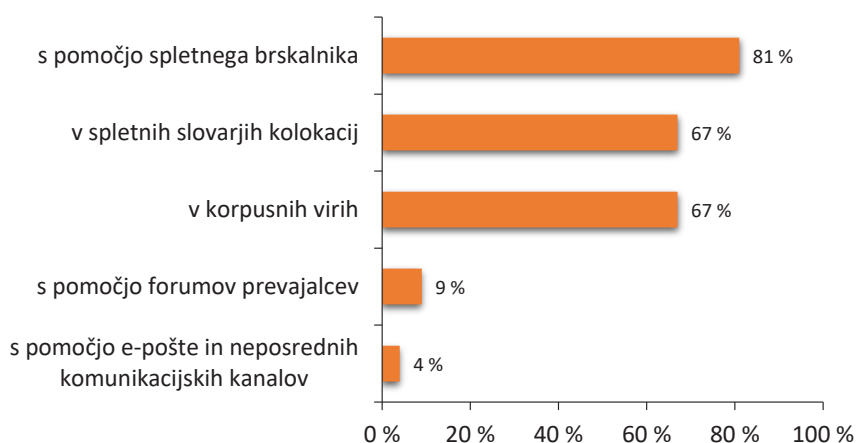
7 ISJFR – Inštitut za slovenski jezik Frana Ramovša



Grafikon 3: Katere slovenske referenčne vire najpogosteje uporabljate? (n = 90)

4.1.4 Iskanje ustreznih kolokacij

Ko smo anketirance spraševali o različnih virih, ki se jih poslužujejo pri iskanju ustreznih kolokacij (Grafikon 4), so odgovarjali, da v večini primerov najprej skušajo poiskati ustrezne rešitve kar s spletnim iskalnikom (81 %). Če na ta način ne uspejo dobiti ustreznih rešitev, iščejo po korpusnih virih (67 %), ali pa po spletnih slovarjih kolokacij (67 %). Zgolj redki se obrnejo po pomoč tudi na forume prevajalcev ali iščejo pomoč prek neposrednih komunikacijskih kanalov.



Grafikon 4: Ko iščete prevodne ustreznice za določene kolokacije, kje najpogosteje najdete ustrezne rešitve? (n = 69)

4.1.5 Strategije iskanja ustreznih prevodnih rešitev za kolokacije

Pri vprašanju o strategijah pri iskanju ustreznih kolokacij so bili odgovori respondentov v odprtem delu vprašalnika zelo raznoliki, med drugim so navajali, da (navedeno verbatim):

- *Ponavadi naprej pogledam v slovar kolokacij, če tam ne najdem ustrezne rešitve, navadno potem iščem rešitev s pomočjo spletnega brskalnika, nato pa še v korpusnih virih.*
- *Napišem jedro kolokacije in prebiram zadetke.*
- *<https://www.freecollocation.com/> in korpusni pristop z določitvijo okolice.*
- *Najprej določim pomen, če ne najdem ustreznice v slovenščini, pogledam v korpus in slovar kolokacij.*
- *S pomočjo iskanja v okolici poiščem kolokacijo v korpusu, nato kolokacije, ki jih tam najdem, še poišem na spletnem brskalniku.*
- *Kolokacijo, ki se mi zdi ustrezna, vnesem v brskalnik, da ugotovim, ali se pogosto pojavlja. Če to iskanje ne obrodi rezultatov, kolokacije iščem še v korpusih.*
- *Kolokacijo poskušam najti najprej na Glosbe⁸. Če dobim primerne zadetke, jih preverim še na spletu (predvsem, kako pogosto se uporabljajo). Če na Glosbe ni zadetkov, pa iščem s pomočjo Googla in raznih enojezičnih slovarjev.*
- *Zelo odvisno od primera. Navadno najprej preverim v slovarju kolokacij in če ne najdem rešitve, jo skušam najti s pomočjo spletnega brskalnika. Vsekakor to iskanje poteka malo zmedeno in vzame kar nekaj časa ...*
- *Za slovenske kolokacije uporabljam iskanje v okolici na Gigafida⁹, za druge pa predvsem besedne zveze, navedene v slovarju. Če ne najdem iskanega, pa vpišem v spletni brskalnik in iščem zadetek, ki se mi zdi, kot da je verodostojen vir.*
- *Kolokacijo vpišem v strojni prevajalnik ali pa slovar (če je krajša in če predvidevam, da jo slovar sploh vsebuje). Ko dobim zadetek, ga še enkrat vpišem v spletni iskalnik in si ogledam najbolj priljubljene zadetke, da preverim, ali je prevod v strojnem prevajalniku/slovarju ustrezen.*
- *Če imam določeno kolokacijo že v mislih in želim le preveriti njeno rabo, to navadno storim kar v brskalniku, če rezultat ni ravno prepri-*

8 <https://sl.glosbe.com/>

9 <https://viri.cjvt.si/gigafida/>

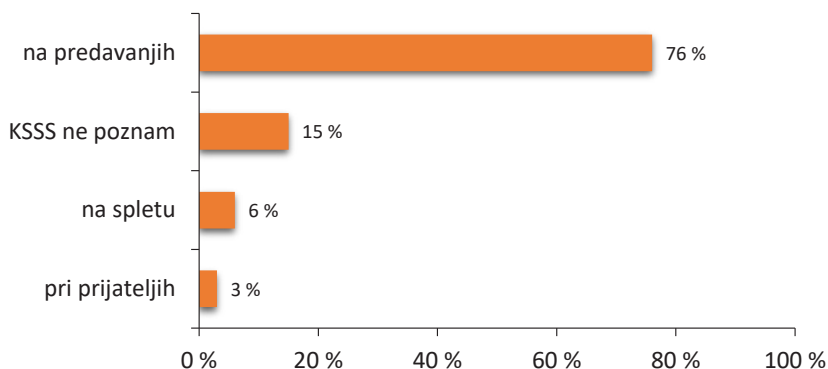
čljiv, preverim še v korpusih ipd. Kadar sem zelo v dvomih, za nasvet vprašam koga od kolegov.

- V brskalniku iščem besedo, ki jo potrebujem, in gledam vire, ki imajo več jezičnih možnosti. Ko najdem prevod, preverim še v enojezičnem slovarju ali slovarju kolokacij.
- Kolokacije vtipkam v splet in poskušam izluščiti njihov pomen iz zadetkov ali (če se mi vsaj malo sanja, kaj iščem) vtipkam vse ideje v korpusni vir, ki mi nato pokaže, kateri bi bil najbolj ustrezen. Po vsem tem presodim, ali sem do odgovora že prišla ali naj iščem naprej.
- Ang-slo: z enojezičnimi angleškimi slovarji raziščem pomen, nato navadno kolokacijo najprej poiščem v Glosbeju in nato preverim z iskanjem na spletu ali v Gigafidi. Včasih iščem kolokacije tudi v slovarju kolokacij, ampak ni prav pregleden. Slo-ang: ne poznam dobro angleških korpusov in kolokacijskih virov, zato navadno vse iščem na spletu.
- Najprej preverim v kolokacijskem slovarju, včasih vpišem v Google in preverim, če so zadetki iz kredibilnih strani.
- Najprej v spletnem slovarju kolokacij napišem stalno besedno zvezo in nato dobljeni prevod vpišem v spletni brskalnik, da vidim, če je prevod zanesljiv.
- Kadar iščem po korpusnih virih, vpišem besedo, za katero ne vem, s čimer kolocira, in odključam besedno(-e) vrsto(-e), ki jo/jih iščem.
- Moja prva izbira je skoraj vedno slovar. V primeru, da se mi najden prevod ne zdi ustrezen za ta določen kontekst, iščem naprej po korpusih, forumih, člankih itd.

4.1.6 Poznavanje KSSS

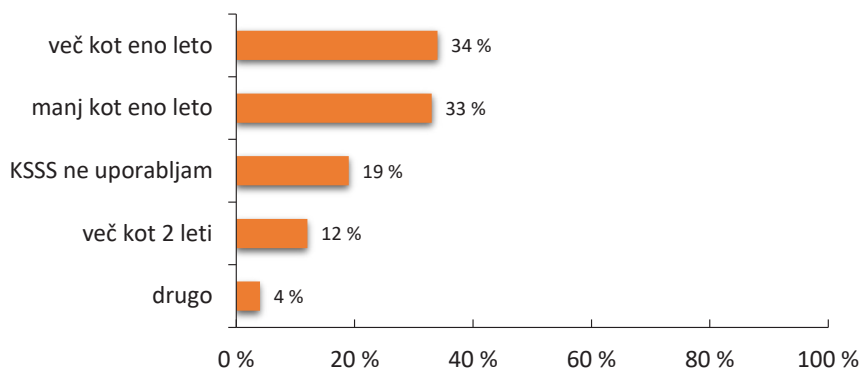
Večina respondentov je na vprašanje o tem, kje so spoznali KSSS (Grafikon 5), odgovorila, da na fakulteti (76 %), najprej pri predmetu *Uvod v prevajanje in informacijsko družbo*, kasneje pa tudi pri kakšnem izmed prevajalskih modulov, nekaj se jih je z njim seznanilo prek spleta ali pa so za ta jezikovni vir slišali od prijateljev.

Kljub temu, da se študenti Oddelka za prevajalstvo v Ljubljani s tem jezikovnim virom seznanijo med pedagoškim procesom že v prvem letniku dodiplomskega študijskega programa, je 15 % respondentov odgovorilo, da KSSS ne poznajo.



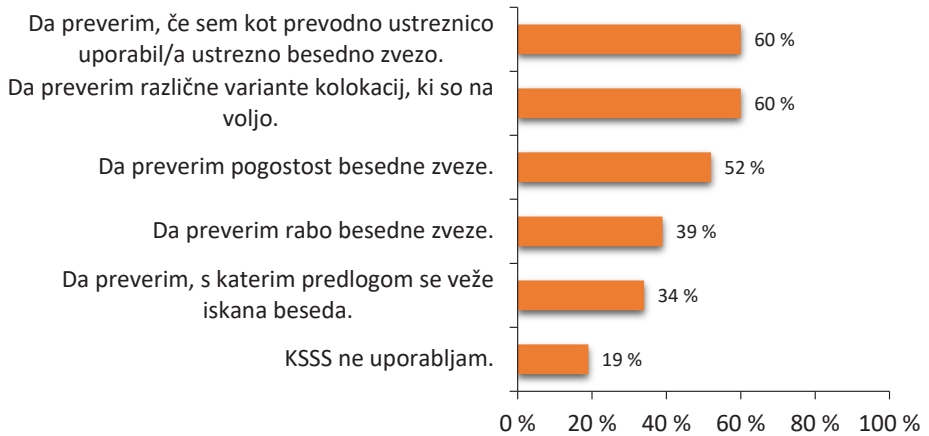
Grafikon 5: Kje ste se seznanili s KSSS? (n = 67)

Na vprašanje o tem, kako dolgo že uporabljajo ta jezikovni vir (prim. Grafikon 6), je 34 % respondentov odgovorilo, da ga uporabljajo že več kot eno leto, 33 % ga uporablja manj kot eno leto, 12 % vprašanih pa že celo več kot dve leti. Je pa kar 19 % respondentov odgovorilo, da KSSS pri svojem prevajalskem delu ne uporabljajo.



Grafikon 6: Kako dolgo že uporabljate KSSS? (n = 67)

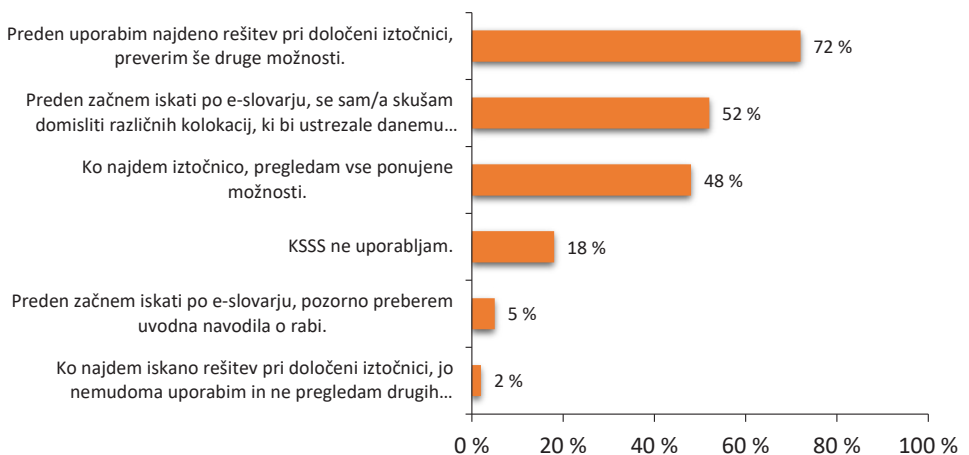
Respondenti so navajali (glej Grafikon 7), da KSSS uporabljajo predvsem, da preverijo različne variante kolokacij, ki so na voljo (60 %), da preverijo ustreznost besedne zveze kot prevodno rešitev (60 %), da preverijo samo pogostost besedne zveze (52 %), nekateri pa tudi, da preverijo, s katerim predlogom se veže iskana beseda (34 %).



Grafikon 7: Zakaj uporabljate KSSS? (n = 67)

4.1.7 Strategije iskanja po KSSS

Kot je razvidno iz Grafikona 8 je največ respondentov navedlo, da preverijo tudi druge možnosti, ki so ponujene v KSSS, preden uporabijo najdeno rešitev pri določeni iztočnici (72 %); 52 % respondentov je odgovorilo, da se skušajo sami domisliti različnih kolokacij, ki bi ustrezale sobesedilu, preden sploh začnejo iskati po elektronskem slovarju; 48 % respondentov dodaja, da pregledajo tudi vse ostale ponujene možnosti, ko najdejo določeno iztočnico.



Grafikon 8: Kakšne so vaše strategije iskanja po KSSS? (n = 65)

Več o tem, ali in na kakšen način študenti prevajalstva dejansko uporabljajo KSSS je predstavljeno v naslednjem razdelku, kjer je prikazan prevajalski proces na izbranem številu študentk in študentov, ki so se prostovoljno javili k sodelovanju v tem delu raziskave.

4.2 Rezultati opazovanja prevajalskega procesa in rabe KSSS

V drugem delu raziskave je sodelovalo 15 študentov, in sicer osem študentov 2. letnika in sedem študentov 3. letnika dodiplomskega študija. Od tega je v drugem letniku sodelovalo šest študentk in dva študenta, ki študirajo na jezikovnih smereh slovenščina-angleščina-nemščina (7) in slovenščina-angleščina-francoščina (1), v tretjem letniku pa sedem študentk, ki so vpisane na smereh slovenščina-angleščina-nemščina (3) in slovenščina-angleščina-francoščina (4). Kot je bilo že omenjeno, vsi študenti Oddelka za prevajalstvo v prvem letniku spoznajo KSSS in so seznanjeni tako z njegovo vlogo kot tudi z različnimi načini uporabe tega slovarskega vira.

Sledi analiza podatkov, ki smo jih pridobili s snemanjem dogajanja na zaslonu med prevajalskim procesom, analiza prevodnih rešitev ene od kolokacij ter analiza rezultatov, ki smo jih pridobili s pomočjo polstrukturiranega intervjuja po prevajalski nalogi.

4.2.1 Prevodne rešitve za kolokacijo št. 4

Z vidika prevajanja se je poved, v kateri se nahaja kolokacija *release information*, izkazala za najbolj zahtevno, saj je že izvirnik nekoliko problematičen. Samostalniška besedna zveza *the way* v funkciji premege predmeta, ki sledi glagolu *criticise*, se namreč najpogosteje uporablja v smislu »kritizirati način, na katerega je bilo nekaj storjeno«, v analiziranem primeru pa ima drugačen pomen, saj poved govori o tem, da so znanstveniki kritizirali dejstvo, da je bila ta informacija sploh posredovana medijem, ne pa način, kako je bilo to storjeno.

V Tabeli 1 so prikazane prevodne rešitve, ki so bile označene kot neustrezne (II-1, II-5 in II-7) oz. delno ustrezne (III-2, III-3 in III-5), in jezikovni viri, ki so jih študenti uporabili med prevajanjem. V prvo skupino spadajo tiste, ki izkazujejo skladijsko neustreznost v ciljnem jeziku ali pa gre za pomensko neustrezno prevodno rešitev, četudi je bila upo-

rabljena slovenska kolokacija z visoko pogostostjo. Kot zgolj delno oz. pogojno ustrezne smo opredelili prevode kolokacij, ki so se pomensko preveč oddaljili od izvirnika ali pa jih zaznamuje skladijsko netipična oblika. V to skupino spada tudi delna prevodna rešitev (III-7), kjer ni prevedena celotna poved, temveč zgolj kolokacija, ki je sicer ustrezna.

Študentskim prevodnim rešitvam smo za primerjavo dodali še strojne prevode, ki jih sicer pri prevajanju te povedi ni uporabil nihče, čeprav pri prevajanju niso imeli omejitve virov. Prevoda Google Translate in DeepL sta v celoti identična, uporabljata ustrezno kolokacijo, ki pa je glede na izvirnik že nekoliko interpretativna – pri tej prevodni rešitvi je kritika znanstvenikov še bolj uperjena proti medijem in ne pre-nagljeni odločitvi njihovih stanovskih kolegov. S tega vidika je prevodna rešitev, ki jo ponudi eTranslation, nekoliko bližje izvirniku. V vseh treh strojnih prevodih pa je *the way* preveden zelo dobesedno s samostalnikom *način*.

Tabela 1: Prikaz neustreznih oz. delno ustreznih prevodnih rešitev in jezikovnih virov, ki so jih študenti uporabili pri iskanju prevodnih rešitev za kolokacijo št. 4

	Kolokacija št. 4	Viri	Uspešnost strategije iskanja po KSSS
IZVIRNIK	Some scientists have criticised the way the information was released to the media before it had been thoroughly confirmed.		
II-1	Nekateri znanstveniki so skritizirali način, kako je bila informacija, prej ko je sploh bila temeljito potrjena, deljena z mediji.	brez virov	/
II-5	Nekateri znanstveniki so skritizirali način izdaje podatkov medijem, še pred tem, ko so bili potrjeni.	angleški kolokacijski slovar ozdic.com	/
II-7	Nekateri znanstveniki so kritizirali način, da so novico objavili v javnosti, še preden so jo temeljito preverili.	spletni an-sl slovar Pons, KSSS	?
III-2	Nekateri znanstveniki kritizirajo način, ki je bil uporabljen za posredovanje informacij javnosti, preden so bile te temeljito preverjene.	veliki an-sl slovar (Amebisov pregledovalnik podatkovnih zbirk ASP32), KSSS	?

	Kolokacija št. 4	Viri	Uspešnost strategije iskanja po KSSS
IZVIRNIK	Some scientists have criticised the way the information was released to the media before it had been thoroughly confirmed.		
III-3	Nekateri strokovnjaki so kritizirali način, na katerega so bile informacije sporočene medijem, preden so bile potrjene.	Evrokorpus, KSSS	Ne
III-5	Nekateri znanstveniki so kritizirali dejstvo, da so mediji objavili informacijo preden bi jo lahko popolnoma potrdili.	korpus Gigafida	/
III-7	*kako je bila informacija posredovana	KSSS	Da
Google Translate	Nekateri znanstveniki so tako kritizirali način, na katerega so bile informacije objavljene v medijih, še preden so bile temeljito potrjene.	/	/
DeepL	Nekateri znanstveniki so kritizirali način, na katerega so bile informacije objavljene v medijih, še preden so bile temeljito potrjene.	/	/
e-translation	Nekateri znanstveniki so kritizirali način, kako so bile informacije posredovane medijem, preden so bile temeljito potrjene.	/	/
<p><i>Opombe.</i> * = prevod zajema zgolj kolokacije in ne celih povedi ; / = KSSS ni bil uporabljen; ? = Najdena kolokacija je primerna, a je prevodna rešitev celotne povedi neustrezna.</p>			

Iskalni nizi študentov, katerih prevodne rešitve so bile ocenjene kot neustrezne oz. delno ustrezne:

II-1: prevedeno brez uporabe jezikovnih virov

II-5: uporabljen angleški kolokacijski slovar ozdic.com z iskalnima nizoma *information* in *released information*; KSSS: ni bil uporabljen

II-7: angleško-slovenski slovar Pons za iskanje ustreznih besed *release*, *some* in *thoroughly*; KSSS: samostalnik *novica*, filter »z glagoli«

III-2: uporabljen veliki angleško-slovenski slovar (Amebisov pregledovalnik podatkovnih zbirk ASP32) za glagol *criticise*; KSSS: samostalnik *informacija*, brez filtrov

III-3: Evrokorpus z iskalnima nizoma *released information* in *information released*; KSSS: *izdaja podatkov* (brez zadetkov)

III-5: uporabljen slovensko-angleški spletni slovar Pons za glagol *criticise*, korpus Gigafida z iskalnima nizoma *izdati informacijo* in *objaviti informacijo* ter SSSS za glagole *kritizirati*, *grajati* in *izdati*; KSSS: ni bil uporabljen

III-7: uporabljen samo KSSS: samostalnik *informacija*, filter »z glagoli«

Pri prevajanju kolokacije št. 4 zgolj en študent ni uporabil nobenih jezikovnih virov, pri štirih med uporabljenimi jezikovnimi viri ni bilo KSSS, deset pa jih je KSSS uporabilo. Med šestimi prevodnimi rešitvami, ki smo jih označili kot neustrezne oz. zgolj delno ustrezne (II-1, II-5, II-7, III-2, III-3 in III-5), pri polovici študenti niso uporabili KSSS.

V Tabeli 2 so prikazane prevodne rešitve, ki so bile označene kot ustrezne, in jezikovni viri, ki so jih študenti pri tem uporabili.

Tabela 2: Prikaz ustreznih prevodnih rešitev in jezikovnih virov, ki so jih študenti uporabili pri iskanju prevodnih rešitev za kolokacijo št. 4

	Kolokacija št. 4	Viri	Uspešnost strategije iskanja po KSSS
IZVIRNIK	Some scientists have criticised the way the information was released to the media before it had been thoroughly confirmed.		
II-2	Nekateri znanstveniki so kritizirali, kako je informacija lahko prišla v medije preden je bila temeljito preverjena.	KSSS	Da
II-3	Nekateri znanstveniki so kritizirali dejstvo, da so bile informacije posredovane medijem, še preden so bile temeljito potrjene.	KSSS, spletni an-sl in sl-an slovar (Pons)	Da
II-4	Nekateri znanstveniki so kritizirali način, kako so bile medijem posredovane informacije, predno so bile natančno potrjene.	KSSS	Da
II-6	Nekateri znanstveniki so kritizirali način, kako so bile informacije posredovane medijem, preden so bile dokončno potrjene.	KSSS, spletni an-sl slovar (Pons)	Da
II-8	Nekateri znanstveniki so močno kritizirali dejstvo, da so informacije prišle v medije, preden so bile dokončno potrjene.	Fran, Gigafida	/

	Kolokacija št. 4	Viri	Uspešnost strategije iskanja po KSSS
IZVIRNIK	Some scientists have criticised the way the information was released to the media before it had been thoroughly confirmed.		
III-1	Nekateri znanstveniki kritizirajo dejstvo, da je bila informacija posredovana medijem, še preden so jo v celoti potrdili.	KSSS, Gigafida, veliki an-sl slovar (Amebisov pregledovalnik podatkovnih zbirk ASP32)	Da
III-4	Nekateri znanstveniki so kritizirali način, kako so bile informacije posredovane medijem, še preden so bile potrjene.	sl-an vzporedni korpus Linguae, Vejice 1.0	/
III-6	Nekateri znanstveniki so kritizirali, da je bila informacija posredovana medijem, preden je bila dodobra potrjena.	KSSS	Ne

Opomba. / = KSSS ni bil uporabljen

Med osmimi študenti, katerih prevodne rešitve so ustrezne, jih je KSSS uporabilo šest, od tega pet uspešno in ena neuspešno.

II-2: KSSS: glagol *preveriti*, brez uporabe filtrov; iskanje uspešno, raba preverjena tudi z zgledi

II-3: KSSS: samostalnik *informacija*, filter »s pridevniki«, brez uspeha; glagol *posredovati*, brez uporabe filtra; iskanje uspešno – *posredovati informacijo*

II-4: KSSS: samostalnik *informacija*, filter »z glagoli«; iskanje uspešno – *posredovati informacijo*, raba preverjena z zgledi

II-6: KSSS: samostalnik *informacija*, brez uporabe filtrov; kliku na kolokacijo *dobiti informacijo*, aktiviranje filtra »z glagoli«; iskanje uspešno – *posredovati informacijo*

II-8: KSSS: ni bil uporabljen

III-1: KSSS: samostalnik *informacija*, brez uporabe filtrov, klik na vrstico kolokacij z glagoli; iskanje uspešno – *posredovati informacijo*, pogostost kolokacije preverjena v korpusu Gigafida (*informacija* + filter »z glagoli«)

III-4: KSSS: ni bil uporabljen

III-6: KSSS: samostalnik *informacija*, brez uporabe filtrov; iz dogajanja na zaslonu ni mogoče razbrati, ali je nefiltriran seznam kolokacij kakor koli vplival ali pripomogel k prevodni rešitvi

4.2.2 Analiza mnenj uporabnikov

V tem delu so predstavljeni rezultati raziskave, ki smo jih pridobili s pomočjo vnaprej pripravljene vprašalnika in polstrukturiranega intervjuja, kjer smo s pomočjo vodenih vprašanj subjektom raziskave po opravljenem snemanju prevajalskega procesa skušali dopolniti informacije o samem poteku prevajanja in rabi KSSS pri iskanju prevodnih rešitev za označene kolokacije.

Respondente¹⁰ smo vprašali, v kolikšni meri so seznanjeni s KSSS in kako dolgo ga že uporabljajo: vsi sodelujoči so odgovorili, da so seznanjeni s tem jezikovnim virom, ki se jim zdi koristen za prevajalsko delo. Sledi navedba posameznih vprašanj in odgovori posameznih respondentov.

4.2.2.1 Zakaj se vam KSSS zdi koristen?

(odgovori respondentov so podani verbatim):

II:-1 *Ponudi izbor možnih kolokacij, izmed katerih nato lažje izberem tisto pravo.*

II-6: *Dober vir, da dobiš kakšno idejo; povezava do Gigafide (daje možnost odločanja glede pogostosti).*

III-1: *Pregleden, hitrost iskanja kolokacij.*

III-2: *Preglednost iskanih nizov.*

III-3: *Ker se pri iztočnici vidi kontekst, ki ga lahko primerjaš z dano situacijo. Vidi se tudi pogostost uporabe. Daje možnost iskanja po osnovnih oblikah posameznih besed.*

4.2.2.2 Katere informacije, ki jih KSSS v zdajšnji obliki še ne vsebuje, bi lahko prispevale k izboljšani uporabniški izkušnji s KSSS?

II-1: *Da bi bile vse kolokacije pregledane (višja stopnja zanesljivosti).*

II-4: *Da bi bila pogostost (št. zadetkov) izražena numerično (ob samem zadetku, da ne bi bilo treba iti gledat v Gigafido).*

II-8: *Oznaka o zaznamovanosti kolokacij, povezava do drugih slovarjev znotraj virov CJVT za iskano besedo.*

III-1: *Bolj jasno pokazana pogostost posameznih kolokacij.*

¹⁰ II (1-8) = študenti 2. letnika, III (1-7) = študenti 3. letnika

- III-2: Nekateri vnosi niso pregledani (zanesljivost); filtri – da bi se bilo mogoče vračati, ne samo ožati nabora (da ni treba iti še 1x od začetka).
- III-3: Podrobnejše informacije o virih, iz katerih so kolokacije pridobljene. Zoženje iskalnih parametrov. Nabor kolokacij ni dovolj širok.
- III-4: Velik del vnosov še vedno ni pregledan – včasih naletiš na »nenavadne« primere (zmanjša zanesljivost; zaupanje v sam vir, zlasti slabo za tiste, ki se jezika še učijo).
- III-5: Nekateri besede še nimajo kolokacijskega para (premajhen obseg gesel).
- III-6: Več kolokacij; da bi bila pogostost posameznih kolokacij bolj vidna.
- II-7: Razvrstitev po »tematikah« (filtri, npr. medicina).

Od 15 subjektov raziskave zgolj dva nista seznanjena s tem, da lahko uporabniki tudi sami prispevajo k osveževanju informacij v KSSS s tem, da glasujejo o primernosti že vnaprej strojno pripravljenih kolokacij oz. kolokatorjev (prim. Arhar Holdt, 2021). Zanimivo je, da se jim ta možnost zdi pozitivna, »Ker vsi uporabljamo jezik in ga s tem tudi sooblikujemo« (II-2).

4.2.2.3 KSSS obstaja le v elektronski različici. Se vam zdi zgolj digitalna različica prednost? Če ja, zakaj?

- II-1: Bistveno hitreje in dostopno (lažje posodabljanje).
- II-8: Slovar je možno hitro posodabljati, mesto je neomejeno, zato lahko pridobimo številne podatke, ki jih drugače ne bi mogli (kot na primer vir kolokacije). Uporablja se lahko na vsaki napravi.
- III-1: Hitrost dostopa, napredne možnosti iskanja, preglednost, primeri, povezave.
- III-3: Lažja dostopnost (sama naprava/telefon), lažje iskanje (vpis ključne besede vs. listanje), lažje posodabljanje.
- III-5: Hitrost iskanja; rezultate je lažje primerjati z drugimi viri; hiperpovezave, ki pomagajo pri iskanju; »cross-referencing« (vs. Fran).

4.2.2.4 Kaj vam je pri KSSS najbolj všeč oz. kaj se vam zdi najbolj uporabno?

II-1: Zelo pregleden. »Highfive« za barvo.

II-3: Možnost filtriranja (abecedni vrstni red ali vezljivost).

II-4: Preglednost samega vmesnika.

II-7: Pohvalil bi skalo, ki nakazuje pogostost uporabe.

