Our Words in a State of Emergency: Psychological–Linguistic Analysis of Utterances on the COVID-19 Situation in the Czech Republic

Dalibor Kučera

Abstract The study focuses on psychological–linguistic analysis of utterances provided by \( N = 2522 \) respondents aged 18–89 years in the period of March–May 2020, for the research of JUPSYCOR (Psychological Impacts of the Coronavirus Epidemic in the Czech Republic). The utterances relate to the interpretation of the state of emergency, the COVID-19 epidemic, and its subjectively perceived impacts. Simultaneously, the study examines the relationship between the analysed texts and the results of the SEHW (Scales of Emotional Habitual Subjective Well-being) questionnaire, which determines the valence of experienced emotions. The aim of the study is to analyse the lexical and morphological layers of the utterances, especially which specific words resonated in the individual questions, what is their emotional load, and which linguistic features of the texts may refer to the respondents’ positive/negative emotional response. One of the outputs based on the results of the quantitative analyses determines that the most distinctive words are connected to negative emotions and most frequently relate to social environment, anxiety, and inhibition. Furthermore, the study proves a positive correlation between a fear scale and a higher occurrence of future tense and use of emotionally negatively loaded words, especially in women. Numerous differences among the individual age and gender cohorts were also proved. The significance of the study lies predominantly in the combination of the linguistic and psychological levels of the analysis, in the utilization of two mutually complementary utterances, and in the presentation of new insights on how people use words when they face an unexpected and emotionally disturbing situation.

Keywords COVID-19 · Utterances · Text analysis · Emotions · Subjective experience

Introduction

The situation of the COVID-19 pandemic that occurred in the spring of 2020 significantly impacted the lives of people in many aspects of their everyday lives. The individual countries’ measures as they were striving to contain the epidemic differed both in timing and severity. As for the Czech Republic (EU), the reaction was relatively fast and strong. Between 10 and 16 March 2020, severe restrictions were adopted, especially restrictions on free movement, closure of schools, shops, restaurants, bans on sporting, cultural, and other activities. As of 19 March, all persons were obligated to wear a face mask in public. The borders were closed completely and Czech nationals staying abroad were predominantly repatriated. All these measures were highly progressive and radical within the context of the situation in Europe, although similar restrictions were subsequently adopted by numerous other countries.

The state of emergency had a major impact on the behaviour and emotional experience of people. Certain studies published so far show an extensive psychological impact, e.g. Twenge and Joiner (2020) demonstrated that the level of mental distress in the US population is three times higher than in 2019 or 2018 using the Kessler-6 scale. Wang et al. (2020) reported that 53.8% of all respondents in their Chinese study scored as medium to
serious psychological impact, out of whom 28.8% scored as medium to high seriousness and 8.8% scored as very serious anxiety level symptoms. The first studies from the Czech Republic also refer to an increased level of anxiety and fear (e.g. 40% of the respondents experienced the fear that they or their loved ones will fall ill with COVID-19 with serious symptoms; Rabušic, 2020). The emergency brought significant changes on the level of interpersonal contact as well. Personal meetings were substituted by phone calls and video conferencing, in the media as well as in day-to-day communication, such words as “face mask”, “lockdown”, and “quarantine”, very rare up to that point, resonated strongly. However, experiencing negative emotions did not necessarily influence only the explicit content of the communication (language content), but also language style (i.e. lexical and morphological layers of the communication; Chung & Pennebaker, 2007).

The present study aims to analyse selected parameters of communication in the form of thematic verbal utterances which might be significant in relation to the psychological and social aspects of the emergency state of COVID-19 epidemic in the Czech Republic. The basic research hypothesis postulates that an unexpected and emotionally disturbing situation will be reflected in the content and form of people’s utterances on the situation (for similar studies, see, e.g. Fiehler, 2002; Peters et al., 2009; Sun et al., 2019; Bernard et al., 2016). This hypothesis is important not only in terms of a possible description of emotional experiences of speakers at a specific time, but also in terms of verifying the potential of an analysis of psychological processes from a quantitative linguistic perspective, i.e. through natural language processing.