III-3 To, da slovar ponudi različne kolokacijske pare, informacije o pogostosti rabe in tipu medija.

III-4: Da lahko zožimo izbor zadetkov.

III-6: Hitrejše iskanje; bolj natančno iskanje (ctrl + »f«); možna uporaba več različnih virov istočasno; možnost integracije virov v druge vire.

4.3 Diskusija rezultatov

Uporabniška študija je pokazala, da večina študentov pri prevajalskem delu redno uporablja predvsem elektronske jezikovne vire (podobne ugotovitve najdemo že v Hirci, 2013 in Gorjanc, 2014), kar ne prese- neča, saj se študenti prevajalstva med študijskim procesom pri različ- nih predmetih in prevajalskih seminarjih seznanijo z uporabo najra- zličnejših prevajalskih pripomočkov. Zdi se, da to tudi na splošno velja za mlajšo generacijo prevajalske skupnosti, ki preferira različne oblike spletnih virov in orodij (prim. izsledke Šorli, 2018; za različne uporabni- ške skupine tudi Arhar Holdt, 2021).

Obenem je študija pokazala, da pri reševanju prevodnih težav pri prevajanju kolokacij iz angleščine v slovenski jezik študentje posegajo po raznovrstnih jezikovnih virih in priročnikih, vendar pa so strategije posameznikov pri iskanju prevodnih ustreznih kolokacijskih parov iz- jemno raznolike. Če bi iskalne strategije, ki so jih respondenti navajali v odprtem delu vprašalnika, skušali nekako strniti in posplošiti, bi jih lahko združili v nekaj skupnih načinov iskanja kolokacij, in sicer:

- a. Iskano kolokacijo vpišejo neposredno v spletni iskalnik, nato pre- verijo rabo v enojezičnih slovarjih, pogosto tudi v kolokacijskih slo- varjih ali v korpusnih virih (npr. v korpusu Gigafida).
- b. Iskati začnejo po kolokacijskih slovarjih (npr. na <http://www.free-collocation.com>), nato preverjajo rabo še po vzporednih besedilih v korpusnih virih ali na spletu.

- c. Rabo kolokacije preverijo na slovarski platformi Glosbe, nato dodatno preverijo rabo na spletu ali v korpusu Gigafida.
- d. Iskane kolokacije vtipkajo v strojni prevajalnik (Google Translate ali DeepL) in nato rabo preverijo v spletnih iskalnikih.

Za ta del uporabniške študije povsem primerljivih izsledkov drugih študij ne najdemo, saj ostale raziskave poleg prevajalcev vključujejo tudi druge ciljne skupine (prim. Arhart Holdt, 2020; Pori idr., 2020). Kljub temu pa se je tako v naši študiji kot tudi v drugih uporabniških študijah izkazalo, da se uporabnikom KSSS zdi dobrodošel in koristen jezikovni vir (prim. pogled uporabniške skupine desetih prevajalcev in lektorjev – žal iz rezultatov ni moč razbrati, koliko je bilo enih in koliko drugih, ki so bili vključeni v raziskavo – ki se jim zdi KSSS neprecenljiv vir jezikovnih informacij (Pori idr., 2020, str. 183)).

V drugem delu raziskave, ki se je osredotočala na prevajalski proces ciljne skupine študentov prevajalstva in rabo KSSS se je izkazalo, da so študenti večinoma dobro seznanjeni s KSSS in ga večina tudi uporablja pri svojem prevajalskem delu. Obenem pa je študija pokazala, da je poglobljena napaka študentov z neustreznimi oz. zgolj delno ustreznimi prevodnimi rešitvami za izbrane kolokacije v večini primerov v tem, da so izvornik obravnavali preveč dobesedno, pri čemer niso upoštevali širšega konteksta, ki bi jim omogočil pravilno interpretacijo, čeprav so imeli na voljo celotno besedilo in vse jezikovne vire.

Z vidika rabe KSSS smo pri študentih opazili dva trenda. Nekateri so v KSSS želeli zgolj preveriti, ali določena kombinacija besed predstavlja kolokacijo oz. kako pogosta je. To je bilo razvidno iz dogajanja na zaslonu, saj so določeno sosledje besed najprej zapisali kot del prevodne rešitve in ga šele nato preverili v KSSS – ta trend je prevladoval pri študentih, katerih prevodne rešitve so bile ocenjene kot ustrezne. Drugi način rabe KSSS je bil manj strukturiran, saj so študenti vnesli jedro, nato pa pregledovali celoten seznam kolokacij, pri čemer se je zdelo, da ne iščejo točno določene kolokacije.

Iz tega je mogoče izpeljati vsaj dve različni razlagi:

1. študenti so točno vedeli, katero kolokacijo želijo preveriti, vendar niso večji rabe filtrov, ki bi jim omogočili, da želeno kolokacijo hitreje najdejo;

2. študenti niso imeli v mislih nobene konkretne kolokacije in so želeli pregledati širok seznam najrazličnejših kolokacijskih rešitev, izmed katerih bi nato lahko izbrali tisto, ki bi se jim glede na kontekst zdelo najprimernejša.

Poleg prej omenjenih trendov iskanja ustreznih kolokacij je bilo mogoče opaziti tudi primer neuspešnega iskanja, kjer se eden od sodelujočih ni zavedal ustroja KSSS, saj je kot iskalni niz vpisal celo kolokacijo namesto zgolj jedra, kjer bi nato z rabo dodatnih filtrov lahko hitro našel potrditev iskane kolokacije. To se v določenem pogledu sklada z ugotovitvami Pori idr. (2020, str. 191), ki izpostavljajo, da je pri delu s študenti treba pozorno in konsistentno spremljati njihovo delo in jih opremiti z jasnimi navodili – tako je v izobraževalnem procesu nujno vedno znova vključevati ta aspekt tudi pri osveščanju o rabi KSSS.

V polstrukturiranem intervjuju po prevajalski nalogi, ki nam je omogočil dopolnitev vrzeli pri razumevanju dogajanja na zaslonu, je bilo poleg pozitivnih lastnosti KSSS izpostavljenih tudi nekaj predlogov za izboljšave. Nekatero podobno predlogo za izboljšave med različnimi uporabniškimi skupinami lahko najdemo tudi npr. v Arhar Holdt (2020) in Pori idr. (2020), kjer je med drugim izpostavljeno, da bi bilo dobro v KSSS dodati frekventnost kolokacij, možnost izvoza podatkov in možnost vračanja na osnovno stran zgolj s klikom na kolokacijo. Navajamo nekaj študentskih predlogov za izboljšave KSSS, in sicer:

- *da bi bile vse kolokacije pregledane in bi se s tem zvišala stopnja zanesljivosti najdenih informacij (manjša stopnja zanesljivosti niža zaupanje v sam vir, kar je še zlasti slabo za tiste, ki se jezika še učijo);*
- *da bi bila pogostost (št. zadetkov) bolj vidna, jasno izražena, morda numerično ob posameznem zadetku;*
- *da bi bila zaznamovanost kolokacij posebej označena;*
- *da bi KSSS imel neposredno povezavo do drugih slovarjev znotraj virov CJVT za iskano besedo;*
- *da bi se pri filtrih bilo mogoče vračati, ne samo ožati nabora;*
- *da bi bilo navedeno nekaj informacij o tem, od kod so kolokacije pridobljene;*
- *da bi se lahko iskalne parametre ožalo;*
- *premajhen obseg gesel (nabor kolokacij ni dovolj širok, nekatere*

besede še nimajo kolokacijskega para), da bi bile kolokacije razvrščene tematsko (s pomočjo različnih filtrov, npr. za medicino, pravo, marketing).

Pričujoča uporabniška študija je s pomočjo triangulacije podatkov (prim. poziv Šorli (2018, str. 248), da bi veljalo več napora vložiti v »kombinirane raziskave, ki bi uporabljale tako metode opazovanja kot eksperimente in anketiranje«), ki smo jih pridobili s kombinacijo različnih metodoloških pristopov, kjer so se informacije dopolnjevale (vprašalnik, opazovanje prevajalskega procesa in polstrukturirani intervju po prevajalski nalogi) potrdila, da so študenti prevajalstva v Ljubljani s KSSS ne le seznanjeni, temveč ta jezikovni vir večinoma tudi aktivno uporabljajo pri svojem prevajalskem delu. Vendar pa je obenem treba opozoriti, da je opazovan vzorec majhen in je težko posploševati izsledke, zato bi bilo nujno vložiti več naporov v dodatno raziskovanje te uporabniške skupine.

Obenem se je v študiji izkazalo, da nekateri uporabniki KSSS premalo poznajo njegov ustroj in napredne možnosti, ki jih ta jezikovni vir ponuja, zato njegovega potenciala ne morejo oz. ne znajo docela izkoristiti. Izkazalo se je tudi, da je poglobitveni dejavnik, ki vpliva na uspešnost rabe določenega prevajalskega vira, jezikovno znanje, ki omogoča kritično presojo o tem, katero slovarsko informacijo je najbolj smiselno sploh iskati in katero nato uporabiti v določenem kontekstu. Tako študija izpostavlja, da je nujno dodatno osveščanje študentov prevajalstva za kompetentno rabo ustreznih jezikovnih virov.

5 Sklep

Od bodočih prevajalcev se pričakuje, da so v kar največji meri digitalno osveščeni, opremljeni z ustreznim tehnološkim znanjem, predvsem pa seznanjeni z relevantnimi prevajalskimi pripomočki in jezikovnimi viri, ki jim lahko olajšajo prevajalsko delo. Zato je smiselno že v času študijskega procesa vlagati čim več naporov v to, da študentsko populacijo pripravimo na prevajalski trg dela in jih primerno digitalno usposobimo, obenem pa neprestano spodbujamo njihovo kritično razmišljanje in raziskovalno radovednost, da v prevajalskem procesu vedno skušajo najti optimalne prevodne rešitve, ki so najprimernejše v danem času in kontekstu.

V članku je predstavljena uporabniška študija med študenti prevajalstva v Ljubljani, kjer smo s pomočjo kvantitativne analize vprašalnika želeli osvetliti uporabniške izkušnje študentov z različnimi jezikovnimi viri, še posebej pri prevajanju kolokacij, zato smo se v kvalitativnem delu analize osredotočili na uporabniške izkušnje s KSSS. Rezultati študije so pokazali, da se študenti zavedajo prednosti reševanja prevodnih težav s pomočjo jezikovnih priročnikov in prevajalskih virov. Pri prevajanju kolokacij in razreševanju prevodnih zagat pri prenosu iz angleškega v slovenski jezik se primarno naslanjajo na spletne kolokacijske slovarje, ostalih virov informacij se poslužujejo manj pogosto. Pri prevajanju kolokacij pogosto posegajo tudi po *Kolokacijskem slovarju sodobne slovenščine*, s katerim se seznanijo že med študijskim procesom. Vendar pa njihovo iskanje ne pripelje nujno do ustreznih prevodnih rešitev. Nedvomno so kolokacijski slovarji eden od ključnih gradnikov uspešnega prevajalskega dela, zato bi bilo smiselno v luči rezultatov ankete v pedagoškem procesu vlagati dodatne napore predvsem v razvijanje kritičnega ovrednotenja najdenih jezikovnih informacij in v ozaveščanje bodočih prevajalcev o tem, kako uspešno iskati po relevantnih jezikovnih virih, da lahko na koncu najdejo žanrsko in kontekstualno smiselne in ustrezne prevodne rešitve.

Zahvala

Prispevek, ki je nastal v okviru raziskovalnega programa št. P6-0215 (*Slovenski jezik – bazične, kontrastivne in aplikativne raziskave*), je sofinancirala Javna agencija za raziskovalno dejavnost Republike Slovenije iz državnega proračuna.

Literatura

- Alves, F. (ur.). (2003). *Triangulating Translation: Perspectives in process oriented research*. Amsterdam/Philadelphia: John Benjamins Publishing.
- Arhar Holdt, Š. (2015). Uporabniške raziskave za potrebe slovenskega slovarpisja: prvi koraki. V V. Gorjanc, P. Gantar, I. Kosem in S. Krek (ur.), *Slovar sodobne slovenščine: problemi in rešitve* (str. 136–148). Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani.

- Arhar Holdt, Š. (2020). How Users Responded to a Responsive Dictionary: The Case of the Thesaurus of Modern Slovene. *Rasprave: Časopis Instituta za hrvatski jezik i jezikoslovlje*, 46(2), 465–482. doi: 10.31724/rihjj.46.2.1
- Arhar Holdt, Š. (2021). Razvrstitev kolokacij v slovarskem vmesniku: uporabniške prioritete. V I. Kosem (ur.), *Kolokacije v slovenščini*. 1. izd. Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani. Pridobljeno s <https://ebooks.uni-lj.si/ZalozbaUL/catalog/view/318/465/6968>
- Arhar Holdt, Š., & Čibej, J. (2020). Rezultati projekta »Slovar sodobne slovenščine: od skupnosti za skupnost«. V D. Fišer in T. Erjavec (ur.), *Jezikovne tehnologije in digitalna humanistika: zbornik konference: 24.–25. september, 2020, Ljubljana*, (str. 3–9). Ljubljana: Inštitut za novejšo zgodovino.
- Arhar Holdt Š., Čibej J., & Zwitter Vitez, A. (2015). S pomočjo uporabniških jezikovnih vprašanj in mnenj do boljšega slovarja. V V. Gorjanc, P. Gantar, I. Kosem in S. Krek (ur.), *Slovar sodobne slovenščine: problemi in rešitve* (str. 196–214). Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani. Pridobljeno s <https://ebooks.uni-lj.si/zalozbaul//catalog/download/15/47/500-1?inline=1>
- Arhar Holdt, Š., Čibej, J., Dobrovoljc, K., Gantar, P. Gorjanc, V., Klemenc, B., ..., & Robnik Šikonja, M. (2018). Thesaurus of Modern Slovene: By the Community for the Community. V J. Čibej, V. Gorjanc, I. Kosem in S. Krek (ur.), *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts* (str. 401–410). Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani. Pridobljeno s <https://e-knjige.ff.uni-lj.si/znanstvena-zalozba/catalog/view/118/211/3000-1>
- Arhar Holdt, Š., Dobrovoljc, K., & Logar, N. (2019). Simplicity matters: user evaluation of the Slovene reference corpus. *Language resources and evaluation*, 53(1), 173–190. Pridobljeno s <https://link.springer.com/article/10.1007/s10579-018-9429-8>
- Arhar Holdt, Š., Kosem, I., & Gantar, P. (2016). Dictionary user typology: the Slovenian case. V T. Margalitadze idr. (ur.), *Lexicography and linguistic diversity: proceedings of the XVII EURALEX International Congress, 6–10 September, 2016* (str. 179–187). Tbilisi: Ivane Javakhishvili Tbilisi State University. Pridobljeno s http://euralex.org/wp-content/themes/euralex/proceedings/Euralex%202016/euralex_2016_015_p179.pdf
- Atkins, B. T. S. (ur.). (1998). *Using Dictionaries*. Tübingen: Niemeyer.
- Atkins, B. T. S., & Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. New York: Oxford University Press.

- Atkins, B. T. S., & Varantola, K. (1998). Monitoring dictionary use. V B. T. S. Atkins (ur.), *Using Dictionaries: Studies of Dictionary Use by Language Learners and Translators* (str. 83–122). Tübingen: Max Niemeyer Verlag.
- Béjoint, H. (1981). The Foreign Student's Use of Monolingual English Dictionaries: A Study of Language Needs and Reference Skills. *Applied Linguistics*, 2(3), 207–221.
- Béjoint, H. (2000). *Modern Lexicography. An Introduction*. Oxford: Oxford University Press.
- Bernardini, S. (2001). Think-aloud protocols in translation research. Achievements, limits, future prospects. *Target* 13(2), 241–263.
- Campoy Cubillo, M. C. (2001). Dictionary use and dictionary needs of ESP students: An experimental approach. *International Journal of Lexicography*, 15(3), 206–228.
- Čibej, J., Gorjanc, V., & Popič, D. (2015). Vloga jezikovnih vprašanj prevajalcev pri načrtovanju novega enojezičnega slovarja. V V. Gorjanc, P. Gantar, I. Kosem in S. Krek (ur.), *Slovar sodobne slovenščine: problemi in rešitve* (str. 168–181). Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani. Pridobljeno s <https://ebooks.uni-lj.si/zalozbaul/catalog/download/15/47/498-1?inline=1>
- Fišer, D. (2008). Recent trends in the translation industry in Slovenia. *The Journal of Specialised Translation*, 10, 23–39. Pridobljeno s https://www.jostrans.org/issue10/art_fiser.pdf
- Gantar, P., Krek, S., & Kosem, I. (2021). Opredelitev kolokacij v digitalnih slovarskih virih za slovenščino. V I. Kosem (ur.), *Kolokacije v slovenščini* (str. 15–41). Ljubljana: Filozofska fakulteta Univerze v Ljubljani; Institut Jožef Stefan. Pridobljeno s <https://ebooks.uni-lj.si/zalozbaul/catalog/download/318/465/6969-1?inline=1>
- Gorjanc, V. (2014). Slovar slovenskega jezika v digitalni dobi. V I. Grahek in S. Bergoč (ur.), *E-zbornik Posveta o novem slovarju slovenskega jezika na Ministrstvu za kulturo*. Ljubljana: Ministrstvo za kulturo RS. Pridobljeno s <https://www.dlib.si/stream/URN:NBN:SI:DOC-UANHTHW/3e8cf8a5-2840-4186-aeef-cdee17d5de71/PDF>
- Gorjanc, V., Gantar, P., Kosem, I., & Krek, S. (ur.). (2015). *Slovar sodobne slovenščine: problemi in rešitve*. Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani. doi: 10.4312/9789612379759
- Grad, M., & Hirci, N. (2022). Raba kolokacijskega slovarja sodobne slovenščine pri prevajanju kolokacij. V D. Fišer in T. Erjavec (ur.), *Jezikovne tehnologije in digitalna humanistika: zbornik konference: 15.–16. september, 2022*,

- Ljubljana, Slovenija (str. 63–70). Ljubljana: Inštitut za novejšo zgodovino. Pridobljeno s https://nl.ijs.si/jtdh22/pdf/JTDH2022_Proceedings.pdf
- Granger, S., & Paquot, M. (ur.). (2012). *Electronic Lexicography*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199654864.003.0001
- Gries, S. T. (2013). 50-something years of work on collocations: What is or should be next ... *International Journal of Corpus Linguistics*, 18(1), 137–166.
- Hansen, G. (ur.). (1999). *Probing the Process in Translation: Methods and Results*. Copenhagen: Copenhagen Business School.
- Hansen, G. (2003). Controlling the process: Theoretical and methodological reflections on research into translation process. V F. Alves (ur.), *Triangulating Translation: Perspectives in process oriented research* (str. 25–42). Amsterdam/Philadelphia: John Benjamins Publishing.
- Hausmann, F. J. (1984). Wortschatzlernen ist Kollokationslernen. Zum Lehren und Lernen französischer Wortverbindungen. *Praxis des neusprachlichen Unterrichts*, 31(4), 395–406. Dortmund: Lensing.
- Hirci, N. (2007). *Učinkovitost uporabe sodobnih prevajalskih virov pri prevajanju v nematerni jezik*. (Doktorska disertacija). Ljubljana: Filozofska fakulteta Univerze v Ljubljani.
- Hirci, N. (2009a). Uporaba sodobnih prevajalskih virov pri izobraževanju prevajalcev. V V. Mikolič (ur.), *Jezikovni korpusi v medkulturni komunikaciji* (str. 57–87). Koper: ZRS Koper, Založba Annales.
- Hirci, N. (2009b). Empirične raziskovalne metode za opazovanje prevajalskega procesa. V N. Kocijančič-Pokorn (ur.), *Sodobne metode v prevodoslovnem raziskovanju* (str. 60–78). Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani. Pridobljeno s <https://ebooks.uni-lj.si/zalozbaul/catalog/download/29/77/745-1?inline=1>
- Hirci, N. (2013). Changing trends in the use of translation resources: the case of trainee translators in Slovenia. *ELOPE*, 10(2), 149–165. Ljubljana: Slovensko društvo za angleške študije.
- Humblé, P. (2001). *Dictionaries and Language Learners*. Frankfurt am Main: Haag und Herchen.
- Jääskeläinen, R. (1999). *Tapping the process: an explorative study of the cognitive and affective factors involved in translation*. Joensuu: Joensuun yliopisto.
- Jakobsen, A. L. (2003). Effects of think-aloud on translation speed, revision, and segmentation. V F. Alves (ur.), *Triangulating Translation: Perspectives in process oriented research* (str. 69–95). Amsterdam/Philadelphia: John Benjamins Publishing.

- Jensen, A., & Jakobsen, A. L. (2000). Translating under time pressure. V A. Chesterman, G. S. S. Gallardo in Y. Gambier (ur.), *Translation in Context* (str. 105–116). Amsterdam/Philadelphia: John Benjamins Publishing.
- Kocijančič Pokorn, N. (2016). Nič več obljubljen dežela: dinamični premiki na slovenskem prevajalskem trgu in področju izobraževanja prevajalcev. *Vestnik za tuje jezike*, 8(1), 9–21. Pridobljeno s <http://revije.ff.uni-lj.si/Vestnik/article/view/7174/6878>
- Kosem, I. (ur.). (2021). *Kolokacije v slovenščini*. Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani. doi: 10.4312/9789610605379
- Kosem, I., Gantar, P., Krek, S., Arhar Holdt, Š., Čibej, J., Laskowski, C. A., Pori, E., ..., & Ljubešić, N. (2019). *Collocations dictionary of modern Slovene KSSS 1.0*. Ljubljana: Centre for Language Resources and Technologies, University of Ljubljana. Pridobljeno s <https://www.clarin.si/repository/xmlui/handle/11356/1250>.
- Kosem, I., Krek, S., Gantar, P., Arhar Holdt, Š., Čibej, J., & Laskowski, C. (2018a). Kolokacijski slovar sodobne slovenščine. V D. Fišer in A. Pančur (ur.), *Zbornik konference Jezikovne tehnologije in digitalna humanistika, 20. september–21. september, 2018, Ljubljana, Slovenija* (str. 133–139). Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani. Pridobljeno s <http://nl.ijs.si/jtdh18/JTDH-2018-Proceedings.pdf>
- Kosem, I., Krek, S., Gantar, P., Arhar Holdt, Š., Čibej, J., & Laskowski, C. (2018b). Collocations dictionary of modern Slovene. V J. Čibej, V. Gorjanc, I. Kosem in S. Krek (ur.), *Proceedings of the 18th EURALEX International Congress: lexicography in global contexts, 17–21 July 2018, Ljubljana* (str. 989–997). Ljubljana: Znanstvena založba Filozofske fakultete. Pridobljeno s <https://ebooks.uni-lj.si/zalozbaul/catalog/view/118/211/3000-1>
- Krek, S., Arhar Holdt, Š., Erjavec, T., Čibej, J., Repar, A., Gantar, P., Ljubešić, N., Kosem, I., & Dobrovoljc, K. (2020). Gigafida 2.0: the reference corpus of written standard Slovene. V N. Calzolari (ur.), *LREC 2020: Twelfth International Conference on Language Resources and Evaluation: May 11–16, 2020, Marseille, France* (str. 3340–3345). Paris: ELRA – European Language Resources Association. Pridobljeno s <http://www.lrec-conf.org/proceedings/lrec2020/LREC-2020.pdf>
- Livbjerg, I., & Mees, I. M. (2003). Patterns of dictionary use in non-domain-specific translation. V F. Alves (ur.), *Triangulating Translation* (str. 123–136). Amsterdam/Philadelphia: John Benjamins. Publishing: John Benjamins Publishing.
- Logar Berginc, N. (2009). Slovenski splošni in terminološki slovarji: za koga? V M. Stabej (ur.), *Infrastruktura slovenščine in slovenistike. Obdobja 28* (str.

- 225–231). Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani.
- Logar Berginc, N., Grčar, M., Brakus, M., Erjavec, T., Arhar Holdt, Š., & Krek, S. (2020). *Korpusi slovenskega jezika Gigafida, KRES, ccGigafida in ccKres: gradnja, vsebina, uporaba*. Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani. Pridobljeno s <https://ebooks.uni-lj.si/ZalozbaUL/catalog/view/233/333/5394>
- Mackintosh, K. (1998). An empirical study of dictionary use in L2-L1 translation. V B. T. S. Atkins (ur.), *Using Dictionaries* (str. 123–149). Tübingen: Niemeyer.
- Magay, T., & Zigány, J. (ur.). (1990). *Papers from the Third International EURALEX Congress*. Budapest: Akademiai Kiado.
- Mikolič, V. (ur.). (2009). *Jezikovni korpusi v medkulturni komunikaciji*. Koper: ZRS Koper, Založba Annales.
- Mikolič, V. (2015). Slovarski uporabniki – ustvarjalci: ustvarjati v jeziku in z jezikom. V V. Gorjanc, P. Gantar, I. Kosem in S. Krek (ur.), *Slovar sodobne slovenščine: problemi in rešitve* (str. 182–195). Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani. doi: 10.4312/9789612379759
- Mikolič Južnič, T. (2009). Vzporedni korpus – prevajalsko orodje in orodje za jezikoslovne analize. V V. Mikolič (ur.), *Jezikovni korpusi v medkulturni komunikaciji* (str. 117–132). ZRS Koper: Založba Annales.
- Müller-Spitzer, C., Kopleinig, A., & Töpel, A. (2012). Online dictionary use: Key findings from an empirical research project. V S. Granger in M. Paquot (ur.), *Electronic Lexicography* (str. 425–457). Oxford: Oxford University Press.
- Müller-Spitzer, C. (ur.). (2014). *Using Online Dictionaries*. Berlin, Boston: De Gruyter Mouton.
- Pisanski Peterlin, A. (2003). Uporaba novih tehnologij pri jezikovnem pouku. *Jezik in slovstvo*, 48(3–4), 103–112. Ljubljana: Slavistično društvo Slovenije. Pridobljeno s <https://www.dlib.si/stream/URN:NBN:SI:DOC-ZUKI-W49K/da6aa8c2-e33a-4fe5-8124-07274d7a8706/PDF>
- Pori, E., Čibej, J., Kosem, I., & Arhar Holdt, Š. (2020). The attitude of dictionary users towards automatically extracted collocation data: A user study. *Slovenščina 2.0: Empirical, Applied and Interdisciplinary Research*, 8(2), 168–201. doi: 10.4312/Slo2.0.2020.2.168-201
- Pori, E., Kosem, I., Čibej, J., & Arhar Holdt, Š. (2021). Evalvacija uporabniškega vmesnika Kolokacijskega slovarja sodobne slovenščine. V I. Kosem (ur.), *Kolokacije v slovenščini* (str. 235–268). Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani. Pridobljeno s <https://ebooks.uni-lj.si/ZalozbaUL/catalog/view/318/465/6973>

- Roberts, R. P. (1992). Translation pedagogy: strategies for improving dictionary use. *TTR: Traduction, Terminologie et Redaction*, 5(1), 49–76.
- Roberts, R. P. (1997). Using dictionaries efficiently. *38th Annual Conference of the American Translators Association, San Francisco, California* (str. 283–296).
- Rozman, T. (2004). Upoštevanje ciljnih uporabnikov pri izdelavi enojezičnega slovarja za tujce. *Jezik in slovstvo*, 49(3/4), 63–75. Pridobljeno s <http://www.dlib.si/stream/URN:NBN:SI:doc-J53E16B1/fa0dd317-179a-4743-ad15-6e56fd00c4cf/PDF>
- Sánchez Ramos, M. M. (2005). Research on Dictionary Use by Trainee Translators. *Translation Journal*, 9(2). Pridobljeno s <https://translationjournal.net/journal/32dictuse.htm>
- Stabej, M. (2009). Slovarji in govorci: kot pes in mačka? *Jezik in slovstvo*, 54(3/4), 115–138. Pridobljeno s <https://www.jezikinslovstvo.com/pdf/2009-03-04-Razprave-Marko-Stabej.pdf>
- Starren, P. in Thelen, M. (1990). General dictionaries and students of translation: A report on the use of dictionaries in the translation process. V T. Magay in J. Zsigány (ur.), *Papers from the Third International EURALEX Congress* (str. 447–458). Budapest: Akademiai Kiado.
- Šorli, M. (2018). Uporabnik jezikovnih virov in orodij kot objekt ali subjekt jezikovne politike? – področje prevajanja in tolmačenja. *Slavia Centralis*, 11(2), 233–250. Pridobljeno s <https://www.dlib.si/stream/URN:NBN:SI:DOC-042353EB/f7387c8e-6882-4233-a62c-66ca01a6c016/PDF>
- Šorli, M., & Ledinek, N. (2017). Language policy in Slovenia: language users' needs with a special focus on lexicography and translation tools. V I. Kosem, C. Tiberius, M. Jakubíček, J. Kallas, S. Krek in V. Baisa (ur.), *Proceedings of eLex 2017: Lexicography from Scratch, 19–21 September 2017, Leiden, Netherlands* (str. 377–394). Pridobljeno s <https://elex.link/elex2017/wp-content/uploads/2017/09/paper23.pdf>
- Tarp, S. (2009). Reflections on Lexicographical User Research. *Lexikos*, 19(1), 275–296.
- Varantola, K. (1998). Translators and their use of dictionaries. V B. T. S. Atkins (ur.), *Using Dictionaries* (str. 179–192). Tübingen: Niemeyer.
- Vintar, Š. (1999). Računalniške tehnologije za prevajanje. *Uporabna informatika*, 7(1), 17–24.