The hypothesis is followed by three research questions:

(1) Which words resonate most in the thematic utterances, i.e. which lexical–semantic basis do people use to describe the existing situation and the emotional experience thereof?

(2) What are the specifics of the utterances on the lexical–morphological level in terms of respondents’ gender and various age cohorts?

(3) In which manner is the lexical–morphological level of the utterances influenced by the respondents’ current emotional experience?

The study consists of performing a series of psychological–linguistic analyses that relate to the utterances of the respondents (N = 2552) to two posited questions: “What does the current situation mean to you: Has it changed your life in any way? If so, how?” (this question focused primarily on the interpretation of the situation) and “How do you currently experience the situation: What do you consider the worst? On the other hand, what helps you?” (this question focused primarily on the impact of the situation and the related coping strategies). The data further include demographic descriptors of the respondents and the results of the SEHW (Scales Emotional Habitual Subjective Well-being) questionnaire which determines the valence of the experienced emotions.

**Verbal Communication Analysis: Psychology of Word Use**

A person’s verbal communication is the subject of study in several disciplines and especially a subject of long-term research in psychology (Gray, 1991). The relationships between specific communication patterns and a person’s interpersonal and intrapersonal functioning have been established in a large number of studies, e.g. screening and diagnostics of disorders through the analysis of speech products (Crystal, 1987), revealing the identity of anonymous communication (Matoušková, 2013), the prognosis of an author’s text or communicator’s behaviour (Canter & Youngs, 2009), or automatic extraction of opinions and attitudes from a text (Rodríguez-Puente et al., 2016). Studies inquire into the specific linguistic markers of gender (e.g. Shoev et al., 2016), emotionality (e.g. Brewer & Gardner, 1996), relationships (e.g. Newman et al., 2008), temperament (e.g. Mairesse et al., 2007; Schwartz et al., 2013), or pathological characteristics (e.g. Demjén, 2014).

Psychological analysis of language use usually differentiates between what a person says (language content) and how the person says it (language style; Chung & Pennebaker, 2007). The importance of language variability of a single person (language style) was repeatedly described on the level of general language usage (e.g. Chen & Bond, 2010), but also in specific word manipulations (e.g. Ireland & Mehl, 2014; Newman et al., 2008).

In terms of the relationship between specific linguistic features (language style) and the characteristics of the communicator, it is most often cited that, for example, women more frequently use “involved” parts of speech (e.g. more pronouns, present tense verbs) in comparison with men who prefer “informative” language (e.g. more nouns, long words, numbers, articles, prepositions; Biber, 1991; Newman et al., 2008). Women also more often use words in first person singular (Mehl & Pennebaker, 2003), negations and verbs (Newman et al., 2008), and more frequently express emotions and self-disclosure tendencies in the text (Holtgraves, 2014). In terms of communicators’ age, the documented differences include, for example, a higher ratio of words with positive emotional load and future tense in older people (Pennebaker & Stone, 2003). If we focus on the connections between language style and the specifics of emotional experience (emotional state), numerous approaches attempting to successfully detect emotions in the text have been introduced (see, e.g. Pang &
et al., 2016) and the occurrence of pronouns (the same resulted in higher use of negatively loaded words; Bernard emotionally loaded words (negative emotional experience documented, for example, a close connection between language analysis software (Linguistic Inquiry and Word detection of words, most frequently using the LIWC language, 2015; see below). This research meta-analysis; Holtzman, 2017). These conclusions are also supported by studies focusing on the issue of manifestations of depression and anxiety (e.g. Anderson et al., 2008; Arntz et al., 2012).

The aforesaid methods of quantitative natural language processing, in which both lexical–semantic and morphological analyses are employed, are substantial for our research. As in the above-mentioned studies, we also apply computational and statistical methods to search for relationships between language style and descriptors of the emotional experience of respondents (based on their psychological test results). In this case, however, we use an updated set of techniques, designed with respect to the Czech language specifics and higher linguistic variability.