The use of language resources and the *Collocations Dictionary of Modern Slovene* among students of translation

The article presents a user study conducted with trainee translators from the University of Ljubljana on their use of language resources and translation tools. In the digital age, future translators are expected to be tech savvy and well-informed about the relevant sources of information. A survey was designed therefore to explore trainee translators' attitude towards English and Slovene language resources and their strategies to overcome translation challenges, while a follow-up study focuses on the translation process when translating collocations from English to Slovene and their use of relevant language resources, especially the *Collocations Dictionary of Modern Slovene*. The results of the study indicate that the trainee translators are aware of the benefits that relevant language resources and other translation tools bring to the translation process. The study also revealed that introducing the *Collocations Dictionary of Modern Slovene* into the translation process has been fairly successful. However, additional emphasis should be placed in translator training on the advanced features this collocations dictionary offers as well as on developing appropriate search skills. Moreover, trainee translators' critical assessment of the solutions should also be encouraged to facilitate finding translation solutions best suited to a given context.

Keywords: user study, use of language resources and translation tools, translating collocations, *Collocations Dictionary of Modern Slovene*, raising awareness of trainee translators

Priloga I: Besedilo prevajalske naloge z označenimi kolokacijami (443 besed)

Scientists revise asteroid warning

Astronomers say reports that the Earth could be struck by a small asteroid in 2030 are **(1) wildly exaggerated**.

Less than a day after **(2) sounding the alert** about asteroid 2000SG344, a **(3) revised analysis** of the space rock's orbit shows it will in fact miss the Earth by about five million kilometres. However, astronomers will continue to monitor the asteroid, which was picked up in September and thought to be 30-70 metres across.

Some scientists have criticised the way the **(4) information was released** to the media before it had been thoroughly confirmed.

(5) Threat rating

Asteroid 2000SG344 is the first object to have a threat rating greater than zero on the 0-10 Torino scale of dangerous objects from space. It was spotted on 29 September by astronomers David Tholen and Robert Whiteley using the Canada-France-Hawaii 3.6-metre telescope on the island of Hawaii.

Shortly thereafter, pre-discovery observations taken in May 1999 by the Linear sky survey were also identified.

On Friday, the International Astronomical Union **(6) issued an alert** saying that the object had about a 1-in-500 chance of striking the Earth on 21 September 2030. No object has ever been rated with so high a chance of impact. Were it to strike our planet, the **(7) results** would be **devastating**, with an explosion greater than the most powerful nuclear weapon.

Sky survey data

But after the announcement, astronomers began looking at sky survey data to see if the object had been picked up but not recognised in earlier observations. This turned out to be the case and these past observations allowed a more accurate calculation of the asteroid's orbit to be made.

The result: in 2030, the space rock will miss us by about five million kilometres, or 12-times the distance from the Earth to the Moon. The new orbit reveals **(8) a slight risk** of a collision with the Earth about 2071, but it is thought that when the orbit is better known this risk will disappear as well. Currently, asteroid 2000SG344 is about 15 million kilometres away and getting more distant.

‘Premature and alarmist’

Because 2000SG344 is in a similar orbit to the Earth, it has been suggested that it might be an old Saturn upper-stage rocket of the type that was used in the early Apollo Moon missions. If it is manmade and did strike Earth, the **(9) effects** would be very **local** and **limited**.

Some scientists have criticised the IAU and Nasa for releasing warnings about the asteroid only for those **(10) warnings** to be **re-scinded** less than a day later. Benny Peiser of Liverpool John Moores University, UK, said it was “extremely unwise, premature and alarmist”.

Priloga II: Polstrukturirani intervju po prevajalski nalogi

Ste seznanjeni s KSSS? Ga pogosto uporabljate?

- ja
- ne

Se vam KSSS zdi koristen pri prevajalskem delu?

- ja
- ne

Če ja, zakaj?

Če ne, zakaj?

Katere informacije, ki jih KSSS v zdajšnji obliki še ne vsebuje, bi lahko pripravile k izboljšani uporabniški izkušnji s KSSS?

Ali veste, da lahko tudi sami prispevate k novim vnosom?

- ja
- ne

Se vam to zdi pozitivno?

- ja
- ne

Zakaj?

Vas to motivira, da bi morda tudi sami dodali kakšna gesla ali jih dopolnili?

- ja
- ne

Če ste odgovorili z »ja«, na kratko opišite, kaj vas je do tega privedlo.

KSSS obstaja le v elektronski različici. Se vam zdi zgolj digitalna različica prednost?

- ja
- ne

Zakaj?

Kaj vam je bilo pri slovarju najbolj všeč oz. se vam je zdelo najbolj uporabno?

Kaj se vam zdi, da slovarju manjka? Kaj bi si želeli spremeniti?

From verbal to adjectival: evaluating the lexicalization of participles in an Estonian corpus

Geda PAULSEN

Institute of the Estonian Language; University of Uppsala

Maria TUULIK

Institute of the Estonian Language

Ahti LOHK

Institute of the Estonian Language; Tallinn University of Technology

Ene VAINIK

Institute of the Estonian Language

This study addresses categorization issues related to adjective candidates in Estonian, focusing on the category of participles. The aim of the analysis was to assess the ranges of the prototypical adjective and to determine its degree of deviation on the prototypicality scale. The investigation was based on a group of validated adjectives – selected adjectives included in the *Basic Estonian Dictionary* – and two control groups of more and less lexicalized participles. We tested seven morphosyntactic corpus patterns characteristic of adjectives. The test patterns were based on the prototypical features of the adjective, as well as

Paulsen, G., Tuulik, M., Lohk, A., Vainik, E.: From verbal to adjectival: evaluating the lexicalization of participles in an Estonian corpus. Slovenščina 2.0, 10(1): 65–97.

1.01 Izvirni znanstveni članek / Original Scientific Article

DOI: <https://doi.org/10.4312/slo2.0.2022.1.65-97>

<https://creativecommons.org/licenses/by-sa/4.0/>



on observations made in the actual lexicographic analysis. To assess the sample words and determine the significance of the test patterns from the point of view of defining adjectivity, we used deviation analysis. The results of this study can be applied to establish a measure of adjectivity for lexicographic judgments when distinguishing, for instance, lexicalized participles from regular ones.

Keywords: corpus linguistics, lexicography, Estonian language, adjective, participle, deviation analysis

1 Introduction

A morphosyntactic analysis, from the corpus linguistics perspective, is a basic operation using inflectional paradigms and a base lexicon in part-of-speech disambiguation of the tokens in a text. For lexicographers, who first determine a word's lexical categorization, the morphosyntactic behavior of a lexeme in its natural contexts is essential information when judging its lexical classification. The data available in corpora yield potential new headwords, but automatic pre-processing is needed in order to make properly weighed decisions about the lexical affiliation of potential new lexemes, as the amount of material may be overwhelming.

In this study, we applied language technology and statistical analysis in order to aid lexicographers in structuring potential headwords. Our target lexical category is adjectives – we attempt to establish the ranges of similarity to the prototypical adjective based on a representative sample of predefined adjectives and to identify degrees for classifying a word (form) as an adjective. The background assumption driving the analysis is that an adjective is not a clearcut but rather a prototype-based category. To ascertain the adjectival core, we developed an evaluation methodology to assess the similarity of a word to an adjective, based on morphosyntactic corpus behavior. In other words, we seek to determine the tolerance ranges of the parameter values that distinguish adjectives from other words and can be used for comparison with the corresponding values of unclear cases.

To test the characteristic attributes of the Estonian adjective, we use a set of corpus patterns based on parameters that include morphological and syntactic features highlighted in the linguistic literature and

detectable in the corpus. The current study is also a re-evaluation of the test patterns used in our previous work (Tuulik et al., 2022). To capture a wider scope of adjectival corpus behavior, we introduce a new pattern: the predicative pattern.

We start our investigation with the rationale of the study and an overview of the Estonian adjective, as well as the participial categories, as described in Section 2. Here we identify the most relevant morphosyntactic properties of adjectives and participles as described in the literature. We proceed with the formation of a random sample of 100 adjectives from the headwords of the Estonian Basic Dictionary (Kallas et al., 2014; see also Kallas and Tuulik, 2011, Kallas et al., 2014), constituting the test group of “adjectives” as validated by lexicographers. To compare the prototypical adjectives with a close, but less clearly adjectival category, we contrast adjectives with participles displaying different degrees of lexicalization. To that end, we created two control groups of equal size: 1) participial independent headword candidates from the lexicographic database of Ekilex, and 2) a sample of regular verbal participles formed of common verbs. The hypothesis behind the composition of the two participial groups is that the participles group of the Ekilex entries are more lexicalized and resemble the reference group of validated adjectives to a larger extent than the regular participles.

To capture the adjectival corpus behavior, we elaborated test patterns detectable in the Estonian National Corpus 2019 and extracted frequency data on the sample words. The test samples, test patterns, methods applied to the data extraction and statistical processing are described in Section 3. Section 4 is devoted to the analysis of the extracted data. The absolute frequency data will be relativized, and the measurement of the adjectival corpus behavior and its limits are described in Section 4.1. After establishing the tolerance ranges of the validated adjectives, the respective values of the control groups will be related to these limits and the degrees of deviation will be calculated in Section 4.2. In Section 4.3, we evaluate the differentiation efficiency of each pattern. A concluding discussion of the results is given in Section 5.

2 Background

2.1 Adjectives and lexical decategorization in Estonian lexicography

This study was motivated by a challenge in the Estonian lexicography: the need to add PoS labels to a vast number of still under-specified keywords of the *Combined Dictionary of the Institute of the Estonian Language* (CombiDic). The current direction in Estonian lexicography is a unification of lexical resources (dictionaries and term bases) into a central superdictionary, the online public dictionary CombiDic. This process is supported by the dictionary writing system Ekilex. At the same time, lexicographic work is moving constantly towards a higher degree of automation and processing of corpora (Tavast et al., 2018; Koppel et al., 2019; Tavast et al., 2020).

A result of the automated processing of lexicographic data is that the lexical database of Ekilex includes automatically generated lists of dictionary entry candidates, requiring assessment of their degree of grammaticalization and/or lexicalization. The data are integrated from different sources,¹ containing words or word forms with different lexicographic statuses:

- a) those not included in the CombiDic (*the CombiDic candidates*),
- b) those included as headwords but without information about their lexical category (*under-specified headwords*),
- c) those included in the CombiDic as headwords with PoS label(s) (*PoS-tagged headwords*).

Providing the underspecified Ekilex entries with PoS tags and assessing the CombiDic candidates for their potential status as lexical entries is an urgent lexicographic issue. Today, 72% (N = 255 691) of the total number of the public CombiDic keywords are missing PoS tags.² A survey of Estonian lexicographers (Paulsen, Vainik, Tuulik, Lohk 2019;

1 For instance, the participle forms included in one of the control groups (the Ekilex participles) of this study derive mainly from the databases of the *Estonian Collocations Dictionary* (2019) and the *Estonian-Russian Dictionary* (2018).

2 This value stems from an excerpt from all Ekilex databases (dictionaries, term bases and phrase collections) done by Kaur Männiko on 24. 1.2022.

Paulsen, Vainik, Tuulik 2020) revealed a need for automatic corpus-based solutions to determine the word class affiliation of a lexeme when there is more than one possible interpretation. Adjectives were pointed out as one of the most complicated categories, in particular the specification of participle forms as either verbal or adjectival (Paulsen et al., p. 188–189).

In a previous study (Tuulik et al., 2022), we tested six morphosyntactic corpus patterns that could differentiate adjectives from other words in 12 groups of words. Six groups of the selected words represented “neighboring” categories of adjectives (prototypical adjectives, less prototypical adjectives, adjectival participles, substantival adjectives, adverbial adjectives, and non-declinable adjectives; with regard to those categories closely related to adjectives, see Vainik et al., 2020, p. 122–123). In addition, we used six control groups representing clear cases of other word classes (verbal participles, substantives, adverbs, verbs, proadjectives and ordinals). All test groups contained 10 words each.

The tested parameters were, to different degrees, able to differentiate adjectival morphosyntactic behavior (Tuulik et al., 2022, p. 295–298). The next step was to measure the scalability of the adjectival behavior on a more representative set of adjectives and establish the tolerance ranges of prototypical adjectives. Since the sample of prototypical adjectives in the previous study was rather small ($N = 10$), the parameters had to be tested on a larger sample of adjectives that represent the best examples of their category. In this study, we focused on the overlapping area between adjectives and verbs. We chose participles as the contrasting test group to the prototypical adjectives for three reasons:

- 1) as a morphosyntactically close lexical group to adjectives, it is theoretically significant to examine where exactly participles differ from adjectives,
- 2) participles constitute one of the most problematic categorization issues for lexicographers,
- 3) participles are substantially represented among those words without clear lexicographic status in the Ekilex database ($N = 1,542$ in January 2022).

Theoretically, we rely on the prototype-based approach to linguistic categories. The latter was initially employed in the study of the internal structure of categories in experimental psychology by Eleanor Rosch (1973; 1975; 1978), and was also found useful in lexical semantics (see e.g., Berlin and Kay, 1969; Geeraerts, 1989). Hence, we assumed that the boundaries of prototype-based categories were not definite, and the members of a category might have different statuses: there might be more typical, “better” examples of a category. By a prototypical adjective we mean a lexeme displaying (to a certain extent) the morphological, syntactic, and semantic properties ascribed to this lexical category in the linguistic literature.

How can one then tell adjectives and participles apart? A prototype can be instantiated by the “best example” or described via a bundle of features, none of which is necessary nor sufficient to define the whole category. The present study is a test of the adjectival core features and the possibility of distinguishing more and less adjectival corpus behavior. As lexicographers need support in qualitative decision-making, we aimed to enhance the procedures used when setting boundaries. In our analysis, we combined the means of both prototypical and classic categorization, as the (gradual) deviation continuum we developed entailed basically binary and privative decisions.

The linguistic properties describing adjectives as a class will be discussed in the next section based on the example of Estonian adjectives. Since the adjective profile will be contrasted with the corresponding patterns of participles, we will also give an overview of Estonian participles.

2.2 The properties of adjectives and participles in Estonian

2.2.1 *The Estonian adjective*

There are no universal criteria for defining adjectives as a word class: adjectives may exhibit properties resembling nouns or verbs or neither of these two major categories. However, a distinguishable adjective class exists in every human language (Dixon, 2004, p. 1). Adjectives do not take major syntactic positions in sentences but occur in an attribu-

tive or predicative relation to the subject or object, modifying the noun. Semantically, adjectives describe nouns and portray their character.

In Estonian, a prototypical adjective is definable by a three-level bundle of features: morphological, syntactic, and semantic. The morphological processes characteristic of adjectives³ involve inflection, forms of comparison, and derivation.⁴ Like other word classes classified as nominals (adjectives, nouns, numerals and pronouns), Estonian adjectives are inflected for case⁵ and number. The adjectival category of comparison involves the comparative suffix *-m* and the superlative suffix *-im*. There are no morphophonological restrictions on forming comparative forms. It is also possible to use the analytic superlative construction *kõige* “most” + comparative form, and some adjectives are only used in this construction (Viitso, 2001, p. 32–35, 42).

The adjective may constitute an adjective phrase by itself, or it may occur together with its modifiers as an attribute (1a), predicative (1b) or predicative adverbial (1c). The adjective is most clearly recognizable when used attributively, which is seen as the primary function of the adjective (Erelt 2017, p. 406). An Estonian adjective used as an attribute is typically prenominal and agrees with its head noun in case and number (as in (1a)), except for the terminative, essive, abessive, and comitative cases, which require the genitive of the adjective attribute (Pajusalu 2017, p. 382; Viitso 2001, p. 35), for instance *rõõmsate lasteta* [glad-GEN child-ABE] “without glad children”. The adjectival predicative modifies the subject most often by using the copula verb *olema* “be”. It usually appears in the nominative case (as in (1b)), but also other grammatical cases and elative occur. The predicative adverbial typically expresses a result state and occurs in the translative, es-

3 It should be mentioned that to simplify the morphological corpus analysis of Estonian, certain forms are treated differently in automatic morphoanalysis than in the traditional grammars: the comparative and superlative forms are analyzed as separate lemmas, and the present participles as adjectives (see e.g. Habicht et al., 2000).

4 Another characteristic of adjectives is the adjectival derivative suffixes (the most frequent are *-ne*, *-line*, *-lik*, *-kas*, *-jas*, *-tu*, *-us*; see Kasik, 2015, p. 348–367, and about adjectival derivation see Vare, 1984).

5 Estonian has 14 nominal cases: three grammatical (nominative, genitive and partitive) and 11 semantic or adverbial cases: illative (ILL), inessive (INE), elative (ELA), allative (ALL), adessive (ADE), ablative (ABL), translative (TRA), terminative (TERM), essive (ESS), abessive (ABE), and comitative (COM) (e.g. Viitso, 2003, p. 32).

sive or nominative cases in connection with a range of verbs of change⁶ (Erelt, 2017b, p. 281–287; Erelt, 2017c, p. 289; Erelt, 2017a, p. 405). An adjective can also take a modifying adverb (1d).

- (1a) *Mahl* *tehti* *hea-de-st* *õun-te-st*
 juice was made good-PL-ELA apple-PL-ELA
 “The juice was made of good apples.”
- (1b) *Õun* *on* *väga* *hea.*
 apple is very good-NOM
 “The apple is very good.”
- (1c) *Jutt* *muutus* *igava-ks.*
 chat became boring-TRA
 “The chat became boring.”
- (1d) *üsna mugav*
 “pretty cozy”
- (1e) *palju mugava-m;* *mugava-im*
 much cozy-CMPR cozy-SUP
 “much cozier” “coziest”

The semantic properties of an adjective affect its ability to take comparative and superlative forms: the adjective allows for comparison if it encodes a scalar (degree) property,⁷ e.g. the adjective *mugav* “cozy” in (1e). Comparison forms are thus not used with all adjectives, even when there are no morphophonological constraints (Viht and Habicht, 2019, p. 27). The distinction between a relative (scalar) and absolute (non-scalar) property also influences the structure of the adjective phrase: scalar adjectives can be modified by adverbs of intensity (*väga soe* “very warm” cf. *?väga lingvistiline* “very linguistic”) (see Erelt, 2017a, 406–408). However, the distinction is not absolute, as the ability of non-scalar adjectives to be modified by adverbs is not impossible in particular contexts, as example (2) shows:

6 These verbs include: *kujunema* “turn”, *muutama* “change”, *minema* “go”, *saama* “get”, etc.

7 This dichotomy corresponds to the distinction between classifying adjectives and qualifying (attributive) ones: the former categorize the entity denoted by the noun as belonging to a certain type or class, while the latter describe the entity (e.g. Warren, 1984).

(2) *Lehed on südajad, alumised on peaaegu kolmnurksed*⁸

“The leaves are heart-shaped; the low ones are almost triangular.”

The combination of semantic, morphological, and syntactic properties that define the Estonian adjective may lack some of the general features characterizing adjectives. There are words labeled as adjectives in Estonian that do not fulfill the agreement condition of the “best example of an adjective”, because they are non-declinable (e.g., *kulla lapse-d* [dear child-PL] “dear children”). Moreover, other word classes may behave as adjectives in certain aspects. For instance, even though comparison is basically an adjectival property, nouns may adapt comparison forms in suitable contexts (*elu on lill* “life is (like) a flower”: *elu on lillem* “life is more (like) a flower”). Nevertheless, the most distinctive example of a category carrying several semantic and morphosyntactic properties of adjectives is the basically verbal class of participles.

2.2.2 *The Estonian participle*

Participles are non-finite verbal forms situated on the border between verbs and adjectives. This implies that the participle suffixes function partly as grammatical, partly as lexical categories (e.g. Viht and Habicht 2019, p. 37), positioned between regular verbal endings and derivative suffixes that yield new lexemes. Estonian participles are related to verbs via inflection for voice and mood. Both present and past participles also show adjectival properties by functioning as attributes or predicatives in a sentence. Common to all participles is that it is possible to regularly form comparative and superlative forms of them (Kerge, 1998; Erelt, 2003, p. 63; Kasik, 2015, p. 369). An important distinction between present and past participles can be made by the verbal and nominal poles: while the non-declinable past participles⁹ occur together with finite verb forms of the verb *olema* “be” (in compound

8 This example is taken from the corpus ENC 2019, subcorpus Web 2013.

9 On rare occasions, the past participle can inflect when used as a postposed attribute, agreeing with its head in case and number. Since this use is rather exceptional, we do not expect it to significantly influence the results. An example of a postposed participle is shown below:

inimese-l, tõrjutu-l ja allasurutu-l, on raske

person-ADE ostracize-PTCP-ADE and stifle-PTCP-ADE is difficult

“the person, ostracized and stifled, has difficulties”

tenses and negation), the present participles in Estonian show properties of nominal categories as they can be inflected for case and number. Modifiers characteristic of activity rather than the result of activity (or the possibility of those modifiers, in particular agentive, temporal and manner adverbials) may incline the interpretation towards the verbal (Erelt, 2017, p. 220). The participial endings, according to the tense and voice categories, for the verb *sööma* “eat” are presented in Table 1:

Table 1: *The Estonian participles*

	Personal	impersonal
present participles	<p><i>v</i></p> <p><i>õun-a</i> <i>söö-v</i> <i>laps</i> apple-PART eat-PTCP child “a child eating an apple”</p>	<p><i>dav/tav</i></p> <p><i>hommiku-l</i> <i>söö-dav</i> <i>õun</i> morning-ADE eat-PTCP apple “the apple being eaten in the morning”</p>
past participles	<p><i>nud</i></p> <p><i>laps on</i> <i>söö-nud</i> <i>õun-a</i> child be-3SG eat-PTCP apple-PART “the child has eaten the apple”</p>	<p><i>dud/tud</i></p> <p><i>hommiku-l</i> <i>söö-dud</i> <i>õun</i> morning-ADE eat-PTCP apple “the apple that has been eaten in the morning”</p>

The present impersonal participle form *söödav* is a good example of the decategorization patterns of participles: this form has the status of a headword in the CombiDic¹⁰ as an adjective meaning “edible, satisfying, palatable”, and even as a noun meaning “edibles”. The adjectival reading enables this participle to obtain the interpretation of a predicative, not compound tense form. Semantically, the lexeme *söödav* shows the abstraction tendency of adjectivized participles when it comes to the concept of time: a characteristic of adjectivization is that the situation or property can be generalized to “at any time, always” (Kerge, 1998, p. 78; Erelt, 2017d, p. 823). The detachment from the verbal paradigm is complete when the participle receives an independent meaning with respect to its verb base (Kasik, 2015, p. 70).

The questions a lexicographer deals with when categorizing participle forms are: How can we distinguish verbal and nominal participles according to their morphosyntactic behavior? When can we say that a

10 <https://sonaveeb.ee/search/unif/dlall/dsall/s%C3%B6%C3%B6dav/1>

participle has distinctively become an adjective? We propose that, in practice, it is a matter of scaling the relative proportion of occurrences in a text corpus in respect to one or another pole. We do not expect the differentiation to be straightforward, but rather a question of tendencies.

3 Material and methods

The analysis of adjectives and participles was based on the morpho-syntactic patterns identifiable in an annotated corpus. In the compilation of the test patterns, we aimed to capture the most salient attributive adjectival sequences, but also the most central non-attributive constructions. These patterns are presented in Section 3.1.

The patterns were tested on validated adjectives: the relative frequencies of corpus patterns of this group represented the reference point for further analysis and could be compared with the respective values of two control groups of participles. The principles behind the selection of the three test groups are discussed in Section 3.2. Section 3.3 presents details of the data extraction procedures. The method we used for assessing the distance of a participle from the prototypical adjectival behavior was deviation analysis, as explained in Section 3.4.

3.1 Catching adjectivity. The test patterns

The extraction of corpus sequences capturing the morphosyntactic behavior of the Estonian adjective is based on seven fixed patterns. The patterns are based on properties typical of the adjective, definable by two main parameters: the attributive and non-attributive adjectival functions. The test sequences must also be extractable by the corpus tagging system.

Most of the patterns reflect the properties assigned to adjectives in the linguistic literature. The third pattern is inspired by practical lexicographic work, and the fourth pattern has grown out of the analysis of corpus material. The seventh, the predicative pattern, is an addition to our previous investigation of adjectives (Tuulik et al., 2022, pp. 283–285). The term *test word* refers to any test word inserted into the search for the respective pattern. Six of the patterns are sequences; the comparative pattern counts, and thereby confirms, the existence

of the comparative form of a test word in the corpus. The test patterns used in this study are as follows:

- 1) The attribute pattern (ATTR) targets the sequence of the test word immediately preceding a noun. This pattern is based on the tendency of an adjective to modify the noun as an attribute. The collocational sequence ADJ_NOUN presumably reflects the most frequent use of adjectives (e.g. *väike laps* “little kid”).
- 2) The agreement pattern (ATTR/AGR) is a sequence of the test word in the same case and number as the following noun. It tests the agreement of the test word and head noun, based on the ability of adjectives in the attributive function to agree in case and number with their head nouns (*väikes-te-l kivi-de-l* [little-PL-ADE kid-PL-ADE] “on the small rocks”).
- 3) The sentence starter pattern (ATTR/ST) sets a syntactic restriction on the attribute phrase: the test word followed by a noun must be located at the beginning of a sentence. The purpose of this pattern is to differentiate verbal participles from adjectivized ones, for instance *Tuleval suvel...* “In the upcoming summer” is quite natural, but *Oleval suvel...* “In the being summer” is not.
- 4) The four-spot pattern (ATTR/4) measures the occurrence of the test word in a larger pattern, where it modifies a substantive and follows the sequence of an unspecified verb and an unspecified word (verb + X + test word + noun). According to our pilot study compiled to test the parameters (Tuulik et al., 2022), this pattern distinguished the main target of the present investigation – participles – from adjectives. With respect to other categories, this parameter was not as effective. This study will thus indicate whether this pattern should be kept in the test battery.
- 5) The adverb pattern (ADV) ascertains the sequence of an adverb preceding the test word. We expect the ability to take adverbial modifiers to be characteristic of the adjectives in the corpus, particularly with scalar adjectives.
- 6) The comparison pattern (COMP) estimates whether a word yields comparative forms. We restrict this test pattern to comparatives, assuming that the existence of a comparative form is a logical precondition for the possibility of a superlative. Moreover, since the

highest degree of comparison can, in parallel, be expressed by the analytic *most*-construction, involving the adverb *kõige* “most” and the comparative form of the adjective (*kõige väiksem* “most smaller”), the results are not quite representative.¹¹

- 7) The predicative pattern (PRED) targets the sequence of the test word directly after the copula verb *olema* “be” or after *olema* and an adverb. Both patterns are characteristic of the Estonian predicative;¹² of course, these are also potential forms for the compound tense constructions involving participles, but in this case there may be additional sentential elements between the copula and participle. We do not specify the morphological form of the test word here as the Estonian predicative can be marked by several cases (the three grammatical cases and the relative case – see Section 2.2).

As the patterns described above indicate, there are two recurrent structural relations unifying the variables: patterns 1–4 include attributive phrases in a more or less fixed position in the sentence, and patterns 4 and 7 involve the pre-adverbial relationship of the test word (adverb + test word). Of the four attributive patterns, 1 and 2 can be classified as **general attributive patterns**, and 3 and 4 as **complex attributive patterns** with fixed positions in the sentence.

A summary of the division of test patterns according to attributive and non-attributive parameters is given in Table 2; the abbreviation TW stands for test word and the patterns are given in their logical form. Note that the elements in patterns containing TW are consecutive sequences.

11 There is an analytic way in Estonian to express even the comparative relation, using the (also comparative form) word *rohkem* (*rohkem väike* “more little”), but this is not as productive a pattern as the *kõige*-superlative and, although understandable, it is often not quite idiomatic.

12 Since the predicative can occur in four cases (nominative, partitive, genitive and relative), we do not restrict the morphological form of the test word. The predicate of this pattern is limited with *olema* “be” because we need to narrow down the corpus search pattern and focus only on the central case of predicative in Estonian, leaving out the cases of the predicative adverbial that enables the predicative to take the translative, essive and nominative cases in connection with verbs expressing change of state: *saama* “become”, *muutama* “turn”, *kujunema* “develop into, turn”, *minema* “go”, etc.

Table 2: Test patterns and parameters

parameter	pattern name	abbreviation	search pattern
attributive	general attributive	attribute pattern	ATTR (TW ^ noun)
		agreement pattern	ATTR/AGR (TW ^ noun) _{agreement}
	complex attributive	sentence starter pattern	ATTR/ST (TW ^ noun) _{sentence start}
		four-spot pattern	ATTR/4 verb ^ x ^ (TW ^ noun)
non-attributive		adverb pattern	ADV (adverb ^ TW)
		comparison pattern	COMP ∃ comparative
		predicative pattern	PRED ∀ olema “be” ^ TW ∀ olema “be” ^ (adverb ^ TW)

3.2 The sample: validated adjectives and control groups of participles

The data set used in this study contained three sample groups. The reference group of our study, which we also call *the validated adjectives group*, consisted of a selection of lexicographically verified adjectives: 100 words extracted by random sampling from the 554 adjectives included in the *Basic Estonian Dictionary*.¹³ We expected the adjectives included in this dictionary to be the most central and prototypical. To ensure the coherence of the sample, we excluded the lexemes that showed ambivalent behavior regarding their word class affiliation (e.g. *vabatahtlik*, interpretable both as the adjective “voluntary” and the substantive “volunteer”). We also excluded adjectives missing some central adjectival features, such as the non-declinable *eri* “separate, various”.

We used two groups of participles as control groups of less prototypical cases to compare with the reference group. Control group 1 contained participles that by expectation incline towards adjectives, and control group 2 consisted of participles used predominantly in verbal contexts. All the participles were selected by random sampling and checked for their suitability. Both samples included all four participle

¹³ The dictionary includes the 5,000 most frequent Estonian words explained in simple language. In addition to frequency criterion, the words were selected according to their prominence for the learner of Estonian (Kallas et al., 2014: 1109–1110).

types (cf. Table 1 in Section 2.2) – personal present, personal past, impersonal present, and impersonal past participles – with an equal number of each participle type.