The majority of the published studies were conducted in the English language, which brings certain risks to the transferability of the results to other non-English speaking populations. This paper focuses on the analysis of the Czech language. The Czech language, a member of the West-Slavic language group, is spoken by relatively few native speakers (10.7 million native speakers; cf. 379 million first-language speakers in English), but it is the 20th most frequently used language on the internet (W3Techs, 2020). In terms of psychological research, several studies on the Czech language, associated under the CPACT project (Computational Psycholinguistic Analysis of Czech Text; Kučera, 2018b), have been conducted in the past few years, focusing on the relationship between the morphological and lexical aspects of the text and the Big Five personality traits (Havigerová et al., 2018), dominance (Kučera et al., n.d.), and depressivity (Havigerová et al., 2019). The results of these studies imply comparability of a large part of the discoveries with the results of the anglophone studies (see Havigerová et al., 2018). However, the type of the text and the comparability (similarity) of the communication situation (to be more precise, the comparability and variability of the text registers selected for the study) plays an important role (see e.g. Biber, 1993; Kučera, 2020).

Method

Data Collection

The data collection was carried out within the JUPSYCOR project, “Research on the Psychological Impacts of the Coronavirus Epidemic in the Czech Republic” (www.jupsycor.cz). The open online interface was promoted through social networks and e-mails sent by cooperating institutions. The interface enabled two types of data collection—one for individual respondents and the other one for assistant interviewers who received online training and collected utterances from other respondents (i.e. the respondents who agreed to participate but would not be able to participate online were interviewed in the very same format through a phone call and their utterances were typed into the web form; see below). The respondents were fully informed about the nature of the survey, asked for informed consent, and participated voluntarily. They provided demographic data, answered two open questions related to their perception and experience of the COVID-19 situation (i.e. utterances), and completed self-reporting questionnaires capturing their emotional states in recent days. The data were collected for 68 days, from 18 March to 25 May 2020. This period was chosen with reference to the development of the pandemic situation in the Czech Republic (from the adoption of major restrictions in March to their decrease in May; Vlada, 2020) (Fig. 1).

Sample

A total of 2552 respondents participated in the research, men and women aged 18–89 years. This sample was obtained via opportunity sampling (see Data collection). The respondents were divided into six categories based on age (age groups; Table 1). We also included the respondents’ representation in terms of their highest achieved level of education (primary school, secondary school, university) and social classification (student, pensioner/retired, other) (Table 2). For the needs of further analysis, the respondents were divided into six groups according to age and gender (cohort groups, Table 3).

Due to the data collection method, the sample could not be balanced with respect to various demographic variables or time of participation (see CSU, 2020). The sample has a significantly increased proportion of females (79%) and young persons under 26 years (42%). The sample also shows an increased proportion of people with a university degree (38% compared to 22% of the population aged 25–64 years as reported by the National Education Fund in 2015; NVF, 2015).
The respondents were asked two questions focusing predominantly on (Q1) the interpretation of the COVID-19 situation by the respondent and (Q2) the negative and positive impacts on the respondents and their coping strategies. The respondents could write any utterance in reply to this question (no min./max. length of the text was specified). These utterances were typed into a web form.

- Q1: What does the current situation mean to you: Has it changed your life in any way? If so, how?
- Q2: How do you currently experience the situation: What do you consider the worst? On the other hand, what helps you?
As stated above, some utterances are based on a literal transcription of the respondent’s utterance by the assistant interviewers (N = 561); however, most were entered directly by the respondent (N = 1991). Only editing of the materials was performed solely in relation to typos in the text.

**Linguistic Analysis**

The basis of the linguistic analysis is the use of three referential dictionaries, SYN2015, SENS, and LIWC2007, and a set of linguistic applications for the analysis of text in the Czech language (PMA).

SYN2015: Representative Corpus of Contemporary Written Czech (Krˇen et al., 2016) is a 100-million-word corpus. It was created as a representation of printed language from 2010 to 2014, containing a wide range of text types (fiction, professional literature, newspapers, etc.). The corpus is lemmatized, morphologically and syntactically annotated by a combination of stochastic and rule-based methods. In terms of this study, SYN2015 was used for frequency analysis.