Control group 1, *the adjectivizing participles in Ekilex*, consisted of 100 participles that were CombiDic candidates or under-specified headwords of CombiDic. We expected most of these forms to behave as adjectives in the corpus texts and potentially to be tagged accordingly in the database. The random sample of Ekilex participles was extracted from the Ekilex database (N = 1,543).¹⁴

Control group 2, *the regular participles*, contained 100 participles for which we expected as little adjective behavior as possible. The verbs functioning as bases for the participles in this group were selected by random sampling from the approximately 1,000 verbs included in the *Basic Estonian Dictionary*. The four types of participles were then formed and manually checked for their verbal use and sufficient frequency in the corpus.

The composition of the test groups was planned with the expectation that the morphosyntactic test patterns described in Section 3.1 would be able to distinguish the groups from each other. In other words, we hypothesize that the adjectivizing participles group of the Ekilex entries resembles the reference group of validated adjectives to a larger extent than the regular participles in the morphosyntactic corpus patterns we focus on.

3.3 The corpus extraction process

We extracted the data from the Estonian National Corpus 2019¹⁵ (ENC 2019; see also Koppel and Kallas, 2022). The 1.5 billion token corpus ENC 2019 was pre-tagged, lemmatized, and disambiguated with the EstNLTKv.1.6 program, a natural language toolkit explicitly developed

14 The PoS-tagging status of CombiDic headwords changes along with the updating of the dictionary. This was extracted 24. 1. 2022.

15 The ENC-corpora are stored in the corpus query system Sketch Engine (Kilgarriff et al., 2004; Kilgarriff et al., 2014). We use the files of the ENC2019 uploaded from Sketch Engine to the home page of the Center of Estonian Language Resources. The frequency results of the Sketch Engine and CELR files may differ by up to 1%, as the last uses a slightly different approach by rejecting the data from broken sentences (Neeme Kahusk, personal communication). The ENC2019 subcorpora are available at <https://entu.keeleressursid.ee/shared/7769/N66ZdfvwzQuXWIVjnhVuX74oWmi1zrruZ1VpN8QE1Hj6jbfq5oMBxm8YQDrugyM>

for the Estonian language and written in the Python programming language, executing basic NLP tasks (Orasmaa et al., 2016, p. 2460, Laur et al., 2020). In the tagging process, EstNLTK uses the tag set of the Vabamorf morphoanalyzer, which combines rule-based and probabilistic models, and its lemma disambiguation system based on the Vabamorf lexicon. According to Kaalep et al. (2012), EstNLTK’s lemma disambiguation precision is around 0.94.

We applied a code¹⁶ written in Python programming language for automatic data extraction. Table 3 presents the logical expressions used in data extraction.

Table 3: *The logical expressions used in test pattern extraction*

pattern name	search pattern	logical expression in the programming language of Python
ATTR attribute	TW + noun	<code>i < sent_len - 1 and lemmas[i].lower() in test_words and postags[i+1] == "S"</code>
ATTR/AGR agreement	<code>[TW + noun]_{agr}</code>	<code>i < sent_len - 1 and lemmas[i].lower() in test_words and postags[i+1] == "S" and forms[i] == forms[i+1]</code>
ATTR/ST sentence starter	<code>[TW + noun]_{st}</code>	<code>i < sent_len - 1 and i == 0 and lemmas[i].lower() in test_words and postags[i+1] == "S"</code>
ATTR/4 four-spot	verb + X + TW + noun	<code>i < sent_len - 3 and postags[i] == "V" and lemmas[i+2] in test_words and postags[i+3] == "S"</code>
ADV adverb	adverb + TW	<code>i < sent_len - 1 and postags[i] == "D" and lemmas[i+1].lower() in test_words</code>
COMP Comparison	\exists comparative	<code>lemmas[i].lower() in comp_words</code>
PRED predicative	<i>olema</i> "be" + TW <i>olema</i> "be" + (adverb) + TW	<code>i < sent_len - 2 and lemmas[i].lower() == "olema" and postags[i+1] and lemmas[i+2].lower() in test_words</code>

The frequency detection of test patterns was restricted by the limits of sentences, and those test pattern occurrences crossing the sentence boundaries were not considered. The test pattern identification counted lemmas of the test words (`lemmas[i].lower() in test_words`) with the exception of test words with the endings "dud", "nud" and

16 https://github.com/ahtilohk/PSG227/blob/main/Test-patterns_occurrences_in_ENC2019_without_estnltk_corpus_processing_module.py

“tud”. These are the cases of the non-inflected past participles and for those only text words were considered.¹⁷ In the extraction of the comparison pattern, we used a general code that searched for the occurrences of comparative forms on the basis of a manually composed list.¹⁸ The test words were untagged throughout the extraction process, i.e., their tagging status was unspecified.

It should be borne in mind that the frequency results directly depend on the quality of the tagging system used, which in this study was based on the Vabamorf morphological analyzer, as incorporated in the EstNLTK program. We are aware of the possibility that tagging and disambiguation errors (e.g. ambiguities caused by inflectional homonymy¹⁹) may have affected our analysis. We did not manually correct the shortcomings of the automatic analysis because a lexicographer would receive a statistical analysis based on the very same corpus processing methods when using a potential application based on this model.

Since the absolute frequencies of test words adopting the corpus patterns were not comparable, we operated with relative frequencies: the absolute frequencies matching the test pattern requirements were divided by the general lemma frequencies.

3.4 Deviation analysis as a similarity measure

To identify the dissimilarity between the morphosyntactic behavior of the prototypical adjectives and the two control groups of participles, we employed a method that we call deviation analysis. It can be used for the systematic comparison of the measurements of a target phenomenon with the respective measurements of a standard. There is no predefined formula in this method, and the relevant parameters are measured and compared one-by-one. Based on the measurements, a

17 Instead of `lemmas[i].lower()` in `test_words` in the logical expression `len(word) > 3` and `word[-3:]` in `[“nud”, “dud”, “tud”]` and `word` in `test_words` are used.

18 To implement the comparison test pattern, we manually formed comparative forms of all of the words in our samples, following the comparative formation rule: the singular genitive form of the test word + *m* (e.g., *väike* “small” : *väikse* [small.GEN] : *väiksem* “smaller”).

19 As a morphologically rich language, form homonymy is not rare in Estonian and the forms of different lemmas may coincide, e.g. the form *koon* can be analyzed as the nominative case form of the noun *koon* “snout”, and the third person present indicative form of the verb *kuduma* “drink” (“I knit”).

range of tolerance can be specified to decide the acceptability of the rates of the target phenomenon as compared to the standard.

In this study, we took the relative frequencies of the corpus patterns as relevant measurements and defined a range of tolerance for every pattern based on the respective values of the reference group of the validated adjectives.²⁰ The results for the control group words could then be subjected to deviation analysis. Additionally, the counts of deviating criteria per word allowed us to establish a scale of dissimilarity to the corpus behavior of adjectives. By specifying the ability of each pattern to exclude regular participles from the adjectival tolerance ranges, we evaluated the differentiation efficiency of each pattern.

4 Deviation analysis of the sample words

In this section, we present the results of the corpus extraction data based on the seven test patterns and the three sample groups. The relativized frequency statistics of the 100 validated adjectives are provided in Subsection 4.1, and these data are the basis for defining the tolerance ranges of adjectival behavior (4.2). Next, we relate the results of the control groups to the tolerance ranges, which enables us to define the deviation ranges and establish the degree of deviation of the control groups in relation to the reference group (4.3), and to assess the efficiency of the test patterns (4.4).

4.1 The test results for the validated adjectives

The variation of the relative frequency results based on the 100 words of the validated adjectives group according to the outcomes in the seven test patterns is presented in a box plot in Figure 1, while the descriptive statistics behind the variation are presented in Table 4, below.

20 For an approach that treats the values of the respective test patterns as a joint measure of overall similarity vs difference, see Vainik et al. (in press).

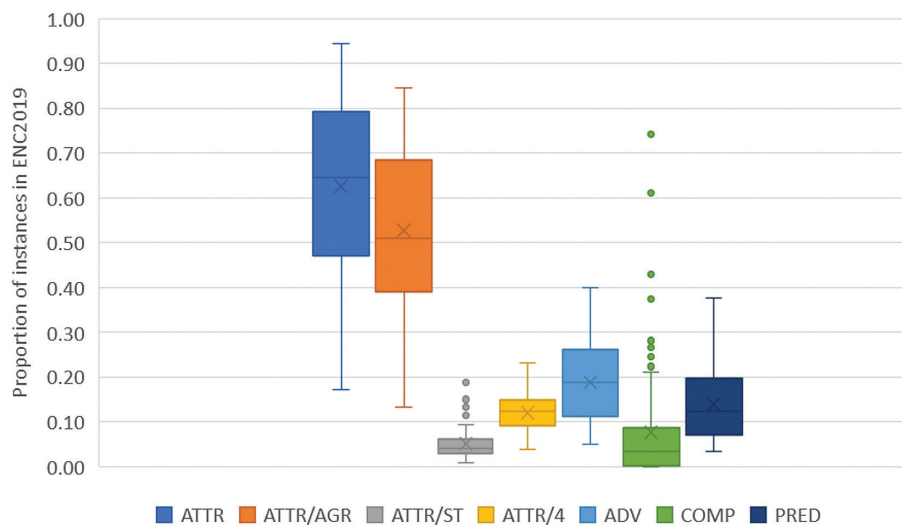


Figure 1: The division of the data of the test group of validated adjectives by the test patterns.

Table 4: Descriptive statistics of the test group of validated adjectives by the test patterns

	ATTR attribute	ATTR/AGR agreement	ATTR/ST sentence starter	ATTR/4 four-spot	ADV adverb	COMP comparison	PRED predicative
min	0.173	0.132	0.009	0.038	0.049	0.000	0.034
max	0.944	0.846	0.193	0.230	0.399	0.742	0.376
ave	0.626	0.527	0.051	0.120	0.188	0.076	0.138
stDev	0.192	0.185	0.034	0.037	0.084	0.120	0.074
median	0.644	0.510	0.042	0.123	0.187	0.034	0.124

Figure 1 and Table 4 show the average and median rates and the ranges of variation across the seven test patterns. The variation spans are notably wide, particularly in the two general attributive patterns: the attribute and the attribute agreement sequences.²¹ The highest average and median values belong to the general attributive patterns, which is in accord with the assumption of the attributive function’s prevailing status for adjectives (cf. Section 2.2; Erelt, 2017, p. 406). The high scores of the general attribute patterns (ATTR and ATTR/AGR) in-

²¹ For more about the correlation relations between the test patterns, see Vainik et al., (submitted).

dicate that the adjectives can quite freely function attributively in different positions of a sentence. The frequency of the complex attributive patterns (ATTR/ST and ATTR/4) is considerably more restricted, as these patterns combine multiple conditions besides the sequence of the test word and a noun (cf. Table 3 in Section 3.1).

The patterns with the most discrepant results are the four-spot-pattern and the comparison pattern, revealing several outliers. The non-attributive patterns – the adverb pattern (ADV), the comparison pattern (COMP), and the predicative pattern (PRED) – overall demonstrate relatively low levels of relative frequencies and variation ranges. The average rate of comparative forms is strikingly low, which is unexpected given the assumed prototypical nature of the validated adjectives. There are seven distinctly non-scalar adjectives without any occurrences in the comparison pattern, for instance *kahekordne* “double, two-floored” and *eelmine* “previous”. The outliers deviating from the general tendency, i.e. the adjectives with exceptionally high results in the comparison pattern, are *kõrge* “high”, *lihtne* “simple”, *täpne* “precise” and *lahja* “lean”.

4.2 Setting the adjectival limit ranges

The marginal rates of relative frequencies in the validated adjective group lay the foundation for postulating ranges of tolerance for the test patterns. The maximum and minimum values of the patterns (see Table 4), except for the comparison pattern, serve as the highest and lowest values of the corresponding tolerance ranges. The evaluation of the comparison pattern differs from other patterns: here we estimate the absolute frequency of a word’s comparative form in the corpora. We consider an absolute frequency of higher than five occurrences to be a sign of non-occasional comparison formation, and hence not deviating from the adjective range.

To sharpen the contrast of adjectives from regular participles, we qualitatively adjusted the limits of the attribute pattern and the sentence starter pattern. In this process, we excluded a few highly deviating adjectives from the ranges of these patterns to better capture the essence of prototypical adjectives. The exclusion was done by comparing the test groups and considering a pattern’s ability to differentiate

validated adjectives from regular participles individually. For example, setting the minimum value for the attribute pattern (ATTR) from 0.173 to 0.246 (by excluding the result of one validated adjective²²), allowed us to differentiate 13 words from the regular participles group that would fit in the tolerance range if this one distinct adjective were not excluded. The ranges of tolerance are presented in Table 5.

Table 5: *The ranges of tolerance. Limits of the prototypical adjective for the seven test patterns*

test patterns	adjectival ranges (relative frequencies)
ATTR – attribute	0.246–0.920
ATTR/AGR – agreement	0.132–0.846
ATTR/ST – sentence starter	0.015–0.193
ATTR/4 – four-spot	0.038–0.230
ADV – adverb	0.049–0.399
COMP – comparison	*5 <
PRED – predicative	0.034–0.376

* The range of comparative forms is calculated in absolute frequencies

The analysis below, addressing the degrees of deviation from the adjective behavior and the differentiation efficiency of the test patterns, is based on the ranges defined in Table 5.

4.3 Assessing the control groups. Defining the deviation scales

Using the ranges of tolerance established in the previous section, we analyzed whether the test words fit into the limits set by the validated adjectives. To do that we counted both inclusive and non-inclusive results regarding the tolerance ranges for each pattern. To illustrate the analysis, we present the results for six control group participles across all seven test patterns in Table 6.

²² This word is *kade* “envious”, which occurs extraordinarily rarely in the attributive function compared to other validated adjectives.

Table 6: Deviation analysis of six test words from the control groups

group**	test word	ATTR	ATTR/ AGR	ATTR/ ST	ATTR/ 4	ADV	COMP	PRED	total
Ekilex	<i>kleepuv</i> “sticking”	+	+	+	+	+	+	+	0
regular	<i>korratav</i> “being repeated, repeatable”	+	+	-	+	+	+	+	1
Ekilex	<i>maksustatud</i> “taxed”	+	-	-	+	+	-	+	3
regular	<i>kaevav</i> “digging”	+	+	-	+	+	-	-	3
Ekilex	<i>soostunud</i> “consented; swampy”	+	-	-	-	-	-	+	5
regular	<i>naerdud</i> “laughed”	-	-	-	-	-	-	+	6

** Ekilex = control group 1, the adjectivizing participles in Ekilex;
regular = control group 2, the regular participles

Let us now consider how much all of the words of the three test groups deviate from the tolerance ranges. Table 7 presents the test words measured according to the number of deviating patterns. As the table shows, 89 validated adjectives result in zero deviation and 10 adjectives deviate by one pattern, while one adjective deviates by two patterns. Based on the results for all three test groups, we define three degrees of deviation from the adjectival behavior: the no deviation (0–1 patterns deviating from the tolerance ranges), low deviation (2–3), and high deviation (4–7) scale.

Table 7: The deviation scale

no. of patterns deviating	0	1	2	3	4	5	6	7
adjectives (%)	89	10	1					
Ekilex participles (%)	6	27	39	16	9	3		
regular participles (%)	1	16	14	20	16	21	11	1
degrees of deviation	no deviation		low		high			

Although 99% of the validated adjectives fall within the highest prototypicality degree of no deviation, there are words that do not

score in all patterns even in this group, which is explainable by the adjustments of adjectival ranges described in Section 4.1. The test words with one deviating pattern are adjectives that do not form comparatives due to semantic restrictions (five adjectives, e.g. *ühetoaline* “one-roomed” and *vasak* “left”) or reach the lower limit of the tolerance range in the attribute pattern (two adjectives, *kade* “envious” and *sõjaline* “military”) or in the sentence starter pattern (five adjectives, e.g. *selge* “clear” and *vajalik* “necessary”). The validated adjective deviating in two test patterns is *kade* “envious”, an adjective favoring non-attributive usage.

Comparing the results for all three test groups proves our hypothesis: the adjectivizing participles in Ekilex correspond to the adjectival behavior to a larger extent than the regular participles group. The deviation analysis shows that 33% of the Ekilex participles and only 17% of the regular participles match the no-deviation space with no or one pattern deviating from the tolerance ranges. Altogether 88% of the adjectivizing participles in Ekilex and 51% of the regular participles fall within the low or no deviation space. Only 12% of the Ekilex participles but 49% of the regular participles are situated at the high deviation level. Note also that all of the test words from the Ekilex group show at least two patterns within the ranges of tolerance.

4.4 Estimating the efficiency of test patterns

In this section, we evaluate the efficiency of the seven test patterns, i.e. the ability of each pattern to exclude regular participles from the tolerance ranges defined on the basis of the variation scope of the validated adjectives (see Table 5 in Section 4.1). The efficiency is assessed by the extent of the difference between the control group and reference ranges. Basically, the bigger the gap between the results of the two groups, the better the corresponding pattern’s efficiency.

First, we calculate the differentiation efficiency of the patterns by comparing the results for the validated adjectives and regular participles group in terms of how many test words fit into the reference ranges of corresponding patterns. The results are presented in Table 8, with the patterns ordered from stronger to weaker efficiency, from left

to right. The values of the adjectives²³ represent 100%, and the corresponding ratio the result of the regular group – the gap between these two is the difference (the results for adjectives minus participles). For the collation of the data, the table also includes the results for the adjectivizing participles in the Ekilex group that fall, as we hypothesized, in between the validated adjectives and regular participles group.

Table 8: *The differentiating efficiency of test patterns; regular participles versus adjectives*

	COMP comparison	ATTR/ST sentence starter	ATTR/AGR agreement	ATTR attribute	ATTR/4 four-spot	PRED predicative	ADV adverb
adjectives	94	95	100	99	100	100	100
Ekilex participles	26	74	51	93	91	72	89
regular participles	14	25	49	62	63	66	75
difference	85%	74%	51%	37%	37%	34%	25%

According to the data presented in Table 8, the most efficient differentiator (the pattern that leaves the most regular participles out of the tolerance range) is the comparison pattern, also excluding a significant number of adjectivizing participles in Ekilex. The second strongest differentiator is the sentence starting pattern, which places 74% of Ekilex participles together with validated adjectives and leaves 74% of the words of the regular participles group outside of the tolerance range. The results for the attribute pattern and the four-spot pattern overlap, suggesting that the test battery would not suffer if one of them were excluded (at least for the analysis of participles). Overall, the results indicate that in each test pattern the adjectivizing participles in Ekilex fall between other test groups, exhibiting lower adjectival scores than the validated adjectives, but significantly higher scores than the regular participles.

The jitter plots below illustrate the distribution of the results of a strong differentiator (the sentence starter pattern, Figure 2) and a weak differentiator (the predicative pattern, Figure 3). The values – relative pattern frequencies for each test word – of the three test groups are

23 Note that the reason for the number of adjectives in three test patterns falling under 100 in Table 8 is that the adjectival ranges have been slightly adjusted (cf. Section 4.1 and Table 5).

presented on the x-axis; for the y-axis, the plots show randomly generated values, ensuring that the dots do not overlap.

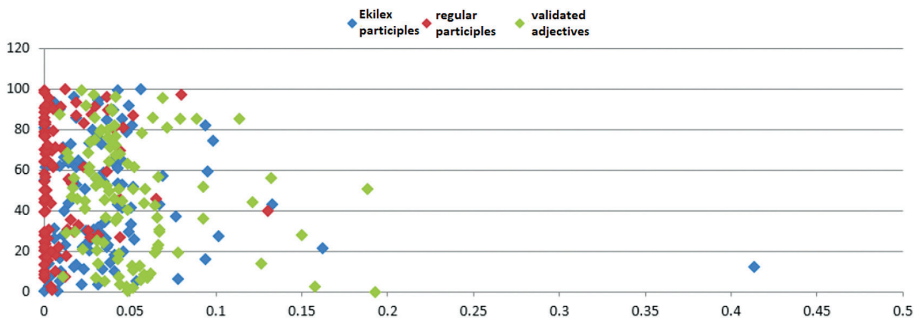


Figure 2: Distribution of results within the sentence starter pattern.

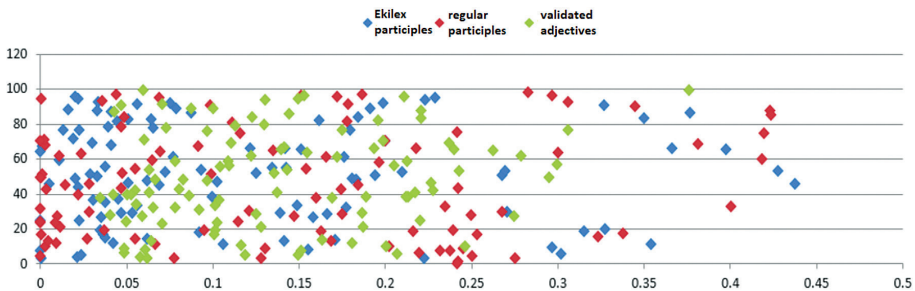


Figure 3: Distribution of results within the predicative pattern.

As the distribution of the results demonstrates, the two patterns differ considerably in their ability to differentiate the test groups. The distribution of the results based on the sentence starter pattern shows the results for the regular participles cumulating near a value of 0, while those for the adjectivizing participles in Ekilex and validated adjectives are split between 0 and 0.2. One of the weakest differentiators, the predicative pattern, spreads the results for the three test groups more evenly over a wider range, from 0 to 0.44.

5 Conclusions

The assessment of the limits of prototypical adjectivity carried out in this study confirmed that it is possible to capture the adjectival corpus behavior by morphosyntactic sequences typical of adjectives. We ap-

proached adjectivity via the most salient morphosyntactic properties of adjectives generalizable by the attributive vs. non-attributive opposition. Operationalizing these main parameters into seven sequential corpus patterns helped us to establish the ranges of variation within the defined limits of tolerance. We can conclude that the test patterns clearly distinguished the group of validated adjectives from the two control groups of participles.

The analysis showed that the validated adjectives in the test group were not homogeneous either: their results spread over three different ranges in terms of the results in the deviation analysis (zero, one or two patterns outside the ranges of the prototypical adjective), showing variance to a certain degree and proving the prototype-based nature of the adjective class. The deviation analysis resulted in a tripartite scale of similarity to adjectives in terms of deviation from the tolerance ranges set according to the variation of the group of adjectives. The overall scale of adjectivity was achieved by calculating the ratio of deviating and coinciding criteria (see the scale of deviation in Table 7). According to the deviation analysis, 12% of the adjectivizing participles in Ekilex and 49% of the regular participles were assessed as highly deviating from the validated adjectives, a result proving that the participles of these two groups differ in degrees of adjectivization. Moreover, and as we hypothesized, the adjectivizing participles in Ekilex (adjective candidates) fell closer to the validated adjectives than the regular participles.

The most accurate differentiation of the regular participles group from the validated adjectives was achieved by the comparison and sentence starter patterns. The results for the validated adjectives indicate that the occurrence of comparative forms is not necessarily frequent even for presumably prototypical adjectives, and thus may leave out words perfectly eligible for adding as dictionary entries. The infrequency of validated adjectives is striking, even in the sentence starter pattern. This indicates that the general adjectival properties (e.g. the simple attribute pattern or the predicative pattern) are not necessarily the clearest distinguishers between adjectival and non-adjectival behavior. The adverb pattern was the weakest differentiator according to the comparison of regular participles and validated adjectives. This result diverges from our pilot study with smaller test groups, in which

the adverb pattern was one of the three best differentiators (Tuulik et al., 2022, p. 296). We can conclude that the best differentiators are not necessarily the most typical adjectival properties (attributive and predicative), but more specific markers of potential morphosyntactic behavior. There are patterns that strongly exclude nonadjectival words (good excluders) and patterns that strongly include (good includers) adjectival words with the notion of prototypical adjectives.

The efficiency analysis revealed that the selection and constitution of patterns used in this study could be elaborated further to optimize the results. The comparison pattern extraction process would be facilitated by developing an automatic generator of comparative forms. Since the two attributive patterns, the attribute and four-spot patterns, show quite similar differentiating results, one of them can be left out of the test battery without weakening the results. The attribute pattern may be preferable as a necessary sequence since it shows slightly better results and is structurally simpler to use in the extraction process. As the results of the attribute agreement pattern were more or less the same in the control groups, due to the identical selection of the declinable present and non-declinable past participles in both, we can conclude that the agreement pattern could be more useful in connection with some other categories, e.g. in the assessment of the adjectival behavior of nouns.

The predicative and adverb patterns also need further adjustment: in their current forms they do not clearly differentiate regular participles from adjectivized ones, or from validated adjectives. One solution would be to include in the extraction code of the predicative pattern certain morphological restrictions by defining the predicative case forms. Moreover, setting the presence of the negation word *ei* “not” in the near context of the test word could also help to highlight verbal uses of participles. The adverb pattern may be elaborated by adding an inclusive search list of intensifying adverbs to the corpus extraction algorithm in order to avoid typical verb modifiers, e.g. adverbs of manner.

Ultimately, it is important to acknowledge the effects such patterns have in concurrence. But how many patterns are necessary to achieve the optimal results? In light of this study, we suggest that a proper test battery should include at least five patterns to capture the morphosyn-

tactic behavior of the versatile class of adjectives. The composition of patterns may be adjusted according to the lexical group targeted for assessment. It is also possible to use different methodological solutions and analyze the results for the test patterns in concurrence instead of a sum of separate values (for a Euclidean distance approach, see Vainik et al., in press).

The use of a quantitative approach can reveal unexpected aspects of a language, and the findings of this study have the potential to contribute to the knowledge of adjectives in Estonian, and also indicate the value of further investigations into this topic. When it comes to the contrasting focus of this study, the Estonian participles, the analysis revealed some similarities with adjectives as exemplified by the two control groups of participles. Another finding is connected to the subtypes of participles: the results of the deviation analysis show that the present participles congregate among the higher scores (in particular passive present participles, such as *hinnatav* “assessable”) and the past participles fall within the lower scores. At least in part this is due to the fact that past participles cannot perform in the attribute agreement pattern, but there may be other factors affecting the similarity to the adjectives. Overall, the reasons for the general tendencies as well as outliers in the data deserve a closer, qualitative analysis.

In our opinion, the results of this study can be applied to develop a multi-parameter application for determining the relative adjectivity of a word or a word form, e.g. adjectivizing participles or nominals (for the border areas of adjectives with other lexical classes in Estonian, see Vainik, Paulsen and Lohk 2020). As the morphosyntactic patterns characteristic of a PoS are language-specific, so is the outcome of our examination. The results are, however, also adjustable for the analysis of other languages.

Acknowledgments

This research was supported by Estonian Research Council grant PSG227.

Abbreviations

ABE – abessive case	CMPR – comparative form
ADE – adessive case	ELA – elative case
ADV – adverb pattern	GEN – genitive case
ATTR – attribute pattern	NOM – nominative case
ATTR/AGR – attribute agreement pattern	PART – partitive case
ATTR/ST – sentence starter pattern	PL – plural
ATTR/4 – four-spot pattern	PRED – predicative pattern
COMP – comparison pattern	PTCP – participle
	SUP – superlative
	TRA – translative case

References

- Berlin, B., & Kay, P. (1969). *Basic Color Terms: Their Universality and Evolution*. Berkeley: University of California Press.
- CombiDic = The EKI Combined Dictionary. (2020). Hein, I., Kallas, J., Kiisla, O., Koppel, K., Langemets, M., Leemets T., ..., & Voll, P. Institute of the Estonian Language. Retrieved from <https://sonaveeb.ee> (15. 11. 2022)
- Dixon, R. M. W. (2004). Adjective classes in typological perspective. In R. M. W. Dixon & A. Aikhenvald (Eds.), *Adjective classes: a cross-linguistic typology*. Oxford: Oxford University Press.
- Ekilex. Retrieved from <https://ekilex.eki.ee/> (26. 11. 2021)
- Erelt, M. (2017a). Omadussõnafraas [The adjective phrase]. In M. Erelt & H. Metslang (Eds.), *Eesti keele süntaks* [The Syntax of Estonian] (pp. 405–415). Eesti keele varamu III. Tartu Ülikooli Kirjastus.
- Erelt, M. (2017b). Öeldistäide [The predicative]. In M. Erelt & H. Metslang (Eds.), *Eesti keele süntaks* [The Syntax of Estonian] (pp. 278–288). Eesti keele varamu III. Tartu Ülikooli Kirjastus.
- Erelt, M. (2017c). Öeldistäitemäärus [The predicative adverbial]. In M. Erelt & H. Metslang (Eds.), *Eesti keele süntaks* [The Syntax of Estonian] (pp. 289–299). Eesti keele varamu III. Tartu Ülikooli Kirjastus.
- Erelt, M. (2017d). Sekundaartarindiga laused [Sentences with secondary constructions]. In M. Erelt & H. Metslang (Eds.), *Eesti keele süntaks* [The Syntax of Estonian] (pp. 756–840). Eesti keele varamu III. Tartu Ülikooli Kirjastus.

- Erelt, M. (2017e). Sissejuhatus süntaksisse [Introduction to syntax]. In M. Erelt & H. Metslang (Eds.), *Eesti keele süntaks* [The Syntax of Estonian] (pp. 53–89). Eesti keele varamu III. Tartu: Tartu Ülikooli Kirjastus.
- The Estonian Collocations Dictionary = Eesti keele naabersõnad. (2019). Kallas, J., Koppel, K., Paulsen, G. & Tuulik, M., Institute of the Estonian Language. Retrieved from <http://www.sonaveeb.ee> (14. 11. 2022)
- The Estonian-Russian Dictionary (2019). Laasi, H., Lagle, T., Leemets, H., Liiv, M., Pärn, H., Simm, L., ..., Tubin, V., (Comp.); Liiv, M., Melts, N., Romet, A., Kallas, J., Riikoja, E., Martoja, I., Smirnov, S., ..., & Veskimägi, E. (Eds.). doi: [10.15155/3-00-0000-0000-0000-0001BL](https://doi.org/10.15155/3-00-0000-0000-0000-0001BL)
- Geeraerts, D. (1989). Prospects and problems of prototype theory. *Linguistics*, 27, 587–612.
- Habicht, K., Kaalep, H.-J., Muischnek, K., Müürisep, K., & Rääbis, A. (2000). Kas tegelik tekst allub eesti keele morfoloogilistele kirjeldustele? Eesti kirjakeele testkorpuse morfosüntaktilise märgendamise kogemusest. *Keel ja Kirjandus*, 9, 623–633.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann Publishers.
- Kaalep, H.-J., Kirt, R., & Muischnek, K. (2012). A trivial method for choosing the right lemma. In *Baltic HLT* (pp. 82–89).
- Kallas, J., Tiits, M., Tuulik, M., Koppel, K., & Jürviste, M. (2014). Eesti keele põhisõnavara sõnastik [The Basic Estonian Dictionary]. Tallinn: Eesti Keele Sihtasutus.
- Kallas, J., & Tuulik, M. (2011). Eesti keele põhisõnavara sõnastik: ajalooline kontekst ja koostamispehmoõtted. *Estonian Papers in Applied Linguistics*, 7, 59–75.
- Kallas, J., Tuulik, M., & Langemets, M. (2014). The Basic Estonian Dictionary: the first Monolingual L2 learner’s Dictionary of Estonian. In A. Abel, C. Vettori & N. Ralli (Eds.), *Proceedings of the XVI EURALEX International Congress: The User in Focus, 15–19 July 2014, Bolzano, Bozen* (pp. 1109–1119). Bolzano/Bozen: Institute for Specialised Communication and Multilingualism. Retrieved from https://www.euralex.org/elx_proceedings/Euralex2014/euralex_2014_086_p_1109.pdf
- Kasik, R. (2015). *Sõnamoodustus*. Tartu: Tartu Ülikooli Kirjastus.
- Kerge, K. (1998). Vormimoodustus, sõnamoodustus ja leksikon: oleviku keskõna võrdluse all. Tallinn: TPÜ Kirjastus.
- Kilgarriif, A., Rychlý, P., Smrz, P., & Tugwell, D. (2004). The Sketch Engine. In G. Williams & S. Vessier (Eds.), *Proceedings of the Eleventh EURALEX*

- International Congress, 6–10 July 2004, Lorient, France* (pp. 105–116). Lorient: Université de Bretagne Sud. Retrieved from <https://euralex.org/category/publications/euralex-2004/>
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, 1, 7–36.
- Koppel, K., Tavast, A., Langemets, M., & Kallas, J. (2019). Aggregating dictionaries into the language portal Sõnaveeb: issues with and without a solution. In I. Kosem, T. Zingano Kuhn., M. Correia, J. P. Ferreria, M. Jansen, I. Pereira, J. Kallas, M. Jakubíček, S. Krek & C. Tiberius (Eds.). *Proceedings of the eLex 2019 conference: Smart lexicography, 1–3 October 2019, Sintra, Portugal* (pp. 434–452). Brno: Lexical Computing CZ, s.r.o. Retrieved from https://elex.link/elex2019/wp-content/uploads/2019/09/eLex_2019_24.pdf
- Laur, S., Orasmaa, S., Särg, D., & Tammo, P. (2020). EstNLTK 1.6: Remastered Estonian NLP Pipeline. In *Proceedings of The 12th Language Resources and Evaluation Conference, May 2020, Marseille, France* (pp. 7152–7160). European Language Resources Association (ELRA). Retrieved from <https://aclanthology.org/2020.lrec-1.0.pdf>
- Orasmaa, S., Petmanson, T., Tkatšenko, A., Laur, S., & Kaalep, H-J. (2016). EstNLTK – NLP Toolkit for Estonian. In N. Calzolari, K. Choukri, T. Declerck, M. Grobelnik, B. Maegaard, J. Mariani, A. Moreno, J. Odijk & P. Stelios (Eds.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016), Portorož, Slovenia* (pp. 2460–2466). European Language Resources Association (ELRA). Retrieved from http://www.lrec-conf.org/proceedings/lrec2016/pdf/332_Paper.pdf
- Pajusalu, R. (2017). Nimisõnafraas [The noun phrase]. In M. Erelt & H. Metslang (Eds.), *Eesti keele süntaks* [The Syntax of Estonian]. *Eesti keele varamu III. Tartu Ülikooli Kirjastus* (pp. 379–404).
- Paulsen, G., Vainik, E., Tuulik, M., & Lohk, A. (2019). The Lexicographer’s Voice: Word Classes in the Digital Era. In I. Kosem, T. Zingano Kuhn., M. Correia, J. P. Ferreria, M. Jansen, I. Pereira, J. Kallas, M. Jakubíček, S. Krek & C. Tiberius (Eds.). *Proceedings of the eLex 2019 conference: Smart lexicography, 1–3 October 2019, Sintra, Portugal* (pp. 319–337). Brno: Lexical Computing CZ, s.r.o. Retrieved from https://elex.link/elex2019/wp-content/uploads/2019/09/eLex_2019_18.pdf
- Paulsen, G., Vainik, E., & Tuulik, M. (2020). Sõnaliik leksikograafi töölaual: sõnaliikide roll tänapäeva leksikograafias [On word classes in contemporary lexicography: The lexicographers’ view]. *Estonian Papers in Applied Linguistics*, 16, 177–202.

- Paulsen, G., Vainik, E., Lohk, A., & Tuulik, P. (2021). Catching lexemes. The case of Estonian noun-based ambiforms. *Electronic lexicography in the 21st century*. In I. Kosem, M. Cukr, M. Jakubíček, J. Kallas, S. Krek & C. Tiberius (Eds.), *Proceedings of the eLex 2021 conference: Post-editing lexicography, 5–7 July 2021, Brno, Czech Republic* (pp. 288–311). Brno: Lexical Computing CZ, s.r.o. Retrieved from https://elex.link/elex2021/wp-content/uploads/2021/08/eLex_2021_17_pp288-311.pdf
- Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive Development and the Acquisition of Language* (pp. 111–144). New York, San Francisco/London: Academic Press.
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology, General*, 104(3), 192–233.
- Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, Lawrence Erlbaum, New York.
- Tavast A., Koppel, K., Langemets, M., & Kallas, J. (2020). Towards the Superdictionary: Layers, Tools and Unidirectional Meaning Relations. In Z. Gavriilidou, M. Mitsiaki & A. Fliatouras (Eds.), *Proceedings of XIX EURALEX Congress: Lexicography for Inclusion, Alexandroupolis, 2021, online, Vol. 1* (pp. 215–223). Greece: Democritus University of Thrace.
- Tavast, A., Langemets, M., Kallas, J., & Koppel, K. (2018). Unified Data Modeling for Presenting Lexical Data: The Case of EKILEX. In J. Čibej, V. Gorjanc, I. Kosem & S. Krek (Eds.), *Proceedings of the XVIII EURALEX International Congress: EURALEX: Lexicography in Global Contexts, 17–21 July 2018, Ljubljana, Slovenia* (pp. 749–761). doi: 10.4312/9789610600961
- Tuulik, M., Vainik, E., Paulsen, G., & Lohk, A. (2022). Kuidas ära tunda adjektiivide? Korpuskäitumise mustrite analüüs [How to recognize adjectives? An analysis of corpus patterns]. *Estonian Papers in Applied Linguistics*, 18, 279–302. doi: 10.5128/ERYa18.16.
- Vainik, E., Paulsen, G., & Lohk, A. (2021). A typology of lexical ambiforms in Estonian. In Z. Gavriilidou, M. Mitsiaki & A. Fliatouras (Eds.), *Proceedings of XIX EURALEX Congress: Lexicography for Inclusion, 7–11 September 2021, Alexandroupolis, Greece, Vol. 1* (pp. 119–130). Alexandroupolis, Greece: Democritus University of Thrace. Retrieved from https://euralex2020.gr/wp-content/uploads/2020/11/EURALEX2020_ProceedingsBook-p119-130.pdf
- Vainik, E., Lohk, A., & Paulsen, G. (2021). The Distribution Index Calculator for Estonian. *Electronic lexicography in the 21st century*. In I. Kosem, M. Cukr, M. Jakubíček, J. Kallas, S. Krek & C. Tiberius (Eds.), *Proceedings of*

the eLex 2021 conference: Post-editing lexicography, 5–7 July 2021, Brno, Czech Republic (pp. 121–138). Brno: Lexical Computing CZ, s.r.o. Retrieved from https://elex.link/elex2021/wp-content/uploads/2021/08/eLex_2021_07_pp121-138.pdf

- Vainik, E., Paulsen, G., Tuulik, M., & Lohk, A. (in press). Towards the Morpho-syntactic Corpus Profile of Prototypical Adjectives in Estonian. *Estonian Papers in Applied Linguistics*.
- Vare, S. (1984). *Omadussõnaliited tänapäeva eesti kirjakeeles*. [The adjectival suffixes in contemporary Estonian]. Tallinn: Valgus.
- Vitso, T.-R. (2003). Structure of the Estonian language: Phonology, morphology, and word formation. In M. Ereht (Ed.), *Estonian language* (pp. 9–92). Tallinn: Estonian Academy Publishers.
- Warren, B. (1984). *Classifying Adjectives*. Göteborg: Acta Universitatis Gothoburgensis.

Od glagolskega k pridevniškemu: vrednotenje leksikalizacije deležnikov v estonskem korpusu

Članek obravnava vprašanja kategorizacije, povezana s pridevniškimi kandidati v estonščini, s poudarkom na deležnikih. Cilj analize je bil oceniti razpon prototipskega pridevnika in določiti stopnjo njegovega odstopanja na lestvici prototipskosti. Raziskava je temeljila na skupini potrjenih pridevnikov – izbranih pridevnikov, vključenih v *Osnovni slovar estonščine* – in dveh kontrolnih skupinah bolj in manj leksikaliziranih deležnikov. Preverili smo sedem morfosintaktičnih korpusnih vzorcev, značilnih za pridevnike. Testni vzorci so temeljili na prototipskih značilnostih pridevnika in na opažanjih pri konkretni leksikografski analizi. Za oceno vzorčnih besed in določitev pomena testnih vzorcev z vidika opredeljitve pridevnika smo uporabili analizo odklona. Rezultate študije je mogoče uporabiti za določitev merila pridevniške ustreznosti pri leksikografskih presojah, na primer za določanje leksikaliziranih deležnikov.

Ključne besede: korpusno jezikoslovje, leksikografija, estonski jezik, pridevnik, deležnik, analiza odklona

Okrogla miza »Jezikovna ideologija, poklicni stereotipi ter primeri dobre prakse lektorjev in jezikovnih svetovalcev«

Matejka GRGIČ, Milica ANTIĆ GABER, Vojko GORJANC, Erika KRŽIŠNIK, Tina LENGAR VEROVNIK, Eva VRTAČIČ

V okviru **2. kongresa Lektorskega društva Slovenije**¹ je bila organizirana okrogla miza »Jezikovna ideologija, poklicni stereotipi ter primeri dobre prakse lektorjev in jezikovnih svetovalcev«, ki je potekala 7. oktobra 2022. Na okroglo mizo, ki jo je moderirala **Matejka Grgič**, so bili vabljeni prof. dr. **Milica Antić Gaber**, prof. dr. **Vojko Gorjanc**, prof. dr. **Erika Kržišnik**, doc. dr. **Tina Lengar Verovnik** in doc. dr. **Eva Vrtačič**.

Zapis je pripravil **Martin Justin**. Pri tem so bile izjave za namene lažjega branja skladiščno nekoliko prilagojene značilnostim pisnega jezika, vendar po principu minimalne intervencije. Vabljene govorke in govorniki so zapis pregledali in predlagali (nevsebinske) popravke za boljše razumevanje besedila.

1 Program kongresa je dostopen na: <https://www.lektorsko-drustvo.si/kongres-2022/>.

Grgič, M., Antić Gaber, M., Gorjanc, V., Kržišnik, E., Lengar Verovnik, T., Vrtačič, E.: Okrogla miza »Jezikovna ideologija, poklicni stereotipi ter primeri dobre prakse lektorjev in jezikovnih svetovalcev«. Slovenščina 2.0, 10(1): 98–125.

Druga izvedena dela / Other performed works

DOI: <https://doi.org/10.4312/slo2.0.2022.1.98-125>

<https://creativecommons.org/licenses/by-sa/4.0/>



Matejka Grgič: Dobrodošli na sklepnem dejanju letošnjega lektorskega kongresa, okrogli mizi o lektorjih, lektorica, lektoriranju, lektorskih mitih, ideologijah in vsem, kar spada zraven. Predstavljam današnje gostje in goste.

To so: dr. **Milica Antić Gaber**, redna profesorica na Oddelku za sociologijo Filozofske fakultete Univerze v Ljubljani, sicer tudi gostujoča profesorica na številnih tujih univerzah; koordinira in vodi več raziskovalnih programov in projektov, posveča se študijam spolov in aktualnim vprašanjem v družbi. Tudi jezik je eno od takih aktualnih vprašanj v naši družbi, tako da – dobrodošli na današnjem dogodku.

Naša naslednja gostja je upokojena redna profesorica dr. **Erika Kržišnik**. Bila je ena najbolj prepoznavnih glasov Oddelka za slovenistiko Filozofske fakultete Univerze v Ljubljani. Ukvarjala se je in se še vedno predvsem s frazeologijo, vendar pa je poučevala tudi številne predmete in vsebine, povezane z lektoriranjem, kar pomeni, da je vzgojila več generacij lektorjev in lektorice in nam lahko danes ponudi res širok pogled na to področje.

Naslednji na vrsti je dr. **Vojko Gorjanc**, redni profesor na Oddelku za prevajalstvo Filozofske fakultete Univerze v Ljubljani, kjer je tudi vodja katedre za slovenistiko. Sicer pa se ukvarja predvsem s sociolingvističnimi temami in s kritično analizo diskurza. Zanima ga korpusni pristop k tem in številnim drugim, z jezikom povezanim temam, predvsem k jezikovnim ideologijam in ideologijam standardnega jezika, kar je seveda tudi za vse nas aktualna in žgečkljiva tema.

Z nami je docentka dr. **Tina Lengar Verovnik** s Fakultete za družbene vede Univerze v Ljubljani. Tudi ona sodeluje z univerzami v tujini, predvsem z Univerzo v Celovcu, in pa z Inštitutom za slovenski jezik Frana Ramovša pri ZRC SAZU, kjer je članica pravopisne komisije. Ukvarja se z jeziki medijev, slogom in, kar se mi zdi še posebej zanimivo, z jezikom mlajših generacij, kar je tudi zanimiva tema, ki jo vse premalokrat naslavljamo.

In čisto na koncu je z nami še doc. dr. **Eva Vrtačič**, tudi sama predavateljica na Fakulteti za družbene vede Univerze v Ljubljani. Ukvarja se z antropologijo vsakdanjega življenja in s temo, ki se nam je zdela izjemno zanimiva, saj ima vedno pomembnejšo vlogo v naših življenjih in jih celo že kroji – to je digitalizacija. Tema je pomembna, tudi ko se

pogovarjamo o jezikih, ki vstopajo v naše življenje. Poleg tega se dr. Vrtačič zanima za teme, kot so ideologija, participacija, transparentnost, vse nekako povezane z našim delovanjem in z vsebinami, s katerimi smo se že spopadali na tem kongresu.

Zdaj pa bi za začetek, za ogrevanje, predala besedo kar takole po vrsti. Moje gostje in gost, na kaj vas spomni beseda *lektor* oziroma *lektoriranje*? Geslo tega kongresa je »Lektorju obraz in lektoriranju mesto«. Kdo so zdaj lektorji, kaj je lektoriranje, s čim ga povežete?

Milica Antić Gaber: Ko rečete lektor, pomislim na lektorico, torej na žensko. In najverjetneje se ne motim, če pogledam, kakšna je zasedba tega vašega zbora tukaj. S tem sem hotela povedati to, kar najbrž ni nič novega, da je lektorski poklic očitno eden od feminiziranih poklicev in da je tudi to najverjetneje povezano s statusom lektorstva in lektoriranja na sploh v družbi.

Ko rečete lektoriranje, pa pomislim na poseganje v moj tekst. [smeh] Običajno med našimi kolegicami in kolegi za knjigo, ki smo jo napisali, rečemo, da je to naš otrok. In seveda vsak varuje svojega otroka in svoj tekst. Na nek način se po eni strani veselim poseganja v tekst, zaradi tega, ker se mi zdi, da se vsakič naučim nekaj novega. Po drugi strani pa se bojim, da bo poseg v tekst morda šel, kar zadeva moj odnos do jezika, kam predaleč in da bo potreben dialog med lektorico in menoj. Kar pa je včasih mogoče, vedno pa tudi ne.

Matejka Grgič: Najlepša hvala. Prosim.

Erika Kržišnik: Zame je lektor idealni bralec. Zdaj sem seveda tudi ugotovila, da *lektor* prvotno pomeni bralec, v latinščini, vse drugo je iz tega nastalo kasneje. Ko sem še predavala in imela seminarje iz teme o lektoriranju, je bil moj prvi stavek, ki se ga verjetno marsikdo od vas spomni, da je lektor idealni bralec. In mislim, da to lektor pravzaprav mora biti, da je to njegovo osnovno delo. In sicer, zakaj tako mislim? Zato, ker avtor res gleda na tekst kot na svojega otroka. Se absolutno strinjam in tudi sama sem že doživela, da je kdo tical po mojem tekstu, zato vem, kakšen občutek je to. Tudi študentom sem včasih, kdaj pa kdaj, to ni bilo redno,

rekla, naj zamenjajo svoje lekture med seboj, pa bodo videli, kaj jim bo popravil kdo drug in kako se bodo pri tem počutili.

Lektor je idealni bralec zato, ker moramo misliti na bralca, ne samo na avtorja. Zlasti, odkar imamo digitalizacijo, odkar pišemo tekste direktno v stroj in jih direktno v stroju tudi popravljamo. Jaz sem ena izmed tistih generacij, ki je recimo magistrsko nalogo še natipkala, popravila, nato pa na novo natipkala. Doktorata pa ne več. Hočem reči samo, da sem jaz v svoji glavi doživela, kakšna je razlika. Verjetno bi, če bi se nekdo zdaj s tega vidika začel ukvarjati z besedili, tudi ugotovil, da se začenja spreminjati stil pisanja. Ker če pišeš direktno iz glave, pišeš nujno drugače. Tisto, za kar mora poskrbeti lektor, je, da bralec iz besedila izve oziroma tam prebere tisto, kar je avtor hotel povedati. To je zame idealni bralec.

Matejka Grgič: Super. Najlepša hvala.

Vojko Gorjanc: Hvala lepa. Jaz na nek način lektorja z vidika svojega strokovnega dela umeščam na dve področji. Eno je na čisto raziskovalno področje: ko pripravljamo jezikovne vire, je lektor eden od uporabnikov in vsaj mi se pri virih, ki jih pripravljamo, precej ukvarjamo z uporabniki in želimo tudi vzpostaviti komunikacijo s skupnostjo, ki tovrstne vire uporablja. Drugo pa je, da poskušam opolnomočiti svoje študente, da so čim bolj jezikovno kompetentni. Hkrati pa tudi, da se zavedajo okolja, v katerem delujejo, v katerem so seveda prisotni tudi lektorji, in da so kompetentni s tega vidika, da se znajo z lektorji dobro pogovarjati. Torej, da znajo vzpostavljati komunikacijo, da znajo argumentirati svoje lastne jezikovne odločitve in s tem prisiliti tudi lektorja, da jih mora na podoben način argumentirati tudi on. To se mi zdi, da je moja naloga in da na tak način razmišljam o lektorstvu.

Matejka Grgič: Hvala.

Tina Lengar Verovnik: Ko pomislim na besedo lektor, lektorica, se zraven na žalost še vedno prikrade nek slabšalni prizvok. Ne zato, ker bi sama imela tak odnos do poklica in dela, ki ga lektorji ali lektorice opravljajo, ampak zato, ker sem ga kdaj v svoji dolgoletni, pa ne ravno intenzivni

praksi lektoriranja doživljala. In tudi zaradi tega, ker sem včasih – zadnja leta, mogoče zadnjih 10 let niti ne več toliko, ampak prej pa pogosteje – tudi od študentk in študentov novinarstva in komunikologije kdaj slišala, češ pa saj mi bo lektor, lektorica popravil. Torej, zakaj bi se moral tako zelo truditi s tem – kar je povedal tudi Vojko –, da sem ne samo več v pisanju, ampak tudi razgledan po jezikovnih virih, se pravi, da znam rešiti situacijo, da znam rešiti problem ali pa vsaj vstopiti v komunikacijo s tistim, ki mi besedilo popravlja. Velikokrat sem na žalost naletela na odnos, da je lektor, lektorica v bistvu samo čistilka oziroma čistilec, in sem morala potem tudi sama pri sebi opraviti preskok. Naredila pa sem ga tako, da o svojem delu nisem več razmišljala kot o lektoriranju, ampak kot o jezikovnem pregledovanju. Zato sem tudi vedno, ko sem zadnja leta še kje opravljala to vlogo, prosila, da zapišejo v kolofon, da sem opravila jezikovni pregled, ne pa, da sem lektorirala. Mogoče bi kdo rekel, da gre samo za preimenovanje istega, ampak moj odnos se je na ta način lahko spremenil, zaradi česar mi je lažje. In vem, da je tudi marsikomu med vami, ker smo se o tem že velikokrat pogovarjali. Vem, da poskušate vzpostavljati tudi razliko med različnimi stopnjami svojega dela – tudi na poimenovalni ravni – in da govorite tudi o jezikovnem svetovanju, kar je spet druga stvar.

Erika Kržišnik: *Nomen est omen.*

Tina Lengar Verovnik: Ja, tako, *nomen est omen*, tukaj res mislim, da to drži, moja osebna izkušnja je taka. Na ta način sem lažje tudi zajela to, kar se mi pa zdi, da je idealna lektura. Ne samo takrat, ko sem jaz prvi bralec, ampak tudi takrat, ko moje besedilo bere nekdo drug – se pravi vstopanje v dialog, ki je hkrati tudi učenje. Ko berem besedila, ki jih lektoriram, se marsičesa naučim in sem za to hvaležna. In ker sem po duhu pedagoginja, imam nenehno potrebo tudi po tem, da poučujem nazaj in dajem informacije o tem, zakaj sem nekje nekaj naredila ali pa zakaj česa nisem naredila. Zakaj sem zamenjala besedo, ki je v resnici ne bi bilo treba, ampak se je na primer že tolikokrat ponovila v odstavku, da bi bilo dobro razmisliti o zamenjavi. Banalen primer in vem, da vsi to počnete; prej smo tudi že slišali o takšnem dajanju komentarjev in razlag. Ampak ja, v okviru izraza lektor, lektorica se mi je zdelo, da

tega ne morem tako zelo legitimno početi, ker na drugi strani ni bilo tovrstnega razumevanja.

Matejka Grgič: Hvala.

Eva Vrtačič: Moj osebni odnos do te besede oziroma poklica, ki ga predstavlja, je tudi po zaslugi tega, da Tina [Lengar Verovnik] še vedno kdaj lektorira, izrazito dober in lektorja, lektorice nikakor ne jemljem kot jezikovnega policaja, pač pa kot sogovornika, sogovornico v prizadevanju za to, da bo končni izdelek čim boljši, čim bolj berljiv in razumljiv. Kot antropologinji pa se mi zdi, da je prevladujoč odnos do, kot je na začetku opozorila kolegica, po navadi lektorice, izrazito negativen. In če se malo pošalim, niti ni čudno, zakaj je temu tako. Slovenci se največ šalimo iz dveh kategorij ljudi; ena od njih je policaj, torej avtoriteta, druga pa blondinka, torej ženska. Lektorica pa je oboje. [smeh] Tako da je logično, da imamo ambivalenten odnos do tega poklica.

Matejka Grgič: Ja, prav gotovo drži. Včasih se tudi ne vprašamo, kako zelo je naš odnos do jezika in do jezikovnih poklicev, do jezikovnih storitev, tudi družbeno pogojen. In mogoče nam potem šele pogled z zunanjega gledišča omogoča širšo perspektivo. Recimo meni se, ko na slovensko stvarnost včasih gledam z italijanske perspektive, vedno poraja vprašanje, kakšna pa bi bila slovenska družba, če bi imeli več slovarjev, več pravopisov. Kaj bi se zgodilo v slovenski družbi, ali bi kot država popolnoma razpadli, če bi recimo imeli zdaj, ko se prenavlja pravopis, dva pravopisa, tri pravopise ... ali pa več slovarjev, več slovnice? Bom postavila to vprašanje najprej članici pravopisne komisije.

Tina Lengar Verovnik: V resnici imamo že zdaj dve izdaji hkrati oziroma eno, ki je še veljavna, in drugo, ki šele nastaja. Imamo pravopis iz leta 1962, ki ga marsikdo še vedno upošteva. Tako da že zdaj soobstajajo. Mislim, da ne bi bilo nič narobe, da bi mogoče drug uredniški pristop prinesel kakšne druge rešitve, ne verjamem pa, da bi bile te v vsebinskem smislu tako zelo drugačne. Odvisno od tega, kdo bi bil v uredniškem odboru, s kakšnim gradivom in na kakšen način bi se priprave pravopisa lotili. Je pa vprašanje – mogoče za kolegici družboslovki – ali

bi bila taka pluralnost Slovincem pisana na kožo. Ali bi to, da imamo možnost izbire, dejansko pomenilo, da bi se na priročnike večkrat obračali. Ali gre bolj za to, da bi morale založbe gledati malce dlje in razmisliti, kam in v kakšni obliki integrirati to, kar bodo delale. Ne nujno v neko knjižno obliko, ampak kam drugam. Zadeva postane seveda še bolj kompleksna, ko začnemo razmišljati, da se navadno šolski sistem vseeno odloči za eno samo stvar. Ministrstvo in druge inštitucije bi zagotovo potrjevale uporabo zgolj enega samega pravopisa in hitro lahko pridemo tudi do situacije, kakršno imajo na Hrvaškem, kjer je postala izjemno zapletena. Tukaj pa zadeva zelo hitro postane družbena, ne samo stvar jezikoslovja in jezikoslovcev. Osebno nimam nič proti temu, bila bi celo vesela tega, da bi se še več ljudi ukvarjalo s tem področjem, da bi nas bilo več takih, ki se s pravopisom ukvarjamo raziskovalno. To bi bilo krasno, pa nas je v resnici vedno premalo.

Matejka Grgič: Mogoče bi zdaj naprej slišali vse tri jezikoslovce oziroma jezikoslovke, potem pa bom podala žogico obema družboslovkama, torej sociologiji in antropologiji, da povzameta to jezikoslovno razmišljanje. Gremo kar po vrsti, prosim, Vojko.

Vojko Gorjanc: Hvala lepa. Jaz bi tukaj vendarle izpostavil veliko razliko med pravopisom ter slovnico in slovarjem. Zame bi bilo pravo olajšanje, če bi slovenski prostor dobil več slovarjev in več slovnic. Sem pa precej skeptičen glede več pravopisov. To je po moje spet povezano s tem, kako dojemamo pravila in kako se obnašamo znotraj pravil. Čeprav je prav kovidni čas pokazal, da so družbe precej nepredvidljive, da so se v družbah, kjer smo pričakovali, da se ljudje ne bodo držali pravil, pravil držali in obratno. Ampak tako, več slovnic in slovarjev bi po mojem v slovenskem prostoru prineslo to, da ne bi enega samega izdelka razumeli religiozno, kot Biblijo. Ker bi imeli na voljo več interpretacij istih jezikovnih fenomenov, enkrat z enega, drugič z drugega zornega kota, bi se seveda potem morali tudi sami odločati. In odločali bi se v smislu – kar je prej rekla kolegica – nekega idealnega bralca, glede na to, komu, zakaj je besedilo namenjeno. In imeli bi več informacij, zato da bi lahko lažje funkcionirali v takem okolju.

Kar se tiče pa pravopisa, je to, kar je rekla kolegica, gotovo res, in hrvaški prostor je gotovo tak. Drugo pa je na primer portugalski prostor, ki ga jaz malo bolje poznam, ker tam delam kot profesor, imam svoj predmet in precej intenzivno živim v tistem okolju. Portugalci so spremenili pravopis pred približno desetimi leti. *Acordo ortográfico* je bil narejen za vsa portugalsko govoreča okolja, v različnih okoljih pa je bil različno sprejet. Na Portugalskem je bil ta diskurz o spremembi pravil izrazito ideološki, tudi če si niso priznali; bil je izrazito postkolonialni, z vidika matere portugalsščine in vseh ostalih portugalsčin. Ampak, Portugalci nikoli ne stojijo pred rdečo lučjo na prehodu za pešce, njim je to samo sugestija. Za razliko od nas, kjer sem sam v Ljubljani že plačal, ker sem šel pri prazni cesti ob rdeči luči čez prehod za pešce. In seveda, tukaj sta ti dve okolji precej drugačni. Tam nikogar ne skrbi, da nekatere založbe še vedno uporabljajo star pravopis in da osrednji dnevnik, *Público*, še vedno uporablja star pravopis, sobivanje funkcionira neproblematično. Nov pravopis pa je povzročil predvsem to, da se otroci v šolah veliko prej in lažje naučijo pisati, zato ker je ukinil vse historične elemente v zapisu, še tiste iz latinščine.