SENS: Dictionary of Emotionally Loaded Words was created by the adjustment of Czech SubLex 1.0 dictionary (Veselovská & Bojar, 2013), performed by the Institute of Theoretical and Computational Linguistics (UTKL; Faculty of Arts, Charles University). The adjustment consisted of deleting 94 words without a sufficient and/or completely obvious emotional load. SENS features 928 words (lemmas) altogether, annotated by a positive, negative, or undetermined emotional load.

LIWC2007: Linguistic Inquiry and Word Count (Pennebaker et al. 2007) is a text analysis program which functions on the basis of the closed-vocabulary approach. Its dictionary is composed of circa 4,500 words and word stems. Each word or word stem is defined by one or more word categories or subdictionaries (see ibid.). For example, the word “cried” is a part of five word categories: sadness (130 Sad), negative emotion (127 Negemo), overall affect (125 Affect), verb (11 Verbs), and past tense verb (13 Past). For this study, synonymous expressions were identified in the dictionary (after translating Czech words into English), specifically in the relevant categories of Psychological Processes (Sects. 121–253) and Personal Concerns (Sects. 354–360). In comparison with the used LIWC2007 version, the newer LIWC2015 version features some modified parts of the dictionary (moreover, certain categories were deleted, e.g. the morphological category of Tense); nevertheless, both versions demonstrate comparable parameters and a high degree of congruity (Pennebaker et al., 2015).

PMA: Prague Morphological Analysis (Hajič, 2001). All obtained texts were further processed via UTKL applications (Jelínek, 2018), collectively termed as PMA. These applications represent a Czech variant of the LIWC (see the comparison in Kučera & Haviger, 2019). However, except for one specific category (Emotions), they primarily focus on morphological analysis. The outcome of this process is the allocation of morphological tags to every lexical unit of the text with an average of 95% accuracy and, in the case of detection of various linguistic variables (e.g. part of speech), as high as 99.5% accuracy (Skoumalová, 2011). This study utilized such linguistic categories that show high compatibility with the anglophone LIWC, i.e. the grammatical categories of Part of speech, Person, Tense, Degree, and Negation, and the semantic category of Emotions that is based on the implementation of the SENS dictionary (see above). All these categories are processed in terms of values of the relative frequency occurrence (i.e. the ratio of the given category to the number of words in the utterance). The overview of the categories and subcategories is included in Table 4.

**Psychological Measures**

To ascertain the respondents’ emotional experience, SEHW: Scales of Emotional Habitual Subjective Well-being questionnaire (Džuka, 2015; Džuka & Dalbert, 2002) were used. The SEHW questionnaire is a ten-item questionnaire focused on the emotional component of

---

**Table 3 Sample: Cohort groups**

| Gender | Cohort group | 1 | 2 | 3 | 4 | 5 | 6 | Total |
|--------|--------------|---|---|---|---|---|---|-------|
| Female |              | 880| 537| 598| 0 | 0 | 0 | 2015  |
| Male   |              | 0 | 0 | 0 | 197| 161| 179| 537   |
| Total  |              | 880| 537| 598| 197| 161| 179| 2552  |

Cohort groups: 1 = females at the age of 18–25 years; 2 = females 26–44 years; 3 = females 45+ years; 4 = males 18–25 years; 5 = males 26–44 years; 6 = males 45+ years
subjective well-being (Diener, 1994), with the Positive Affect Scale consisting of four descriptors (pleasure, happiness, joy, and physical freshness/energy/briskness) and the Negative Affect Scale comprising six descriptors (anger, guilt feelings, shame, fear/anxiety, pain, and sadness/sorrow). Respondents were answering the question: “How often have you experienced these affects in the past few days?”. The explicitness of questionnaire statements and simplicity of answering for respondents were the main reasons for choosing SEHW. The questionnaire has been used in numerous studies on well-being in different populations (e.g. Džuka & Dalbert, 2006; Gurková et al., 2012). The respondents indicated how often they experienced each affect state in recent days on a six-point frequency scale ranging from 1 (almost never) to 6 (almost always). The internal consistency estimate for the Positive Affect Scale was Cronbach’s alpha = 0.83, and for the Negative Affect Scale, it was Cronbach’s alpha = 0.67 in the validation study (Džuka & Dalbert, 2002). In the present study, the respective Cronbach’s alphas were 0.85 and 0.74.