Se mi pa zdi, da smo se, kar bo mogoče sociološki vidik še lahko dodal, Slovenci mogoče navajeni do pravil obnašati drugače. Pravila jemljemo, vsaj v nekaterih kontekstih, veliko bolj resno kot na Portugalskem. Jaz tam nekaj napišem, odpošljem, pa potem vidim, da sem naredil nekaj napak; sam opazim, da sem naredil napake v portugalsčini. Pa jim rečem, če lahko to popravim, in dobim odgovor: »Pa kje je tu kakšen problem, saj se vse zelo dobro razume.« In si rečem: »Dobro, prav, pa ne.« [smeh] To je njihov odnos do jezikovne norme. Precej drugačen kot pri nas. Ko sem včeraj prišel sem, sem prespal v mestu in lastnica penziona je imela nekako malo ruski naglas. Takoj, ko me je pozdravila, se je opravičila in rekla: »Oprostite, jaz še ne uporabljam dosledno dvojine.« In sem odgovoril: »Veste, jaz tudi ne, ker moj dialekt pravzaprav nima dosledne dvojine.« [smeh] Česa takega ne boste na Portugalskem nikoli slišali, ker so navajeni na tako različne portugalsščine, na tako različne tujce, ki različno uporabljamo portugalsščino in imajo do tega preprosto drugačen odnos. Skratka, te razlike je v različnih okoljih zanimivo opazovati. Zato se mi zdi, da bi več pravopisov v slovenskem okolju prineslo več škode kot koristi, obratno pa se mi zdi pri slovnici in slovarju.

Matejka Grgič: Najlepša hvala.

Erika Kržišnik: Jaz se strinjam. In sicer zakaj? Zato ker mislimo, da sta slovnica in slovar naravna danost. To imamo. Napisano ali nenapisano, slovenščina ima slovnico in ima slovar. Medtem ko pravopis ni naravna danost, to je pač konvencija, je dogovor, ki seveda do določene mere sledi jeziku – kaj pa vem, postavljanje vejic imamo po sintaksi in tako naprej – ampak vendarle, to je res dogovor. In v tem smislu je smiselno imeti en dogovor, da se ljudje ne izgubljajo pri stvareh, za katere pravzaprav ne vem, če je vredno ali pa potrebno, da se pri njih izgubljajo. Pri slovnici in slovarju pa je drugače. Zlasti pri slovarju: ja, seveda imamo lahko različne, saj so odvisni od tega, komu so namenjeni, katere vrste jezik hočejo predstavljati. To ni nič posebnega in res mislim, da samo pri nas nismo navajeni na več različnih slovarjev.

Medtem ko bi se pri slovnici vprašala, kaj imamo pri tem v mislih, kaj si sploh predstavljamo, da bi različne slovnice lahko bile. Recimo jaz, pardon, tudi jaz v trenutku, ko slišim, reagiram: ko je kolegica [moderatorica] rekla, čakajte, da pogledam, sem si zapisala [smeh iz publike] ... Ne, ne, hočem vam povedati eno stvar, za katero mislim, da bi morala oziroma da bo zanesljivo v kratkem času – v stotih letih sigurno, morda pa že v petdesetih – prišla v slovnico. Kolegica je rekla: »O vsemu, kar spada ...« »O vsemu«, se pravi ta »u«, ki je dativ in ne mestnik. Da do tega prihaja, je normalno. In to se bo verjetno standardiziralo. Torej, a bo ena slovnica to naredila 20 let prej, če bi bile različne, ali bo naredila 50 let prej ali bo naredila 3 dni prej, je druga stvar. Ampak taka stvar, tretji in peti sklon – v nemščini med tretjim in petim sklonom že zdavnaj ni več razlike. Razlike tudi v slovenščini ni v ženskem spolu, razlika je samo v moškem, za srednjega pa tako veste, da ga hitro priredimo po moškem. V glavnem, mislim, da se te razlike lahko učimo, lahko u-ja ne pišemo, lahko ga popravljamo, ampak to enačenje je očitno tudi en del naravne slovnice. In v tem smislu so lahko slovnice različne. Da bi si pa vsak posebej izmišljeval slovnico, mislim, da ne bi imelo kakšnega pretiranega smisla. Tudi, če bi jih izdajali.

Matejka Grgič: Najlepša hvala. No, zdaj pa prosim.

Milica Antić Gaber: Zdaj pa najtežja naloga, ali kako? Jaz bom najprej rekla en stavek za uvod, da boste razumeli moj pogled. Morala se bom majčkeno odmakniti od kolegic in kolega jezikoslovc in jezikoslovca, pa bom priznala, zaradi mojega pomanjkljivega znanja na tem področju. Jaz pač nisem jezikoslovka in ne poznam toliko teh pravil, kot jih poznate vi in jih upoštevate. Jaz jih kršim, upam, da me ne boste, Erika, zdaj analizirali, koliko napak bom naredila v svojem govornem jeziku. Iz več razlogov. En od razlogov je ta, da je slovenščina moj drugi jezik, da je nek drugi jezik bil moj prvi, materni jezik, in da sem se slovenščine naučila v obdobju, ko sem bila že jezikovno sposobna, da temu tako rečem. Tako da sem se morala na novo naučiti – seveda v osnovni šoli – tega jezika. Ki sem ga v precejšnjem obdobju svojega življenja, zaradi tega, ker sem pač otrok priseljencev, govorila na način, ki ga v knjigah lepo popisuje Goran Vojnovič. Miksanje, bi rekel Goran. To je en razlog. Drugi razlog je ta, da sem družboslovka, in v družboslovju smo vedno dajali prednost vsebini pred formo. Zato tudi v mojem formiranju v družboslovju na takratnem FSPN-ju² ni bilo toliko pozornosti dane jeziku, kot je verjetno zdaj, ko v študijskem procesu študentkam in študentom ponujajo veliko več. Tretji razlog pa je, ker sem sociologinja in kot sociologinja in kot družboslovka razmišljam o tem, da so pravila seveda zato, da jih upoštevamo, ampak tudi zato, da jih kršimo. Do tega imam torej nek intimen odnos. Seveda je treba vedeti, zakaj se neka pravila kršijo, ne kar tako, ne kar počez, ne kar zaradi tega, ker se mi je danes to zahotelo. Tako da, jaz v jeziku kršim pravila in si včasih vzamem pravico, da to tudi uveljavljam in puščam v svojih tekstih. Predvsem z vidika uporabe novih besed, novih izrazov, pa ne samo zaradi prevajanja iz angleščine, ki ga je v družboslovju veliko, ampak tudi sicer, zaradi mojega odnosa do jezika nasploh in do pomembnosti izražanja v vključujočem jeziku. Zato ga najverjetneje za zelo stroge lektorje in lektorice hudo kršim.

Zdaj pa, kar zadeva tak sociološki pogled. Zdi se mi, da ne bi bilo, tako kot ste rekli vsi trije, čisto nič narobe, če bi imeli več slovarjev. Tako kot pravi Vojko, tudi jaz mislim, da jih že obstaja več, samo da si nekateri tega nočejo priznati. Zelo pomembno je, komu je slovar namenjen,

2 Fakulteta za sociologijo, politične vede in novinarstvo (FSPN), ki se je leta 1991 preimenovala v Fakulteto za družbene vede (FDV).

kako pišemo. Ko pišem tekst za časopis, za kakšno prilogo, je moj jezik drugačen, tudi moj besednjak je drugačen, kot je moj jezik in moj način pisanja pri pisanju znanstvenih tekstov, ker seveda ciljам na drugačno publiko. Tako da se mi zdi, da si je tukaj mogoče izboriti veliko svobode. To me je naučil eden od vaših kolegov, ko mi je lektoriral moj prvi tekst. To je bil moj doktorat, kjer mi je rekel: »Vzemi si svobodo. Izbori si nove izraze.« In to sem tudi naredila, pa je doktorat vseeno šel skozi. [smeh] Zdaj, zakaj je šel skozi, ne vem, morda takrat še ni bilo takšne občutljivosti glede jezika. Ampak kar zadeva slovenski prostor nasploh, pa mislim, da bi delalo kar precejšnjemu delu slovenske populacije probleme, če bi imeli več slovníc in več pravopisov, ker nismo navajeni na pluralnost in se na pluralnost pravzaprav v zadnjem času šele navajamo. Je pa ta recimo prisotna v angleščini: angleščina ni ena, angleščin je tisoč. Dobro, angleščina ima drugačen status, slovenščina je majhen jezik, majhno okolje, zaprto okolje. Ampak vendarle, če jaz poslušam nekoga s Primorske ali s Koroške, bi včasih rabila slovar, da bi se med seboj razumeli. Ampak to je treba privzeti. Mislim, da so vse te razlike v jeziku legitimne in da jih moramo seveda upoštevati in temu primerno ravnati tudi na družbeni ravni.

Matejka Grgič: Najlepša hvala. Prosim, žogico podajam na drugo stran »okrogle mize«.

Eva Vrtačič: Jaz se strinjam s tem, da bi največja težava nastopila pri uporabnikih in uporabnicah, zlasti pri splošnih uporabnicah. Za strokovno javnost je seveda fino, da ima na voljo čim več različnih gradiv in priročnikov, večina ne samo govorenega, ampak tudi pisanega jezika pa se vseeno odvija na drugih nivojih in v drugih javnostih. Za slednje bi si že zdaj kot pedagoginja, kot gledalka in poslušalka medijev želela, da bi več posegali po obstoječih jezikovnih priročnikih. Čeprav sama nimam odklonilnega odnosa do tega, da bi obstajalo mnoštvo različnih priročnikov, se mi vendarle zdi, da je bistveno bolj pomembno tiste, ki obstajajo, približati ljudem, jim jih prikazati kot njihove zaveznike, ne pa sovražnike. Predvsem skozi evolucijo odnosa do jezika, ki bi lahko sestopil s piedestala, naš odnos do njega pa se lahko na novo definira kot sproščen, igriv, to pa potem lahko spet preraste v tako opevano ljubezen do materinščine. Iz nje lahko morda izide novo zanimanje za

pravila, učenje normativne rabe in podobno. Mislim, da ima žuganje s prižnice, pa tudi, če je jezikovno, nasproten učinek. Pa če je ta ena ali pa jih je pet različnih, celo če so vmes protestantske.

Matejka Grgič: No, pri naslednjem vprašanju pa bi začela z našima družboslovkama. In sicer: omenjen je bil covid. Ne bomo ga seveda obravnavali kot medicinski problem, ampak mogoče kot razmislek, kako v družbi funkcionirajo pravila. Kdaj so pravila učinkovita, čemu morajo zadoščati, da delujejo, da jih ljudje spoštujejo, kdaj pa gre vse narobe? Kdaj se zgodi, da tudi takrat, ko vemo, da bi bilo mogoče vseeno fino, če bi ravnali po pravilih, ne ravnamo tako? Kdaj jih mogoče niti ne slišimo? Kdaj pa se mogoče zanalašč odločamo v nasprotju s pravili in rečemo: »Ne, tega se pa ne bom držala, pa tudi če se vse skupaj se-suje.« Skratka, česa smo se naučili, in predvsem, česa se nismo naučili o pravilih po tem, ko smo preživeli oziroma preživljamo epidemijo?

Eva Vrtačič: Mislim, da mora vsako pravilo, da bi bilo uspešno, v smislu, da bi ga ljudje upoštevali, nasloviti neko njihovo tesnobo in jim ponuditi odgovor nanjo. Se pravi, na primer: tvoj test kovid je pozitiven, osami se, da ne boš okužil drugih. Okej, *makes sense*. Razesaja virus, ostani doma po deveti zvečer. Kaj pa vem, če slednje res odgovarja na težavo, ki jo imam v tistem trenutku. Mislim, da v jeziku ni veliko takih pravil, ki bi nam delala težave. Največkrat jih kršimo po nesreči, ne nalašč. Tisti, ki jih kršijo zavestno, pa morda jezik obvladajo in pravila kršijo z razlogom, na primer iz literarnih razlogov. Mislim, da običajen uporabnik, ki naredi neko napako, te ne naredi zato, ker se bi, da rečem grdo, rogal slovenščini, ampak zato, ker ni pozoren, je pozabil pravila ... Saj so nas v šoli kar precej »mučili« s slovnico, ampak še rajši pa so nas »mučili« s književnostjo. Če ob tem odprem še eno temo; ne vem, zakaj slovenščina vsaj v srednji šoli ni razdeljena na dva predmeta. Saj se tudi filozofija dogaja v jeziku, pa ima svoj predmet, zgodovina se dogaja v jeziku, pa ima svoj predmet, zakaj pa je literatura del slovenščine? Imamo relativno kompleksna jezikovna pravila. Če jih človek res hoče usvojiti, obvladati, mora imeti za to poseben interes, ljudem, ki ga nimajo, pa tega ne smemo zameriti. No, ne vem, če sem ravno odgovorila na vaše vprašanje.

Matejka Grgič: Super iztočnica je bila, odlično. Saj tukaj ne gre toliko za to, da odgovarjamo na vprašanja, ampak morda bolj, da odpiramo nova. To, kar ste povedali, je odlično novo vprašanje za kolegico.

Milica Antić Gaber: Jaz bom začela s primerom. Epidemija nas je naučila uporabljati – vsaj tiste, ki jih nismo uporabljali prej, in jaz sem ena med njimi – vse te različne možnosti komuniciranja preko različnih tehnologij. Ena od teh tehnologij je bil Zoom; hvala Univerzi, da nam je to omogočila, da smo se kar nekaj naučili. In jaz, ko sem imela predavanja po Zoom aplikaciji, sem moje študente in študentke, ko sem poskušala vzpostaviti kontakt z njimi, najprej prosila, da imajo vklopljene kamere. Z oddelka za psihologijo so nas sicer opozarjali, da to morda ni najbolje, da gre za vprašanje avtonomije oz. zasebnosti, vpogleda v zakulisje in tako naprej. Zato sem jim rekla, da zaradi mene lahko ozadje zamegljijo, ampak naj prosim ostanejo, da vidim, komu predavam. In to sem naredila zato, da sem vzpostavila kontakt z njimi, da sem predavanja vodila tudi nekoliko interaktivno, ker drugače sploh nisem vedela, ali so še z menoj ali ne. No, in kaj sem ugotovila, ko sem vzpostavila kontakt z njimi? Da so mi začeli govoriti v svojem narečju. Ne vem, očitno je bivanje v domačem okolju, bivanje v njihovih sobah, s starši in tako naprej, pripomoglo k temu, da se niso zavedali, da so v pedagoškem procesu, kjer komuniciramo v knjižnem jeziku. In so začeli govoriti v štajerščini, primorščini, koroščini in tako naprej. Po eni strani se mi je to zdelo fino, po drugi pa sem se pravzaprav zavedala tega, da oni zdaj – to sva se prej pogovarjali z Eriko – situacijo očitno razumejo čisto drugače, ker je bil to nek čisto drugačen proces. Če bi bili v predavalnici, bi zagotovo uporabljali knjižni jezik. To se mi je zdela taka, mogoče tudi z jezikoslovnega in sociolingvističnega vidika zanimiva situacija.

Sicer pa – zdaj sem pa že skoraj pozabila vprašanje, pravila, ne? [smeh] Pravila, ja, seveda, jaz mislim, da pravila upoštevamo takrat, ko jih na nek način ponotranjimo. In da jih ponotranjimo, potrebujemo neko racionalizacijo, zakaj je to pravilo dobro. Ne samo dobro za mene, ampak tudi za druge. V času korona obdobja sem ugotovila tudi, da se je družbeni prostor zelo individualiziral, namesto, da bi se odprli v skupnost. Vse se je zvedlo na vprašanje, zakaj moram jaz? Na primer, zakaj moram jaz zdaj nositi masko? To se je pokazalo v šolskem pro-

cesu, ko so se starši razburili, zakaj mora njihov otrok, če z njim ni nič narobe, nositi masko. Ali pa so trdili, da mu bo to škodovalo, ker bo, kaj že, vdihaval svoj izdihani zrak in tako naprej. Ne razmišljati o tem, kaj to pomeni za druge, tudi za skupnost, to se mi zdi izjemno pomembna izkušnja prav iz tega obdobja. Skratka, moram ponotranjiti to, da je to pravilo dobro za mene in za širšo skupnost in takrat ga bom tudi sama upoštevala. Jaz se dobro spomnim te razlike, kako so – ker se pač seveda s tem tudi nekoliko raziskovalno ukvarjam – različni politiki in političarke v različnih državah nagovarjali tako imenovano ljudstvo, da naj upošteva pravila. Pri nas smo imeli nagovor v stilu: »Upoštevajte navodila, sicer bo sledila kazen.« Oziroma, kako je že rekel eden od govornikov vlade: »Uživajte, dokler lahko.« Namesto, da bi vodili ljudstvo z lastnim zgledom, kot recimo nekatere predsednice vlade po svetu, ki so se pokazale v svojem zasebnem okolju in iz svojih lastnih izkušenj rekly, da bomo malček potrpele, potem bomo pa morda lahko spet normalo zaživele. To je bilo čisto drugače sprejeto kot grožnja, da če ne boste upoštevali pravil, boste pa, ne vem, 400 evrov kazni plačali, če boste jedli rogljiček na cesti. In to se je tudi zgodilo. Se mi zdi, da je ta odnos, ta relacija zelo pomembna v družbi.

Matejka Grgič: Zdaj pa pridemo k jezikoslovkam oziroma jezikoslovcem. Kako pa znamo mi komunicirati jezikovna pravila? Ali mogoče prevečkrat takole zažugamo s prstom? A se mogoče splošna javnost boji predvsem slovenistk in slovenistov? Kaj pravite, profesorica Kržišnik?

Erika Kržišnik: Aja, da se naši odjemalci, da rečem temu tako, bojijo, da bo vse rdeče? To bi morali vprašati koga drugega ...

Matejka Grgič: Mislim, da smo prej slišali neko pričevanje lektorice, ki je povedala, da ji v podjetju, za katero dela, neradi dajo besedila, ker pravijo, da bo koga preveč izpostavila. Kaj pravite o tem?

Erika Kržišnik: Mislim, ne vem natančno, o čem se govori, jaz s svojimi lekturami – in moram reči, da sem jih v življenju naredila kar nekaj – nimam slabe izkušnje. Imam pa dobro: profesor Zadravec, ki sem mu lektorirala tekst v nekem zborniku, je prišel k meni in rekel, da tako,

kot sem jaz prebrala njegov tekst, njegovih člankov ni prebral še nihče. Jaz se niti ne spomnim, ali sem kaj popravila, ali sem malo ali veliko, nimam pojma. Vem pa, da mi je to rekel, in od takrat je moje vodilo, da je lektor idealni bralec.

Kar se pravil, izrečenih tako z žuganjem ali s pridobivanjem, tiče, sem prepričana, da je res, da bolj zaleže pridobivanje kot žuganje. Ker ne vem, o kakšne vrste lekturah govorimo, težko rečem, ali je bilo to žuganje ali kaj drugega. Očitno pa obstaja strah, kaj bo, če mi lektor pogleda tekst, kar verjetno pomeni, da sami nismo prepričani o tem, da dobro pišemo. Ker dejstvo je – da se vrnem – jezik je forma, jezik je oblika, telepatsko ne moremo komunicirati. Spremembe v jeziku so dokončno ustaljene, ko dobijo svoje formalno potrdilo. In mislim, da je med formo in formalizmom velika razlika. Forma nekaj pripoveduje, odstop od pravila tudi nekaj pripoveduje in dokler nekaj pripoveduje, je vse v redu. Če pa ne pove nič, pa mene kot bralko to zmoti. Niti me kot bralko ne jezi to, da je narobe napisano, ne, jezi pa me, da moram iti še enkrat nazaj, da ugotovim, kaj je narobe, da ne razumem. Tule sem si celo prepisala en odlomek iz zadnje Sobotne priloge. A dovolite, da preberem? »Ohranjajo razprave, o katerih so se zdele pred desetletji nujne in so vključevale ...« Tu sta se prekrili dve skladenjski formi, »ki so se pred desetletji zdele nujne« in »o katerih se je pred desetletji zdelo nujno, da se ...«. Izgubila sem dve minuti, da sem došla, za kaj gre, ker je bila seveda poved zelo dolga. Če napake ne bi bilo – to niti ni slovnična napaka, to je pač prekrivanje dveh skladenjskih struktur – tega časa ne bi izgubila. In lektor – zato jaz pravim, da je idealni bralec – lektor pravzaprav sploh ne bi popravil napake, ampak nekaj, česar pisec ni opazil: da sta se mu v glavi dva vzorca – da nekaj poveš na tak ali drugačen način – prekrila in je nastala zmešnjava. To je to.

Matejka Grgič: Najlepša hvala.

Vojko Gorjanc: Jaz se bom vrnil na Portugalsko. Tam so se kovidnih pravil tako dosledno držali, da sem komajda dojel, da je to možno. In še danes, kljub temu da ga imajo bistveno manj kot mi, ne boste v stanovanjski zgrabi, ki ima dvigalo, nikoli več doživeli tega, da bi se kdo z vami peljal v dvigalu, ker so obdržali to previdnost, da se v dvigalih skupaj

vozijo samo sostanovalci. Tudi midva sva se pogovarjala [moderatorici], da so se Italijani držali kovidnih pravil veliko bolj, kot bi kadarkoli kdo pričakoval. Poleg tega, kar je že bilo rečeno, da so ustvarjali diskurz, ki je bil drugačen, da so s pozicij moči komunicirali ljudje, ki so jim v družbi bolj zaupali in so zares vlivali zaupanje tudi s svojim početjem, poleg vsega tega so zato, da so jih lahko ponotranjili, potrebovali razumljiva pravila. Tukaj pa mislim, da imamo mi težavo, ko gre za jezikovna pravila. Naša pravila, če pogledamo slovenski pravopis konkretno, niso tako zelo razumljiva. Od tega, da imamo tam zglede za postavljanje vejice, vzete iz Prešerna, do tega, da so pravila dejansko taka, da potrebujejo interpretata. In zato, da bi lahko ponotranjili pravila, jih moramo razumeti. In to bi moral biti najbrž prvi korak pravopisne komisije, da v prihodnje zelo razmišlja o tem, kako bodo pravila predstavljena.

Erika Kržišnik: Ali lahko samo medklic. Pravopis iz leta 1962 je izšel tudi v skrajšani obliki kot šolski pravopis.

Vojko Gorjanc: Vsi starejši pravopisi so veliko bolj razumljivi, vsi. Še tisti iz 1899 je veliko bolj razumljiv.

Tina Lengar Verovnik: Vojko mi je z jezika vzel tole o razumevanju in razumljivosti. Ampak bom začela z eno anekdoto. Ko smo se lotili prenove pravopisa, smo pri prvem poglavju še s precejšnjim strahospoštovanjem preoblikovali kakšne člene, nismo si upali preveč. Ko pogledam, na kakšen način delamo zdaj, se pravzaprav sploh ne oziramo več na dikcijo obstoječega pravopisa, ampak ga delamo res na novo. Javna razprava, ki je resda v prvi vrsti namenjena strokovno najbolj usposobljenim uporabnikom (ker tudi pravopis, ki ga delamo zdaj, v resnici delamo predvsem za jezikovno izobraženega uporabnika), je namenjena tudi razpravi in povratnim informacijam o tem, ali so pravila dovolj razumljiva, ali so zgledi ustrezno izbrani. Ker zelo pomembna stopnja je tudi, ali so vzorčni zgledi res taki, da omogočajo še boljše razumevanje. Hvaležni smo za vsakršne odzive. Tudi mi se v tem smislu razvijamo, se pravi, se odmikamo od dikcije, ki je bila v pravopisu iz leta 2001 res pripeljana do skrajnosti.

Ampak če se vrnem nazaj k epidemiji in pravilom, se mi zdi, da nas je naučila, da nobeno pravilo ni čisto brezprizivno. In tega se je dobro

včasih zavedati tudi v jeziku. Da to ni svetinja, ki jo bomo zdaj jemali s takšnim strahospoštovanjem, kot smo se mi lotevali prenove pravopisa, ampak je treba to početi z zdravo mero občutka za tukaj in zdaj, pa tudi za pluralnost, ki jo je prej omenjala profesorica Antić Gaber, ki se zdaj tudi v jeziku kaže veliko bolj kot kadarkoli prej. Prej pravzaprav tega niti ni bilo, saj imamo šele po 90. letih tako veliko produkcijo najrazličnejših javno dostopnih besedil v najrazličnejših registrih pisanja. Pa to, da ni vse formulirano samo v obliki pravila. Tudi mene zelo moti ali pa me v največjo zadrego spravi, ko nekega pravila ne najdem tam, kjer bi ga pričakovala. To bi mi bilo v pomoč, čeprav bi mogoče potem šla še sama preverjati po korpusih. Ampak če se nimam sploh na kaj opreti, pa je to veliko težje.

Hkrati pa čisto vse v jeziku tudi ni v pravilih. Prej se je večkrat omenjalo, kolikor sem ujela razpravo, da so predvsem študenti, pa verjamem, da tudi izkušeni lektorji in lektorice, v zadregi, kako in do kod posegati. In kaj, ko je treba v resnici reagirati glede na žanr, glede na naslovnika, skratka, ko je treba vse te parametre upoštevati tudi pri lekturi. To pa po mojem ni več stvar pravil, ampak socializacije v jezik. Včasih študentom, ko se pogovarjamo o govornem jeziku, rečem, da bodo pri branem besedilu, ki ga bodo pripravili, veliko bolj sledili pravilom, ker je to pač pisano v domeni standarda, tega, kar se da standardizirati. Tudi ko gre za prosto govornen jezik, ni vseeno, kako govorite na radijskem programu ali pa v javni situaciji; velike razlike bodo med temi diskurzi. Ampak dokler ne boste socializirani v to, s pomočjo uredništva, tudi lektorjev, trenerjev in tako naprej, se boste seveda lovili in delali napake, nihče pa vam ne bo dal v roke priročnika, ker ga ni. Tu pa je razlika – skačem nazaj na prejšnjo temo – s pisnimi mediji, kjer pa se dogaja, da nekateri oblikujejo svoje »slogovne« priročnike. In to je tisto, o čemer lahko tudi na ravni pravopisa govorimo kot o variantnosti, razliki. Vendar prinaša to seveda s seboj tudi odgovornosti. Lektorske ekipe znotraj medijev morajo za to, kakšno jezikovno podobo in kakšno jezikovno politiko peljejo, sprejemati tudi odgovornost, in včasih se mi zdi, da je to težko. Včasih bi človek raje, da bi mu pravopis povedal, kako mora biti, da nimaš nobene možnosti izbirati med različnimi odtenki. Ker to po svoje pač prinaša več odgovornosti.

Matejka Grgič: No, mi počasi zaključujemo naše srečanje. Ob koncu bomo seveda dali besedo tudi občinstvu, ker mislim, da je bilo kar veliko zanimivih iztočnic. A dovolite, čisto za konec – smo na dvojezičnem območju. Za trenutek se postavimo v kožo drugega ali v svojo kožo na drugi strani meje: če bi se v Trstu sestali italijanski lektorji (ki jih sicer ni, ampak recimo, da so), bi bilo zame in za celotno slovensko skupnost, ki tam živi, izjemno pomembno, če bi kdo pozdravil v slovenščini, vsaj dve besedi. Zato mi dovolite, da pozdravim v italijanščini.

[pozdrav v italijanščini]

Torej, preden zaključimo, še zadnje vprašanje za čisto vse moje goste in gostje na tej okrogli mizi. Ne več o tem, kaj se dogaja s slovenskim jezikom, ki da je na udaru tujih, bližnjih in daljnih jezikov, ampak česa smo se in česa se lahko še naučimo iz tega, da slovenski jezik prihaja v stik z drugimi jeziki, kako se ta stik odraža v lepih stvareh v našem jeziku?

Eva Vrtačič: Dobro vprašanje. Ne vem, meni se zdi odveč zganjati paniko. Družboslovci in humanisti vemo, da velikokrat moralna panika vznikne zaradi novih tehnologij. Predvsem zaradi tega, ker smo bili mi socializirani v druge tehnologije. Tako kot so sprva za romane govorili, da se ženske, ki jih berejo, povsem izklapljujejo iz realnosti in živijo v nekem fantazijskem svetu, odrezane od sveta, se podobno danes govori o družbenih omrežjih. Podobno se mi zdi tudi z jezikom: slovenščina, kot so se je naučili naši starši, stari starši, se je zelo spremenila, slovenščina, ki smo se je v šoli učili mi, se spreminja pred našimi očmi. To povzroča nek strah, tesnobo, kakorkoli, ki pa je po mojem brezpredmeten.

Matejka Grgič: Hvala.

Tina Lengar Verovnik: Jaz se bom navezala na to, kar ste vi povedali pred tem, se pravi, da prihajate iz prostora za mejo. Že nekaj let poučujem na Univerzi v Celovcu, to je meni tudi bližji prostor, saj prihajam iz slovenskega dela Koroške. Tri mesece sem bila tam tudi na študijskem bivanju in poznam veliko ljudi. Nekaj, česar se premalo zavedamo in se za italijanski prostor, vsaj zdaj, v okviru projekta RSDO, končno dogaja, je to, da v okviru pluralnosti knjižnega jezika obstajajo različne sloven-

ščine tudi v smislu osrednjega in teh drugih prostorov, tudi Porabja po svoje, kjer pa je situacija še bolj specifična. In odkrivati ta jezik, ki ima svoje specifikke, ki se jih včasih nekateri govorci tam žal sramujejo, češ da še vedno vnašajo dialektizme ali pa da je njihov jezik zastarel, v resnici pa gre le za tam živečo in vitalno knjižno slovenščino – to je nekaj, kar je mene obogatilo, pa bi verjetno še marsikoga drugega, če in ko se bo to še malo bolj vtisnilo v našo zavest. Tudi to, kaj prinaša aktivni stik s tujejezičnim okoljem. Se pravi, ne samo to, da mi zdaj zelo enostavno prevzemamo ali pa pač ne prevzemamo. Ali pa se postavljamo na neke obrambne pozicije, ali bomo zdaj v terminologijo nekaj sprejeli ali bomo dovolj hitro tvorili lastni izraz. Ali se bomo, kot je rekla Eva, šli moralno paniko in mlade ves čas obtoževali tega, kako nedomoljubno se vedejo v jezikovnem smislu, čeprav je dokazano, da je prevzemanje iz tujih jezikov definicijska prvina slenga.

Eva Vrtačič: Te lahko zmotim z enim primerom?

Tina Lengar Verovnik: Daj.

Eva Vrtačič. Lektorskim, kako je geografski determinizem postal zemljepisna naddoločenost, recimo.

Tina Lengar Verovnik: Ali pa fizični delavci – telesni delavci. To je primer iz mojega besedila oziroma branja. Ja, skratka, tega stika, tega dnevnega stika, živeti v svojem jeziku kljub temu, da je to nekaj povsem drugega. Ko si v resnici obkrožen z drugim jezikom in ko se dolga desetletja javno o tem v veliki meri sploh ni govorilo. Šele zdaj se na čisto drugačen način v čisto drugačnih okoliščinah ta slovenščina revitalizira ali pa se jo pač uči kot tuji jezik. Vem, da sem zdaj malo zašla.

Matejka Grgič: Ne, super je bilo.

Tina Lengar Verovnik: To je moja asociacija.

Matejka Grgič: Najlepša hvala.

Vojko Gorjanc: Slovenščina in Slovenci smo vedno živeli tukaj, na tem prepihu različnih jezikov in kultur, in po mojem se mogoče niti ne zavedamo ali pa ne ozavestimo tega privilegija. Tega privilegija, da smo nekaj korakov stran od zelo različnih kulturnih okolij, romanskega, germanskega, madžarskega, ki je sploh popolnoma drugačen, in nekega jezikovno-kulturno vendarle bližnjega, južnoslovanskega. To je velika prednost, ki jo marsikdo v marsikaterem okolju – nekdo na jugu Italije, nekdo v Madridu, sredi Španije – nikoli ne more živeti na tak način, kot jo mi. To je velika prednost, o kateri bi bilo dobro tudi govoriti kot o prednosti in ozaveščati, zakaj je ta naša pozicija nekaj, kar je izjemno dragoceno.

Matejka Grgič: Prosim.

Erika Kržišnik: Ja, lahko kar nadaljujem. Absolutno se strinjam s kolegom in sicer bom povedala, zakaj. Ko človek iz svojega prvega jezika, naj bo to dialekt, naj bo to neka mestna govorica ali karkoli, pride v stik z drugimi govoricami, jeziki, kakorkoli že temu rečemo in za karkoli konkretno že gre, je to vedno bogatenje. Nujno. Torej je pravzaprav bogatenje tudi stik med slovenščino in katerimkoli drugim jezikom. To velja, dokler imamo govorci slovenščine do svojega jezika pozitiven odnos. Dolej je to bogatenje. Hvala.

Milica Antić Gaber: Se zelo strinjam. Vsi ste tako lepo povedali, bi se pa navezala na vas [moderatorko], ki ste rekli, da smo tukaj, in sem se spomnila na meje, na geografske meje, jezikovne meje, kulturne meje, ki jih več ali manj vsi prestopamo, celo zelo pogosto. Jaz jih zadnje čase zelo pogosto prestopam, tako kot je Vojko rekel, v vsa ta geografska področja in zadnje čase še nekoliko več v Istri. In v Istri se mi zdi izjemno lepo, zelo lepo se počutim, kadar sem tam, ker slišim tako pogosto prestopanje iz hrvaščine v italijanščino in tudi v slovenščino. In to sporazumevanje v več jezikih, tako kot ste vsi rekli, je tako izjemno velika prednost tega našega prostora, da se mi zdi, da je to treba nekako prevzeti. Če primerjam to s tem, kar se dogaja recimo mojemu sinu, ki že 15 let v živi Franciji s Francozinjo, ki ga je, verjamem, zelo lepo naučila govoriti francoski jezik. Ampak skorajda v vsakem pogovoru z novimi ljudmi, ki jih še ni srečal, ga vprašajo, od kod prihaja. Skratka, takoj uja-

mejo neko izgovorjavo, neke besede, ki niso pravilne, in ga vprašajo, od kod je. Tega, ko sem živela v Angliji, recimo, nisem doživela. Angleži so navajeni, da obstaja tisoče angleščin, Francozi pa zelo striktno jemljejo pravila in svoj jezik, kar verjamem, da je prisotno tudi v slovenščini. Ker se takoj, ko nekdo spregovori z naglasom, kot si ti rekel, Vojko, sprašujemo, od kod je prišel. S tem, da – in zdaj bo prišel ta moj sociološki uvid – smo veliko bolj pozitivno naklonjeni nekemu, ki prihaja iz nekega zahodnoevropskega sveta in jezika, kot nekemu, ki prihaja iz južnoslovenskega okolja, saj slednjemu takoj pripišemo nek negativni priznak. To pa je nekaj, česar bi se lahko znebili, glede na to, v kakšnem geografskem prostoru živimo.

Matejka Grgič: Najlepša hvala. In sedaj predajam ... ne mikrofona, ker mikrofona nimamo, ampak besedo občinstvu. Prosim.

Iz publike: Jaz bi se samo dotaknil opombe, da je pravopis zapleten. Je obsežen, vsekakor, ta zadnji pravopis iz 2001, in ima tudi problematična, torej protislovna mesta, tudi nedorečenosti, ki pustijo [vprašanje] malce odprto, pogosto se pojavlja »navadno je tako«, pa pravzaprav ne vemo, kako si lahko pomagamo s tem »navadno je tako«. Ampak sicer se mi zdi, da so ta mesta bolj posledica neke nepozornosti, ki je bila v tej količini dela in naglici (kljub temu da je nastajal tako dolgo), vseeno razumljiva, ne pa sad nekega pomanjkanja čuta za preglednost. Če naredim neko primerjavo. Povsod je pomembna vaja. Jaz ne pričakujem, da drugi vadijo pravopis, ker pač delajo druge stvari. Ampak, recimo, jaz se veliko vozim s kolesom. In vsak dan gledam gume, sprednje, zadnje kolo. Ko sem se prvič lotil menjanja zračnice, mi je bilo to zelo težko, pa sem se ga lotil samo na sprednjem kolesu. Pa sem petkrat zamenjal zračnico in sem se je lotil še na zadnjem kolesu. Pa nisem bil zadovoljen in sem ga nesel na servis, da to naredi nek strokovnjak. Ampak sem se pa lotil. Hočem povedati, da je stvar mogoče res zastrašujoča, ampak je to vedno odvisno od tega, koliko časa nameniš tej stvari. Hvala.

Vojko Gorjanc: Lahko odgovorim? Jaz mislim, da ne gre za zapletenost, ampak za razmerje med zapletenostjo in razumljivostjo. Meni se zdi logično, da je vsebina zapletena in tako naprej. Ampak kako je ta razložena,

tukaj pa imamo ogromno prostora, da lahko naredimo drugače in bolje. Ker pravopis ni namenjen izključno strokovnim uporabnikom, ampak bi pravzaprav moral biti namenjen vsem pišočim. Tam bi morali najti odgovore na svoja vprašanja. Ali pa jih rabimo več, tako si ti [moderatorka] prej rekla. Saj je bil ves čas ob pravopisu še šolski pravopis. In prvi slovenski pravopis je bil v bistvu šolski pravopis, ker je bil narejen ravno za šole. In tam so pojasnila narejena na ta način, da rečejo: »Postavite v tako in tako okolje, pa boste videli, kako se stvari sklanjajo.« In v tem smislu se mi zdi, da bi morala biti pravopisna pravila bolj razumljiva.

Tina Lengar Verovnik: Če lahko navežem svoj odgovor na tisto prvo vprašanje o več pravopisih v več založbah, se mi zdi, da imajo druge založbe priložnost v tem smislu: ko bo prenova končana, ja, narediti prilagoditve za šolo, oblikovati različne digitalne storitve in tako naprej. To pač ni več stvar Pravopisne komisije, ampak teh, ki se s tem ukvarjajo in so specialisti na tem področju.

Iz publike: Jaz bi se tudi mogoče navezala na vaš [od moderatorke] včerajšnji komentar, kaj je šlo narobe v šolstvu, v pravopisu, da lektorji dobivamo doktorate, ob katerih komentiramo, da so ti ljudje nepismeni. Naj povem še svojo izkušnjo. Zadnje dve leti je šlo skozi moje roke, kot se reče, okoli petsto ljudi, ki so se udeležili mojih delavnic *à la* »ABC pravopisa« oziroma »Lepa slovenščina na delovnem mestu«. Sodelovala sem s podjetjem, kjer je zaposlenih več kot štiristo ljudi. Večina, 70 procentov teh zaposlenih, ima magisterij, recimo 10 procentov doktore, vsi so diplomirani. Izjemno izobraženi ljudje, izjemno. Njih je recimo ob tem, da po »Pozdravljeni« v [primerih tipa] »Pozdravljeni, gospa Erika Kržišnik«, sledi vejica, recimo, zadela kap. Doktorje in magistre. Pri tem, kaj se piše z veliko in z malo, česar jim tudi sama pogosto ne znam dobro razložiti, so blazni problemi. Uporabniki virov, recimo, ki bi bili radi uporabniki pravopisa, ga ne razumejo. To je naš problem. In moram povedati svojo izkušnjo. Meni se zdi lektorski poklic moralno izredno cenjen. Razprava za drugič: morali bi narediti, da bo cenjen še materialno. To je treba. Ampak, ali veste, kaj jaz velikokrat občutim? Zakaj imam jaz, ki pač razumem pravopisna pravila, zdaj to moč, da nadvladam nekega strokovnjaka in šarim po njegovem besedilu, ker

pač on nima tega časa in te dojemljivosti, da bi to razumel? Moral pa bi, ker, kot ste povedali na začetku, pravopis je od nas vseh.

Matejka Grgič: Mogoče še naslednje vprašanje, pa potem bo čas za odgovor.

Iz publike: V bistvu ste na to že nekako odgovorili, moj pomislek je bil predvsem, [kako] narediti pravopis, ki ima dikcijo, ki pravzaprav nagovarja vse. Drugo moje vprašanje pa je, kako zdaj s tem pravopisom doseči vse uporabnike, da se nam ne bo zgodilo to, kar se je pravzaprav zgodilo Portugalcem oziroma Francozom, za katere je pravopis, podobno kot za nas, pač svetinja, in so imeli reformo pravopisa, a je zdaj še vedno zmeda. V bistvu ta pravopis, čeprav naj bi stvari poenostavil in do neke mere jih je, ni dosegel publike v tolikšni meri, kot bi jo dejansko moral. Tako da moje vprašanje tukaj je predvsem to, kako se bo ta nov pravopis posredoval šolam in tudi vsem ostalim, ki se bomo tega morali »naučiti« na novo.

Tina Lengar Verovnik: Odgovora na to v resnici nimam, ker mi nimamo mandata za to, da potem pripravljamo tudi vse odvode, torej šolski pravopis, učbenike in tako dalje. Se pa zelo trudimo, da bi naredili pravopis, ki bo strokovnjakom, ki bodo te stvari delali, razumljiv in čim bolj vseobsegajoč ter da se bo čim manj izogibal mestom, ki se jim je prejšnji pravopis, ki jih ni zaznal ali pa se jim je včasih res namenoma izognil. Torej, to je mogoče en kratek stik v razumevanju tega, kaj mi počnemo. Mi delamo opis oziroma predpis, kakorkoli že kdo hoče to definirati, ki je potem lahko podlaga za ostalo. Naše delo je raziskovalno, hkrati pa se trudimo to ubesedovati v takšni formi, ki je za javnost sprejemljiva in razumljiva. Nimam drugega odgovora na to.

Bi pa se navezala na prejšnje vprašanje. Ne vem, kje bi lahko začeli odmotavati ta klopčič. Eva je že prej dala zelo dobro idejo ločitve pouka književnosti od jezika. Imam svežo izkušnjo, hčerko, ki je letos delala maturo. Ne bom sploh šla v to temo, ampak jezikovnega pouka so imeli vsa leta izjemno malo, na koncu so sredi marca sami prosili profesorja, ali se lahko, prosim, še malce učijo jezika. In ja, poznajo pravopis, vedo, kaj je slovar, kaj je slovnica, ampak samo na deklarativni ravni. Saj se naučijo geselski članek in tako naprej, ampak to ni to. Pri študentih, ki

jih mi dobimo na fakulteto, pogosto niti osnova ni dovolj dobra za to, da jih lahko pripeljemo do večjih piscev, pa nekateri to vseeno postanejo. Imam študente, ki so na začetku, v prvem letniku, res slabi. Ampak jih ne etiketiram, ne obupam nad njimi, ker iz takih lahko po petih letih postanejo izvrstni pisci. Ne vsi, lahko pa. Stvar je v motivaciji. In veliko takih študentov in študentk bi lahko ujeli, če bi na fakultetah imeli tudi specializirane predmete za študente in študentke, ko so dovolj motivirani, da bodo to sploh počeli. To je že bila ideja in ministrstvo nam je pred leti celo plačalo za študijo, ki smo jo s kolegicami opravile ter pregledale študijske programe. Kolegica Nataša Logar je ravno zdaj končala visokošolski učbenik znanstvenega pisanja, recimo, ki nam je krvavo potreben, in mislim, da bo do neke mere pokril to liso; za tiste, ki so zainteresirani, seveda. Ampak ni prepozno na fakultetni ravni, absolutno ne, samo pravega interesa pa v resnici ni. Ministrstvo je svojo domačo nalogo opravilo, kljukica je bila narejena, me smo pregledale in to je ostalo nekje v predalu. Ravno včeraj sem gledala, niti dosegljiva ni več ta študija nikjer, kar umaknili so jo, ko so prenavljali svojo spletno stran.

Iz publike: Jaz bi mogoče samo še nekaj, da osvežim, glede te velike in male začetnice, ker nisem prej znala. Meni je bilo, ko sem izvedela, da se bodo vsa zemljepisna imena pisala z veliko, kot da bi mi nož zarili v srce. Oh, grozno. Potem pa sem imela izkušnjo s temi petstotimi ljudmi, izobraženimi, kot sem rekla. Vemo, od kdaj obstaja to pravilo, trg, mesto, vas, selo, vse z malo. Ampak, koliko procentov mislite, da jih je to vedelo? In to je funkcionalna pismenost, da znaš napisati naslov, na primer nekomu, ki živi v Stari Cerkvi pri Kočevju, da veš, da pišemo Stara Cerkev z veliko. Res majhen procent jih to ve. Doktorjev, magistrrov znanosti, ne po bolonjskem sistemu. In ko sem jaz to videla, sem dojela, zakaj sprememba, zakaj se gre v poenostavljanje tega. Ker ne moremo reči, devetdeset procentov Slovencev je zabitih kot šterne, saj si niti tega pravila ne morejo zapomniti. Ampak vidiš, da je nekaj narobe s tem pravilom, da ga izobraženi ljudje ne moremo ponotranjiti.

Matejka Grgič: Kar nekaj rok je bilo dvignjenih, tako da predlagam, da začnemo pri tisti, ki je prva dvignila roko, in potem pridemo lepo do konca, po vrsti. Ja, prosim.

Erika Kržišnik: Jaz sem hotela vprašati, ali se delajo kakšne velike spremembe. Razen te z veliko začetnico, jaz upam, da ne, saj v 21. stoletju v temelju spreminjati pravopis, ki ga imamo od konca 19. stoletja, ne bi imelo nobenega smisla. To, ali bo pri krajevnih imenih vse z veliko ali ne, to je nekaj bolj ali manj dogovornega. Ne smemo delati znanosti iz tistega, kar je samo konvencija. Moramo znati povedati, da je nekaj bolje, da je šlo toliko in toliko stvari narobe zaradi tega in tega, in mislim, da nihče ne bo imel nič proti in si bomo spremenjeni, poenostavljeni dogovor zapomnili. Drugače pa upam, da ne delamo kakšnih velikih revolucij. Ker se spomnim revolucije v hrvaškem – ali srbskem – pravopisu, ko so šli iz sintaktične vejice, kot jo imamo mi ali pa Nemci, na semantično, ki je taka ali vsaj podobna kot v angleščini. To pa je ogromna razlika. Mene recimo pri branju v angleščini moti nepostavljanje vejic. Ne moti v smislu, da bi lektorirala in postavila vejico; moti me v smislu zadrževanja hitrosti branja.

Milica Antič Gaber: Razumevanja tudi.

Erika Kržišnik: No, saj, branja z razumevanjem, seveda. Tako da upam, da v tem smislu ne bo kakšne velike revolucije.

Tina Lengar Verovnik: Ne, saj mislim, da lahko kolegice in kolegi, ki zelo pridno sodelujejo v javni razpravi, povedo, kakšen je njihov vtis. Ker jaz včasih s te notranje pozicije res ne vem, ali je to za koga revolucija. Ampak – se vam zdi, da so stvari revolucionarne?

Matejka Grgič: No, mogoče bi detajle vseeno pustili za naslednjo okroglo mizo, bi pa dala besedo še razpravljalcema, ki sta se prijavila, Vojko najprej in potem še Eva, bi prosila.

Vojko Gorjanc: Jaz se s tabo [Tino Lengar Verovnik] ne strinjam glede tega, da še ni prepozno. Jaz mislim, da je na univerzi že prepozno. Da bi pravzaprav morali na univerzo študenti prihajati že opismenjeni; se pravi, vsakdo, ki konča maturo, bi moral biti preprosto pismen.

Tina Lengar Verovnik: Ja, v idealnem svetu, se strinjam.

Vojko Gorjanc: Ja, v idealnem svetu, ampak ta idealen svet je bil prisoten, to ni tako zelo idealiziran svet. Jaz sem pravopis na Univerzi v Ljubljani začel učiti pred več kot 25 leti. Pred 25 leti praktično ni bilo treba učiti pravopisa, ali pa minimalno. Jaz s študenti nisem rabil več kot en mesec za to, da je bil pravopis obvladan, zato ker so prišli na univerzo pismeni.

Erika Kržišnik: V tem smislu pismeni.

Vojko Gorjanc: V tem smislu pismeni, pravopisno pismeni, pustimo ostalo. Ampak zdaj, ko pogledamo zdajšnje generacije, zdaj moramo imeti cel predmet v enem letniku, pa še skoraj ni dovolj za to, da opismenimo študente, ki zdaj prihajajo. Se pravi, v 25 letih se nam je zgodila ena ogromna sprememba, od tega, da je bilo treba pravopis majčkeno zmodelirati, do tega, da se ga je treba naučiti praktično od začetka. Torej, nekaj je hudo narobe. Ne vem točno, kaj, kje in zakaj, ampak je narobe. Kar pa se tiče tega vprašanja, kako bomo zdaj naslavljali javnost, jaz misli, da imamo tukaj ogromno prostora. Slovenski prostor v celoti še vedno zelo slabo komunicira znanost. Imamo posameznike, ki so izjemno dobri, kot je Sašo Dolenc, ki se izjemno angažirajo v tem smislu, da znajo in poskušajo naslavljati komuniciranje znanosti, mi pa se razmisleka, kako skomunicirati tako, da bodo ljudje razumeli, še niti lotili nismo dobro.

Matejka Grgič: Prosim.

Eva Vrtačič: Jaz sem hotela samo z zunanje pozicije pripomniti, da že obstajajo dokaj dobra digitalna orodja, ki lahko malenkosti, ki bi se jih bilo treba »napiflati«, uredijo namesto nas. Mislim, da je ključna neka druga vrsta pismenosti. Ta je bolj povezana s tem, kako artikulirati argument ali misel, kako se suvereno gibati znotraj besedja, kompleksnejših stavčnih struktur in podobno. Medtem ko nam na vprašanje, ali tista vejica tam je ali je ni, lahko dokaj natančno odgovori umetna inteligenca. Tudi če ni stoddostno zanesljiva.

Vojko Gorjanc: Jaz mislim, da so stvari zelo povezane. Jaz, ko dobim študentski izdelek, vidim, da študent nima težav samo z enim ali dru-

gim. Se pravi, tisti, ki je dober, je dober v celoti, tisti, ki je slab, je v resnici slab v celoti. Tako da, ni preprosto tako, da bi zdaj vejice popravila umetna inteligenca, ki jih dejansko lahko sorazmerno enostavno, ampak s tem skoraj nič ne rešimo.

Eva Vrtačič: Večinoma se strinjam z vami, sem pa tudi na ravni literature kdaj prišla do kakšne prazne forme, ki je bila izbrušena do zadnje vejice.

Vojko Gorjanc: Ampak tako generalno ...

Eva Vrtačič: Generalno pa drži.

Matejka Grgič: Mislim, da je bila najbolj na udaru ravno naša predstavnicica pravopisne komisije, zato se mi zdi spodobno, da ji damo čisto na koncu še besedo, da ona zaključi to našo današnjo okroglo mizo. Potem se debata nadaljuje, nadaljuje pa se zunaj, na sončni terasi. Tako da, prosim, Tina.

Tina Lengar Verovnik: Zdaj na koncu se bom raje oglasila kot pedagoginja, če smem. Ja, res se mi zdi, da bi bilo treba stvari drugače postaviti že čisto na začetku šolske vertikale, kot tako radi učeno rečemo. Ampak ne vem, kako. Zato se navadno tudi raje ne oglašam prav veliko in nočem biti prav zelo glasna, ker si ne predstavljam, kako. Dejstvo pa je, da so se stvari spremenile, kot pravi Vojko, zelo spremenile, in da si samo želim, da bi se ta mastodont, ki mu rečemo šolstvo, začel premikati v to smer, da bi spet dobivali študente, ki bi bili pismeni. Pa res ne gre samo za pravopis, še zdaleč ne. Tudi oblikoslovje, besedišče, skladnja so velik problem, razgledanost po besedilnih žanrih, vse to. Študentom in študentkam, ki pridejo na novinarstvo, recimo sploh ni nerodno priznati, da ne berejo časopisov in revij. Oni so se prišli zdaj tega šele naučit. Ampak hočem reči, da v trenutni situaciji, ko pač imamo, kar imamo, ni prepozno, če se na fakulteti vseeno posvečamo tem motiviranim študentom, ki so že s tem, da so se vpisali na nek program, izkazali motivacijo za to, da v akademskem okolju, pa če ne drugače, vsaj do magisterija, delujejo tudi kot tvorci besedil. Ker morajo, saj ne

gre drugače, to so primorani početi, pa včasih niso opremljeni za to in jim je hudo. In če jim niti zdaj ne pridemo nasproti, potem smo jih pa res na celotni vertikali pustili na cedilu.

Matejka Grgič: Mislim, da smo z optimizmom začeli in da je dobro, če z optimizmom tudi zaključimo tole današnjo okroglo mizo. Predvsem v tem smislu motivacije, izboljšav, ki so vedno mogoče, pa pogovora, tudi v smislu jezikovne in družbene konvencije ter nekega človeškega, kulturnega sobivanja na vseh ravneh. Tako da bi se vam najlepše zahvalila, da ste bili danes z nami – vsem našim štirim gostjam in gostu: najlepša hvala, da ste sprejeli naše vabilo.

33. evropska poletna šola logike, jezika in informatike ESSLLI 2022

Mojca Brglez

Filozofska fakulteta, Univerza v Ljubljani

1 Uvod

Med 8. in 19. avgustom 2022 je na irski National University of Ireland Galway (v času pisanja že preimenovana v University of Galway) potekala poletna šola ESSLLI¹ oziroma *European Summer School of Logic, Language and Information*, ki pod okriljem evropskega združenja FoLLI² vsako leto poteka v drugem evropskem mestu. Poletna šola, ki je bila letos že 33. zapovrstjo, interdisciplinarno povezuje znanja in veščine iz treh disciplin: logike, jezika in računalništva. Za organizacijo letošnjega dogodka je poskrbel organizacijski odbor pod vodstvom Johna McCraeja (Data Science Institute, National University of Ireland Galway), o programu pa je odločal programski odbor pod vodstvom Larryja Mossa (Indiana University Bloomington).

Poletna šola je namenjena magistrskim in doktorskim študentom ter mlajšim raziskovalcem s področij jezikoslovja, logike, računalništva in informatike. Obiskali sva jo dve doktorski študentki iz Slovenije, Magdalena Gapsa (FF UL) in Mojca Brglez (FF UL).

1 <https://2022.esslli.eu/>

2 <http://www.folli.info/>

Brglez, M.: 33. evropska poletna šola logike, jezika in informatike ESSLLI 2022. Slovenščina 2.0, 10(1): 126–130.

1.19 Recenzija, prikaz knjige, kritika / Review, book review, critique

DOI: <https://doi.org/10.4312/slo2.0.2022.1.126-130>

<https://creativecommons.org/licenses/by-sa/4.0/>



2 Program

Kot vsako leto so bili tudi letos na voljo različni tečajji, razvrščeni v tri interdisciplinarne sklope: logika in jezik, logika in računalništvo, jezik in računalništvo. Vsega skupaj je bilo na voljo 7 temeljnih, 16 osnovnih in 18 naprednih tečajev ter 4 celotedenske delavnice, ki so vključevale predstavitve raziskav in člankov izbrane ožje tematike. Vsak dan je potekala tudi študentska sekcija, v okviru katere so študentke in študenti predstavljali lastno raziskovalno delo. Poleg predavanj so organizatorji priredili še tradicionalno tekmo v (galskem) nogometu, piknik in irski zabavni večer ter štiri večerna predavanja vabljениh predavateljev.

2.1 Tečajji

Tečajji na treh težavnostnih stopnjah so bili tematsko razdeljeni v tri skupine, in sicer LoLa (logika in jezik), LaCo (jezik in računalništvo) ter LoCo (logika in računalništvo). Šola je bila organizirana tako, da je v prvem tednu posamezni tečaj potekal vsak dan ob istem terminu, v drugem tednu pa so bile na voljo nove vsebine oziroma tečajji. Tečajji vseh treh skupin so tako vsakodnevno potekali sočasno v štirih različnih terminih, velikokrat se je bilo zato treba odločiti za enega od dveh tečajev znotraj iste področne skupine.

Za jezikoslovce, predvsem tiste brez predhodne formalno semantične in matematične podlage, so bili najprimernejši in tudi najbolj uporabni tečajji iz nabora skupine jezik in računalništvo, ki je ponujala dva temeljna, šest osnovnih in šest naprednih tečajev. V prvem tednu je tako potekal temeljni tečaj *Hands-on Distributional Semantics for Linguistics using R*, na katerem sta predavateljici Stephanie Evert in Gabriella Lapesa predstavili osnove pridobivanja vektorskih reprezentacij besed na podlagi distribucijske semantike (tj. na podlagi pogostosti pojavitev ali sopojavitvev) ter praktično uporabo vektorjev z vajami v programskem jeziku R. Drugi temeljni tečaj *A Linguist's Guide to Neural Networks*, ki ga je izvedel Tim Van de Cruys, je bil odličen uvod za udeležence z manj računalniškega predznanja, saj je orisal osnove nevronske omrežij, vektorskih vložitev in jezikovnih modelov ter njihovo uporabo pri globokem učenju, ki dandanes prednjači kot ena glavnih metod za računalniško obdelavo naravnih jezikov.

Med tečaji na osnovni ravni je v prvem tednu potekal tečaj *Cognitive and Computational Models of Abstractness*, kjer sta Diego Frassinelli in Sabine Schulte im Walde predstavila koncepta konkretnosti in abstraktnosti besed/konceptov, jih navezala na druge termine, kot so specifičnost, splošnost, predstavljivost, predstavila različne obstoječe baze človeških in polsamodejno pridobljenih ocen abstraktnosti ter ponudila praktične vaje v programskem jeziku R. Na koncu sta predstavila še različne možnosti uporabe ocen abstraktnosti pri drugih nalogah (npr. za samodejno prepoznavanje prenesenega jezika ali slik). Množično obiskan je bil tečaj *Argument Mining between NLP and Social Sciences*, ki sta ga izvedli Gabriella Lapesa in Eva Maria Vecchi. Udeležencem so bile predstavljene osnove teorij logike in argumentacije, odmevne raziskave s področja rudarjenja argumentov, pa tudi tveganje pristranskosti pri samodejnih metodah in umetni inteligenci.

V drugem tednu je navdušil osnovni tečaj o gradnji in hrambi jezikovnih virov *Creating and Maintaining High-quality Language Resources*, ki ga je izvajal Valerio Basile. Predstavil je različne vrste in formate jezikovnih virov, načine izvedbe človeškega označevanja virov, tehnike za poenotenje ocen in metode za izračunavanje strinjanja med ocenjevalci ipd. Zadnji dan je predstavil še izjemno zanimiv in razširajoč pogled, imenovan »perspektivizem«, ki namesto samodejnega privzemanja večinskega mnenja zagovarja ohranjanje izvirnih ocen posamičnih ocenjevalcev pri označevanju subjektivnih fenomenov (kot je na primer označevanje sentimenta), kar nudi vpogled v mnenje manjšin in posameznikov ter do neke mere razkrije subjektivno naravo ocenjevanja.

Večina naprednih tečajev je bila bolj logično- in formalnosemantično naravnanih. Med njimi je bil v prvem tednu na voljo tečaj *Computational Models of Grounding in Dialogue*, pri katerem so se udeleženci seznanili s formalizacijo izražanja skupne podlage oziroma razumevanja v dialogu. V enem od drugih naprednih tečajev, *From Minimal(ist) Formalizations to Parsing: Pros and Cons of a Symbolic Approach in a Deep-Learning Era*, sta Cristiano Chesi in Gregory M Kobele udeležencem predstavila pristope k analizi jezikovnih fenomenov skozi slovnično strukturo (s poudarkom na Chomskyjevi minimalistični slovni) ter rezultate analize primerjala z napovedmi sodobnejših strojnih metod, ki temeljijo na globokem učenju.