Results

Description of Utterances

Table 5 describes the numbers of words, sentences, and tokens (individual occurrences of a linguistic unit) that were recorded within the utterances (texts Q1 and Q2) while employing PMA (Prague Morphological Analysis). All texts which featured at least one word in both utterances (Q1 and Q2) were included in the analyses. The results demonstrate that both men and women wrote...
utterances of similar length, equivalent to short commentaries, slightly longer in the case of text Q2.

**Description of SEHW Questionnaire**

Table 6 features an overview of the SEHW (Scales of Emotional Habitual Subjective Well-being; Džuka & Dalbert, 2002) questionnaire results according to the individual items and the average score for negative emotions (SEHW_N) and positive emotions (SEHW_P). Table 7 features a division of respondents into two groups—low scoring and high scoring in SEHW_N negative emotions (high = mean \( \geq 2.3 \)) and low scoring and high scoring in SEHW_P positive emotions (high = mean \( \geq 3.5 \)), determined according to the sample median.

In previous studies, the positive affect mean score SEHW_P ranged from 3.00 (elderly people; Džuka & Dalbert, 2006) to 3.85 (high school students; ibid.) and negative affect mean score SEHW_N ranged from 2.28 (nurses; Gurková et al., 2014) to 2.96 (elderly people; Džuka & Dalbert, 2006). The values gathered in the present study fall within these ranges.

**Most Distinctive Words: Results of Frequency Lexical Analysis**

After performing the frequency analysis, words occurring most frequently in the utterances (Q1 and Q2) of the whole sample \( (N = 2552) \) were detected, namely in the Part of speech category: Nouns, Adjectives, Verbs, and Adverbs. These words were identified in the frequency dictionary of the SYN2015 corpus, determining the rank in which they occur in this corpus. Subsequently, S/J ratio was calculated (dividing rank 2 in SYN2015 by rank 1 in JUPSYCOR Q1/Q2 texts). For instance, the noun “restriction” is listed with rank 2 = 903 in the SYN2015 corpus, but it ranked rank 1 = 9 in our Q1 responses. It was therefore mentioned approximately 100 \( \times \) more frequently in the utterances than in common Czech communication. The overview in Tables 8 and 9 includes words with S/J ratio \( \geq 25 \). This value was determined ad hoc, regarding the legible

---

**Table 5** Utterances: Q1 and Q2 texts description for the whole sample, females and males

|       | Q1_Words | Q1_Sentences | Q1_Tokens | Q2_Words | Q2_Sentences | Q2_Tokens |
|-------|----------|--------------|-----------|----------|--------------|-----------|
| N     | 2552     | 2552         | 2552      | 2552     | 2552         | 2552      |
| M     | 21.772   | 1.936        | 25.851    | 26.930   | 2.417        | 32.016    |
| Mdn   | 18.000   | 1.000        | 22.000    | 26.000   | 2.000        | 31.000    |
| Mode  | 9.000    | 1.000        | 11.000    | 45.000   | 2.000        | 52.000    |
| SD    | 15.021   | 1.227        | 17.696    | 14.396   | 1.321        | 16.985    |
| Min   | 1.000    | 1.000        | 1.000     | 1.000    | 1.000        | 1.000     |
| Max   | 56.000   | 9.000        | 70.000    | 54.000   | 11.000       | 68.000    |

**Female**

|       | Q1_Words | Q1_Sentences | Q1_Tokens | Q2_Words | Q2_Sentences | Q2_Tokens |
|-------|----------|--------------|-----------|----------|--------------|-----------|
| N     | 2015     | 2015         | 2015      | 2015     | 2015         | 2015      |
| M     | 22.169   | 1.937        | 26.326    | 27.376   | 2.393        | 32.512    |
| Mdn   | 18.000   | 1.000        | 22.000    | 27.000   | 2.000        | 32.000    |
| Mode  | 9.000    | 1.000        | 10.000    | 44.000   | 2.000        | 52.000    |
| SD    | 14.945   | 1.232        | 17.614    | 14.266   | 1.295        | 16.842    |
| Min   | 1.000    | 1.000        | 1.000     | 1.000    | 1.000        | 1.000     |
| Max   | 54.000   | 9.000        | 69.000    | 54.000   | 9.000        | 66.000    |