2.2 Delavnice

Na poletni šoli se je bilo mogoče udeležiti tudi delavnic, ki so potekale vzporedno s tečajji. Format delavnic je bil zasnovan tako, da je organizator delavnice predlagal neko izbrano temo, na delavnico pa povabil raziskovalce, ki se s to temo ukvarjajo, da predstavijo svoje prispevke ter jih izpostavijo morebitnim drugim mnenjem, pripombam in idejam za nadaljnje raziskovanje. Po ena delavnica od šestih je bila izvajana v okviru skupin LoCo in LaLo, kar štiri delavnice pa so bile umeščene v skupino LaCo. Med temi sta Timothée Bernard in Grégoire Winterstein izvedla zanimivo delavnico *Bridges and gaps between formal and computational linguistics*,³ ki je naslovila začetno tesno povezanost ter kasnejša razhajanja med pristopoma formalnega in računalniškega jezikoslovja. Na delavnici *Natural Logic Meets Machine Learning III (NALOMA'22)*,⁴ ki sta jo predlagala Stergios Chatzikyriakidis in Aikaterini-Lida Kalouli, so raziskovalci predstavljali prispevke s področja razumevanja naravnega jezika, na primer analize sentimenta, strojnega prevajanja, povzemanja ipd. Poudarek je bil na uporabi hibridnih modelov, ki kombinirajo sodobno globoko učenje s tradicionalnejšimi simbolnimi pristopi. Tretja delavnica, *The Semantics of Imagination*,⁵ ki sta jo vodila Kristina Liefke in Justin D'Ambrosio, se je osredotočila na zelo nišno tematiko, in sicer globokosemantično analizo angleškega glagola 'imagine' (domišljati si, predstavljati si). Michaël Moortgat in Gijs Wijnholds pa sta na delavnici *End-to-End Compositional Models of Vector-Based Semantics*⁶ udeležence povabila k oddaji prispevkov, ki obravnavajo ali evalvirajo kompozicionalne modele jezika.

2.3 Vabljen predavanja

Na poletno šolo so organizatorji povabili še štiri gostujoče predavateljice, ki so po dnevni tečajih in delavnicah izvedli večerna predavanja. V prvem tednu sta občinstvo nagovorila Ed Curry (Data Science Institute, NUI Galway) in Henriëtte de Swart (Utrecht University); prvi je govoril o pomembnosti skupnih »podatkovnih prostorov« za deljenje podatkov,

3 <https://gdr-lift.loria.fr/bridges-and-gaps-workshop/>

4 <https://sites.google.com/view/naloma22/home?pli=1>

5 <https://www.ruhr-uni-bochum.de/phil-inf/events/index.html.en>

6 <https://compositionalcalculus.sites.uu.nl/workshop/>

druga pa o čezjezikovni semantiki, ki jo lahko odkrivamo prek vzporednih korpusov. V drugem tednu sta predavala Bonnie Webber (University of Edinburgh) in Samson Abramsky (University College London). Prvo predavanje je naslovlilo pristope k odpravljanju nekonsistentnih oznak v korpusih, drugi predavatelj pa je udeležence vpeljal v principe kvantne mehanike in ovrednotil njen pomen v sodobnem računalništvu.

3 Sklep

Poletna šola ESSLLI je neprecenljivo vozlišče za mlade, saj študentom omogoča spoznavanje bodisi bolj splošnih tematik s področja jezika, logike in računalništva bodisi poglobljanje znanja na bolj nišnih področjih. Poleg akumulacije znanja je morda še večja dodana vrednost šole, da študentom z vseh koncev sveta, na različnih ravneh izobraževalne ali raziskovalne poti in z različnimi strokovnimi ozadji, omogoča mreženje, izmenjavo mnenj ter pridobivanje novih idej. To je posebej dragoceno v obdobju pandemije COVID-19, ko so se raziskovalni in izobraževalni dogodki ali dejavnosti v veliki meri preselili v virtualni svet, ki pa žal omogoča le posredno navezovanje stikov in je nezadovoljiv nadomestek diskurza v živo.

Letošnje poletne šole se je udeležilo 207 študentov, 40 udeležencev iz akademskega sveta, 5 iz gospodarstva ter kar 74 predavateljev. Prvič v zgodovini ESSLLI je največji del udeležencev prihajal iz ZDA, kar je presenetljivo, glede na to, da v ZDA prav tako vsako leto poteka ekvivalentna šola NASSLLI. Več informacij o letošnji poletni šoli, programu in tečajih je trenutno še vedno na voljo na spletni strani.⁷ Znano je tudi že, da se poletna šola v letu 2023 seli v Ljubljano, med 31. julijem in 11. avgustom bo potekala na Fakulteti za računalništvo in informatiko Univerze v Ljubljani.

⁷ <https://2022.esslli.eu/>

Konferenca Jezikovne tehnologije in digitalna humanistika 2022

David BORDON

Filozofska fakulteta, Univerza v Ljubljani

1 O konferenci

Septembra 2022 je v prostorih Fakultete za družbene vede Univerze v Ljubljani potekala konferenca Jezikovne tehnologije in digitalna humanistika (JTDH), ki jo je priredilo Slovensko društvo za jezikovne tehnologije (SDJT) v soorganizaciji s Centrom za jezikovne vire in tehnologije Univerze v Ljubljani (CJVT), Inštitutom za novejšo zgodovino (INZ) ter raziskovalnima infrastrukturalnima CLARIN.SI¹ in DARIAH-SI².

Po spletni izvedbi leta 2020 je konferenca, ki se sicer odvija vsaki dve leti, letos ponovno potekala v živo. SDJT organizira konferenco že od leta 1998; do leta 2014 sicer pod drugim imenom – konferenca Jezikovne tehnologije, leta 2016 pa je izvedlo tematsko širitev še na polje digitalne humanistike. Splošna tematska področja konference so jezikovne tehnologije, digitalno jezikoslovje in digitalna humanistika.

Konferenca je mednarodna – od 120 avtorjev prispevkov je bila skoraj tretjina tujih – večina prispevkov pa je bila predstavljena v angleščini, ki je poleg slovenščine tudi uradni jezik konference. V sklopu programa je bilo mogoče prisluhniti študentski sekciji v slovenščini in angleščini, dvema slovenskima in trem angleškimi rednim sekcijam, predstavitvi plakatov tako v angleščini kot v slovenščini in dvema vabljenima predavanjema.

1 <https://www.clarin.si/info/o-projektu/>

2 <http://www.dariah.si/>

Bordon, D.: Konferenca Jezikovne tehnologije in digitalna humanistika 2022. Slovenščina 2.0, 10(1): 131–135.

1.19 Recenzija, prikaz knjige, kritika / Review, book review, critique

DOI: <https://doi.org/10.4312/slo2.0.2022.1.131-135>

<https://creativecommons.org/licenses/by-sa/4.0/>



Predstavitve so bile posnete in so na voljo na spletni strani konference JTDH 2022.³

2 Predkonferenčne delavnice

Konferenčno dogajanje se je začelo že dan pred uradnim odprtjem. Na Inštitutu za novejšo zgodovino (INZ)⁴ so se odvijale praktične delavnice uporabe različnih orodij, ki so namenjena raziskovalkam in raziskovalcem. Ajda Pretnar Žagar je predstavila tematsko modeliranje parlamentarnih razprav na korpusu ParlaMint pred in med epidemijo covid-19 z uporabo orodja Orange. Na drugi delavnici sta Jakob Lenardič in Kristina Pahor de Maiti predstavila raziskovalno infrastrukturo in orodja CLARIN.SI, ki nudi podporo pri ustvarjanju, obdelavi, arhiviranju in ponovni uporabi jezikovnih podatkov.

3 Vabljeni predavanja

Letos sta kot vabljeni predavatelja na konferenci gostovala Eetu Mäkelä in Benoît Sagot. Mäkelä, izredni profesor na Univerzi v Helsinkih in Univerzi Aalto ter tehnološki direktor infrastrukture DARIAH-FI⁵, se ukvarja z interakcijo med računalništvom in humanistiko, obenem pa vodi raziskovalno skupino, ki si prizadeva ugotoviti tehnološke, procesne in teoretične temelje uspešnih računalniških raziskav v humanistiki in družboslovju. V predavanju *Designing computational systems to support humanities and social sciences research* je predstavil izsledke raziskav prej omenjene raziskovalne skupine. Med glavnimi dognanji njegove predstavitve gre izpostaviti, da sodelovanje med humanisti in računalničarji pogosto ne obrodi zelenih sadov, kar je posledica neskladij med disciplinami, denimo različnih tradicij in znanstvenih pristopov, ter različno koncepcijo tega, kateri podatki so pomembni in relevantni za obdelavo. Poudaril je, da je pred začetkom interdisciplinarnih projektov pomembno vse te vidike vzeti v poštev in jih prilagoditi simbiotičnemu sodelovanju.

Drugi konferenčni dan je predaval Benoît Sagot, vodja programske skupine ALMAAnCH iz pariškega raziskovalnega centra Inria (Institut

3 <https://www.sdjt.si/wp/dogodki/konference/jtdh-2022/>

4 <https://www.inz.si/>

5 <https://www.dariah.fi/>

national de recherche en sciences et technologies du numérique).⁶ Je specialist na področju procesiranja naravnega jezika in deluje na mnogoterih področjih digitalnega jezikoslovja. V predavanju *Large-scale language models: challenges and perspective* je predstavil nastanek večjezičnega korpusa OSCAR,⁷ predvsem z vidika prečiščevanja ogromnih količin podatkov, na katerih sloni korpus (podatke pridobivajo iz *dumpov* ameriškega združenja Common Crawl). Govoril je tudi o jezikovnem modelu za francoščino CamemBERT, prvem modelu take velikosti za jezik, ki ni angleščina, o težavah, s katerimi so se srečevali v teku projekta, in načinih, kako so jih premostili. Pri CamemBERTu je vredno izpostaviti dejstvo, da 4 GB (dovolj raznolikih) podatkov zado- stuje za doseganje *state-of-the-art* nivoja kakovosti.

4 Vzporedne sekcije

Jedrni del konference JTDH 2022 so bile zagotovo vzporedne sekcije v slovenskem in angleškem jeziku. Format vzporednih sekcij se je na konferenci prvič izvedel leta 2016, po širitvi na področje digitalne humanistike. V zadnjih dveh izvedbah pred letošnjo, leta 2018 in 2020, pa se je program delil na tematske sklope. Na letošnji konferenci smo lahko v osrednjem delu poslušali predstavitve 70 avtorjev, od tega 12 prispevkov v slovenščini in kar 17 prispevkov v angleščini. Zaradi števila in jezikovne narave prijavljenih prispevkov se je sistem vzporednih sekcij pokazal kot smiseln, razporeditev udeležencev pa zaradi visoke mednarodne udeležbe precej homogena.

Na področju korpusnega jezikoslovja smo lahko spoznali tri nove kor-puse – korpus Trendi⁸, prvi spremljevalni korpus za slovenščino, ki upo-rabnikom nudi podatke o aktualni jezikovni rabi in omogoča diahrone je-zikovne analize, korpus študentskih besedil KOŠ, namenjen pridobivanju empiričnih podatkov o pisni jezikovni zmožnosti študentske populacije, in hrvaški korpus DirKorp, specializiran za govorna dejanja. Z uporabni-škega vidika je bila predstavljena raba *Kolokacijskega slovarja sodobne slovenščine* (KSSS)⁹ pri prevajanju kolokacij iz angleščine v slovenščino. Analiza je bila izvedena na vzorcu dodiplomskih študentov Oddelka za

6 <https://www.inria.fr/fr>

7 <https://oscar-project.org/>

8 <https://sled.ijs.si/korpus-trendi/>

9 <https://viri.cjvt.si/kolokacije/slv/#>

prevajalstvo FF UL – izsledki kažejo, da je sposobnost uporabe slovarja sorazmerna s kakovostjo rešitev, do katerih uporabnik lahko pride, hkrati pa sama raba jezikovnih virov ni zagotovilo, da bo prevodna rešitev ustrezna. Obratno, raba jezikovnih virov ni zagotovilo, da bo prevodna rešitev ustrezna. V debati je bilo izpostavljeno, da slovenski visokošolski univerzitetni programi študente dobro učijo, kako uporabljati jezikovna orodja.

Pri govornih tehnologijah so bili med drugim predstavljeni najnovejši napredki pri samodejni slovenski grafemsko-fonemski pretvorbi ter projekt poravnave zvočnih posnetkov s transkripcijo narečnega govora in petja. Izpostaviti velja prispevek, vezan na izgradnjo govorne baze Artur. Avtorji so se posvetili primerom dobre prakse pri poenotenju metapodatkov med združevanjem različnih govornih korpusov in predlagali načine, kako se v bodoče izogniti neskladjem med metapodatki.

Številni prispevki so se posvečali diskurzu, med zanimivejšimi so bili prispevki o sovražnem in grobem besedišču v odzivnem *Slovarju sopomenk sodobne slovenščine* (SSSS)¹⁰ ter dva, vezana na parlamentarno okolje; prvi se je posvetil pregledu mednarodnih raziskav parlamentarnega diskurza v zadnjih desetih letih, drugi pa je bil vezan na populistični diskurz v slovenskem parlamentu med letoma 1992 in 2018.

Velikega napredka je bilo deležno področje označevanja – na konferenci so bili predstavljeni primeri dobre prakse in optimalne rešitve, uporabljene pri projektu oblikoskladenjskega označevanja korpusa SentiCoref¹¹, ki bo vključen v nov učni korpus za slovenščino (trenutni ssj500k¹²), v sklopu projekta Razvoj slovenščine v digitalnem okolju (RSDO)¹³. Poleg tega so se v okviru projekta RSDO izvajale aktivnosti v povezavi s shemo Universal Dependencies (UD)¹⁴ – raziskovalci so obstoječo infrastrukturo nadgradili in ustvarili dokumentacijo označevalnih smernic UD za slovenščino.

Pri strojnem prevajanju je izstopal predvsem prispevek, ki je predstavil človeško evalvacijo prevodnih rešitev strojnega prevajalnika za jezikovno kombinacijo slovenščina-angleščina, ki nastaja na projektu RSDO. Na področju terminologije pa so bili prispevki vezani predvsem na modele samodejnega luščenja terminov.

10 <https://viri.cjvt.si/sopomenke/slv/>

11 <https://www.clarin.si/repository/xmlui/handle/11356/1285>

12 <https://www.clarin.si/repository/xmlui/handle/11356/1434>

13 <https://slovenscina.eu/>

14 <https://universaldependencies.org/>

5 Študentska sekcija in predstavitve plakatov

Velika dodana vrednost konference JTDH je samostojna študentska sekcija, uvedena že leta 2016, in sekcija s plakati, ki obstaja od leta 2018. Mladim raziskovalkam in raziskovalcem ter študentkam in študentom je tako omogočeno, da se lahko (brezplačno) prijavijo na konferenco – če je njihov prispevek sprejet, ga lahko predstavijo v eni izmed omenjenih sekcij in objavijo v zborniku. Za mnoge mlade raziskovalke in raziskovalce je objava v zborniku JTDH prva resnejša objava znanstvenega prispevka, kar predstavlja velik doprinos, saj so jim na začetku poklicne poti tovrstne možnosti običajno povsem (predvsem finančno) nedostopne.

Samostojna študentska sekcija je po izvedbi ekvivalentna ostalim jedrnim sekcijam, letos so denimo predstavitve trajale 10 minut, sledila pa so vprašanja občinstva. Dinamika sekcije s plakati je nekoliko drugačna, saj se odvija v preddverju, avtorji prispevkov pa so na voljo za predstavitev, vprašanja ali pogovor.

6 Občni zbor SDJT – predstavitev vmesnih rezultatov projekta RSDO in zaključek

Po formalnemu zaključku konference je sledil občni zbor SDJT in predstavitev orodij, ki so nastala v sklopu projekta RSDO, ki je v času pisanja poročila v zaključni fazi. Predstavitve so si sledile po delovnih sklopih, v katerih so bili izpostavljeni strojni označevalnik, metakorpus slovenskega jezika, terminološki portal, modeli strojnega prevajalnika, orodje za prepoznavanje imenskih entitet in koreferenčnosti, ekstrakcijo povezav, baza znanja, orodja za povzemanje besedil ter orodje za semantične premike in diahrone analize.

Zelo plodna izvedba konference – po štirih letih ponovno v živo – je pokazala, da se konferenca JTDH vedno bolj uveljavlja na mednarodnem parketu, obenem pa ohranja svojo dostopnost mlajši generaciji. Tematska raznolikost in bogatost programa ji utrjujejo položaj kot eni pomembnejših pri nas. Čestitke za odlično izvedbo; še na mnoga leta.

Simpozij Obdobja 41: Na stičišču svetov: slovenščina kot drugi in tuji jezik

Ina POTEKO

Filozofska fakulteta, Univerza v Ljubljani

Po dveh letih izvajanja na daljavo je Simpozij Obdobja ponovno zaživel v prostorih Univerze v Ljubljani. Tokratni 41. simpozij v organizaciji Centra za slovenščino kot drugi in tuji jezik (CSDTJ)¹ ter Oddelka za slovenistiko Filozofske fakultete Univerze v Ljubljani je bil prvič v celoti namenjen slovenščini kot drugemu in tujemu jeziku pod naslovom *Na stičišču svetov: slovenščina kot drugi in tuji jezik*. Med 16. in 18. novembrom 2022 se je v devetih sklopih zvrstilo 33 predstavitev, vseh 37 prispevkov pa je dostopnih tudi v zborniku.²

Kot napoveduje že naslov simpozija, je bilo predstavljenih mnogo stičišč: med različnimi jeziki, obdobji, prostori in kulturami, predvsem pa med govorcji slovenščine kot drugega in tujega jezika ter tistimi, ki jo poučujejo. Če smo z drugih slovenističnih simpozijev, namenjenih jeziku, nekoliko bolj vajeni teoretiziranja, so bili prispevki na tem usmerjeni predvsem k praksi. O tem pričajo tudi metodološki pristopi, saj so bili v ospredju vprašalniki, intervjuji, analiziranje besedil govorcev SDTJ, pregled gradiva za poučevanje, predstavitev uspešnih pedagoških praks, nadgrajevanje obstoječih orodij in korpusov ter nastanek novih in primerjalne analize med jeziki.

1 <https://centerslo.si/>

2 Zbornik je prosto dostopen na povezavi: https://centerslo.si/wp-content/uploads/2022/11/Obdobja-41_zbornik.pdf.

Poteko, I.: Simpozij Obdobja 41: Na stičišču svetov: slovenščina kot drugi in tuji jezik. Slovenščina 2.0, 10(1): 136–140.

1.19 Recenzija, prikaz knjige, kritika / Review, book review, critique

DOI: <https://doi.org/10.4312/slo2.0.2022.1.136-140>

<https://creativecommons.org/licenses/by-sa/4.0/>



Simpozij so letos namesto s plenarnimi predavanji otvorili z okroglo mizo, naslovljeno *Slovenščina kot drugi in tuji jezik – resničnost in vizije*, v kateri so sodelovali dr. Mihaela Knez, dr. Stanka Lunder Verlič, prof. dr. Tanja Petrović, prof. dr. Marko Stabej in Magda Stražišar. Razpravljali so o stanju in načrtovanju na področju SDTJ, med drugim tudi o *Akcijskem načrtu za jezikovno izobraževanje* in o trenutno aktualni *Resoluciji o nacionalnem programu za jezikovno politiko*³.

V vabilu sta predsednici letošnjega simpozija dr. Ina Ferbežar in dr. Nataša Pirih Svetina zapisali, da je slovenščina kot drugi in tuji jezik »integralni del slovenske in slovenistične realnosti«. Ob sodobnih migracijskih procesih in trendu priseljevanja je razmišljati o slovenščini tudi kot o drugem in tujem jeziku še bolj ključno, kot je bilo do sedaj. Pogled na slovenščino s perspektive tistih, ki poučujejo tujce, in učečih se slovenščine kot drugega in tujega jezika ne odpira le vprašanj, neposredno povezanih s procesom poučevanja in učenja, ampak sproža tudi poglobljeno refleksijo o jeziku samem in obstoječih praksah slovenistične stroke ter hkrati opominja, da je pri jeziku v ospredju resnični uporabnik.

Marko Stabej je v svojem prispevku ugotavljal, kaj jezikovna variantnost pomeni za tiste, ki se slovenščine učijo kot drugega ali tujega jezika. Pri učenju tujega jezika ima govorec sprva omejene jezikovne zmožnosti, ki pa so pogosto omejene tudi na standardno različico jezika. Če že, varietete usvoji sčasoma, a pogosto predvsem na receptivni, ne pa tudi na tvorni ravni. Prav zato bi v slovenskem prostoru govorkam in govorcem SDTJ morali omogočiti tudi orodja in vire, ki bi bili koristni za učenje in rabo slovenščine nasploh – torej takšne, kot jo govorimo in pišemo vsi govorce in govorce (Stabej, 2022, str. 315–322).

V nekaterih prispevkih so izpostavili težave, s katerimi se soočajo tisti, ki se slovenščine učijo kot neprvega jezika. Med drugim so poudarili vpliv prvega jezika, ki na usvajanje novega lahko vpliva pozitivno, npr. prenos enakega ali podobnega besedišča iz hrvaščine v slovenščino, ali pa predstavlja prepreko. V povezavi s slednjim so v več prispevkih izpostavili glagolski vid, ki je za tiste, katerih prvi jezik ni slovanski, posebej zahtevno področje. Vredno je izpostaviti tudi prispevek Juna Mite, govorca japonščine kot prvega jezika, o trajanju zlogov v slovenščini, ki

³ <http://www.pisrs.si/Pis.web/pregledPredpisa?id=RESO123>

o jeziku razmišlja z nam mogoče nekoliko nenavadnega, a zelo zanimivega vidika (Mita, 2022, str. 239–250).

Govorcev slovenščine kot neprvega jezika je čedalje več, zato so prepričavanja obstoječih pedagoških praks na področju poučevanja SDTJ nujna. Tako se je na simpoziju zvrstilo več primerov dobrih praks, med drugim tudi, kako upoštevati generacijske razlike za uspešno poučevanje. Obenem je bilo izpostavljeno, da učenje jezika ne pomeni le učenja besedišča in slovnice, ampak tudi sociokulturnih elementov, pri čemer imajo posebno mesto tudi stereotipi, skriti v jeziku, in humor. V poučevanje SDTJ se lahko uspešno vključuje tudi literatura. Pri poučevanju slovenščine v tujini in na dvo- ali večjezičnih območjih je med drugim ključno, da obstaja prostor, kjer je mogoče govoriti slovensko, hkrati pa je bilo izpostavljeno, da ponekod poučevanje slovenščine ni uspešno zaradi pomanjkanja ustreznega kadra.

O tem, kdo sploh so uporabniki slovenščine na najnižji ravni, sta se spraševali Ina Ferbežar in Mateja Eniko, ki sta ugotovili, da ob neuspehu pri izpitu na vstopni ravni A1 po SEJO uporabnik ne izgubi le motivacije, ampak tudi pravice, ki izhajajo iz njegovega statusa (Ferbežar in Eniko, 2022, str. 99–108). Kot je v uvodniku zapisala Ina Ferbežar: »Čeprav je želja po vzajemnem razumevanju srž procesov jezikovnega sporazumevanja med ljudmi, srečevanja svetov niso vedno harmonična« (Ferbežar, 2022, str. 9). Kljub temu da so predstavljeni prispevki prikazali pretežno pozitivno sliko odnosa tako do učenja slovenščine kot do tistih, ki se slovenščine učijo kot neprvega jezika, pa v družbi še vedno o(b)stajajo prepreke, povezane s politiko, pričakovanji in zahtevami do govorcev SDTJ ter tudi s stigmatizacijo do določenih priseljencev, o čemer pa je bilo na simpoziju povedanega (pre)malo.

Poučevanje in učenje SDTJ predstavljata precejšen izziv, zato so novosti s področja korpusov, virov in orodij še kako dobrodošle. Špela Arhar Holdt, Iztok Kosem in Mojca Stritar Kučuk so predstavili novo prosto dostopno orodje, prilagojeno za slovenščino, ki temelji na švedskem programu Svala ter je pomembno za področje razvoja korpusov usvajanja jezika (Arhar Holdt idr., 2022, str. 23–30). Sledili sta predstavitvi pisnega korpusa za usvajanje slovenščine kot tujega jezika KOST (Stritar Kučuk, 2022, str. 323–334) in korpusa učbenikov za učenje slovenščine kot drugega in tujega jezika KUUS (Klemen idr., 2022, str.

165–174), na podlagi katerega je nastal tudi izhodiščni seznam jedrnega besedišča za stopnje SEJO A1–B1. V tem sklopu je bila predstavljena še spletna aplikacija Loris za pomoč govorcem slovenščine na območju jezikovnega stika med slovenščino in italijanščino (Popič, 2022, str. 281–290). Prav tako se še naprej razvija aplikacija Slovene Learning Online za samostojno učenje slovenščine (Alič in Jerman, 2022, str. 11–22).

Simpozij Obdobja je potekal v prijateljskem in sodelovalnem duhu, kar dokazujejo tudi mnogi prispevki, napisani in predstavljeni v soavtorstvu. Če si izposodim metaforo iz prispevka Elizabete Jenko z Univerze na Dunaju, ki je za poučevanje glagolskega vida uporabila sliko zahajajočega sonca, pri čemer rumena barva sonca simbolizira omejeno notranjost (dovršnost), oranžna pa večerno nebo neomejene širine (nedovršnost) (Jenko, 2022, str. 146), lahko zaključim, da je 41. simpozij Obdobja kot sonce sicer osvetlil temo slovenščine kot drugega in tujega jezika, o večernem nebu, ki ga obdaja, pa bo še treba razpredati. Upam, da prej kot na naslednjem tovrstnem simpoziju.

Literatura

- Alič, T., & Jerman, T. (2022). Slovenščina kjerkoli in kadarkoli: spletni tečaj za samostojno učenje slovenščine Slovene Learning Online. V N. Pirih Svetina in I. Ferbežar (ur.), *Na stičišču svetov: slovenščina kot drugi in tuji jezik. Obdobja 41* (str. 11–22). Ljubljana: Univerza v Ljubljani, Filozofska fakulteta. Pridobljeno s https://centerslo.si/wp-content/uploads/2022/11/Alic_Obdobja-41.pdf.
- Arhar Holdt, Š., Kosem, I., & Stritar Kučuk, M. (2022). Metode in orodja za lažjo pripravo korpusov usvajanja jezika. V N. Pirih Svetina in I. Ferbežar (ur.), *Na stičišču svetov: slovenščina kot drugi in tuji jezik. Obdobja 41* (str. 23–30). Ljubljana: Univerza v Ljubljani, Filozofska fakulteta. Pridobljeno s https://centerslo.si/wp-content/uploads/2022/11/Arhar-Holdt-et-al_Obdobja-41.pdf.
- Ferbežar, I. (2022). Uvodnik. V N. Pirih Svetina in I. Ferbežar (ur.), *Na stičišču svetov: slovenščina kot drugi in tuji jezik. Obdobja 41* (str. 9–10). Ljubljana: Univerza v Ljubljani, Filozofska fakulteta. Pridobljeno s https://centerslo.si/wp-content/uploads/2022/11/Uvodnik_Obdobja-41.pdf.
- Ferbežar, I., & Eniko, M. (2022). »Lah blatschem gotovina?«: jezikovni profil uporabnika slovenščine na najnižji ravni. V N. Pirih Svetina in I. Ferbežar

- (ur.), *Na stičišču svetov: slovenščina kot drugi in tuji jezik. Obdobja 41* (str. 99–108). Ljubljana: Univerza v Ljubljani, Filozofska fakulteta. Pridobljeno s https://centerslo.si/wp-content/uploads/2022/11/Ferbezar-Eniko_Obdobja-41.pdf.
- Jenko, E. (2022). S Piko Nogavičko po sledih glagolskega vida. V N. Pirih Svetina in I. Ferbežar (ur.), *Na stičišču svetov: slovenščina kot drugi in tuji jezik. Obdobja 41* (str. 143–153). Ljubljana: Univerza v Ljubljani, Filozofska fakulteta. Pridobljeno s https://centerslo.si/wp-content/uploads/2022/11/Jenko_Obdobja-41.pdf.
- Klemen, M., Arhar Holdt, Š., Pollak, S., Kosem, I., Huber, D., & Lutar, M. (2022). Korpus učbenikov za učenje slovenščine kot drugega in tujega jezika. V N. Pirih Svetina in I. Ferbežar (ur.), *Na stičišču svetov: slovenščina kot drugi in tuji jezik. Obdobja 41* (str. 165–174). Ljubljana: Univerza v Ljubljani, Filozofska fakulteta. Pridobljeno s https://centerslo.si/wp-content/uploads/2022/11/Klemen-et-al_Obdobja-41.pdf.
- Mita, J. (2022). Ljubljana, štirizložna? Razmišljanje o trajanju zloga v slovenščini z gledišča govorca japonščine. V N. Pirih Svetina in I. Ferbežar (ur.), *Na stičišču svetov: slovenščina kot drugi in tuji jezik. Obdobja 41* (str. 239–250). Ljubljana: Univerza v Ljubljani, Filozofska fakulteta. Pridobljeno s https://centerslo.si/wp-content/uploads/2022/11/Mita_Obdobja-41.pdf.
- Popič, D. (2022). Digitalna podpora slovenskemu jeziku v stiku. V N. Pirih Svetina in I. Ferbežar (ur.), *Na stičišču svetov: slovenščina kot drugi in tuji jezik. Obdobja 41* (str. 281–290). Ljubljana: Univerza v Ljubljani, Filozofska fakulteta. Pridobljeno s https://centerslo.si/wp-content/uploads/2022/11/Popic_Obdobja-41.pdf.
- Stabej, M. (2022). Jezikovna variantnost in slovenščina kot drugi in tuji jezik. V N. Pirih Svetina in I. Ferbežar (ur.), *Na stičišču svetov: slovenščina kot drugi in tuji jezik. Obdobja 41* (str. 315–322). Ljubljana: Univerza v Ljubljani, Filozofska fakulteta. Pridobljeno s https://centerslo.si/wp-content/uploads/2022/11/Stabej_Obdobja-41.pdf.
- Stritar Kučuk, M. (2022). KOST med korpusi usvajanja tujega jezika. V N. Pirih Svetina in I. Ferbežar (ur.), *Na stičišču svetov: slovenščina kot drugi in tuji jezik. Obdobja 41* (str. 323–334). Ljubljana: Univerza v Ljubljani, Filozofska fakulteta. Pridobljeno s https://centerslo.si/wp-content/uploads/2022/11/Stritar-Kucuk_Obdobja-41.pdf.