**Male**

|       | Q1_Words | Q1_Sentences | Q1_Tokens | Q2_Words | Q2_Sentences | Q2_Tokens |
|-------|----------|--------------|-----------|----------|--------------|-----------|
| N     | 37       | 537          | 537       | 537      | 537          | 537       |
| M     | 20.281   | 1.933        | 24.069    | 25.257   | 2.507        | 30.156    |
| Mdn   | 15.000   | 2.000        | 18.000    | 24.000   | 2.000        | 28.000    |
| Mode  | 7.000    | 1.000        | 11.000    | 7.000    | 2.000        | 14.000    |
| SD    | 15.227   | 1.209        | 17.904    | 14.766   | 1.410        | 17.402    |
| Min   | 1.000    | 1.000        | 1.000     | 1.000    | 1.000        | 1.000     |
| Max   | 56.000   | 9.000        | 70.000    | 54.000   | 11.000       | 68.000    |

Words = number of individual words; Sentences = number of sentences (usually ending in punctuation); Tokens = number of occurrences of specific realizations of words and symbols
arrangement of the list and visualization (approximately 25 words for all parts of speech related to one question). It was ascertained for each word whether it appears in the SENS dictionary (and if yes, the emotional load of the word was included) and whether the same or synonymous word appears in the LIWC dictionary (if yes, semantic–psychological connotations of the word were included).

For the illustration of significant words (in terms of all parts of speech), a visualization of Q1 (Fig. 2) and Q2 (Fig. 3) was performed in a word cloud form. The font size corresponds to the S/J ratio value.

The most significant words appearing are, for instance, “as”, “face mask”, “contact”, and “uncertain”. The significant words with negative emotional load according to the SENS dictionary are in the absolute majority, except for two words with positive load in Q2 (“calmness” and “manage”). In terms of psychological connotations of words according to the LIWC dictionary, words in the categories of Social (9 words), Anxiety (6 words), Inhibition (5 words), and Work (4 words) are in majority.

Comparison of Respondent Groups in Terms of Linguistic Categories Usage

Further analyses were aimed to compare the use of linguistic categories (in what way are the utterances phrased) between men and women (Table 2) and between 6 cohorts (Table 3). A Mann–Whitney U test and a Kruskal–Wallis H test were run to determine whether there were significant differences in the relative frequencies of linguistic categories between these groups. The effect sizes (Cohen’s $d$) of the presented results are within a range that Cohen (1988) reports as a small effect (0.1–0.3), as given in Tables 10 and 11.

The influence of gender on phrasing the utterances was proven in three linguistic categories (POS–V, POS–R, and Em2.-), in both Q1 and Q2. In their utterances, men used a significantly lower number of verbs, fewer prepositions, and fewer emotionally negatively loaded words (Table 10).

The diversity of the cohort groups was proven in nine linguistic categories for Q1 and Q2 simultaneously (Table 11). The groups’ general tendencies in linguistic categories usage (group means/medians) are highly comparable for both texts (Q1 and Q2). The most distinctive in this regard are categories: prepositions (POS–P), used more frequently by younger people (men and women) in contrast to middle-aged people; verbs (POS–V), which are used to a higher degree by young people (especially women), and, on the other hand, less by especially older men; first person (Per–1), once again used primarily by young people, but also older women.

| SEHW questionnaire descriptives (SEHW) |
|----------------------------------------|
| SEHW1 Anger | SEHW2 Guilt | SEHW3 Pleasure | SEHW4 Shame | SEHW5 Energy | SEHW6 Fear |
| N           | 2552        | 2552          | 2552        | 2552         | 2552       |
| M           | 2.587       | 1.898         | 3.472       | 1.587        | 2.250      | 3.049      |
| Mdn         | 3           | 2             | 3           | 1            | 3          | 3          |
| Mode        | 3           | 1             | 3           | 1            | 3          | 3          |
| Std. Deviation | 1.183     | 1.073         | 1.08        | 0.917        | 1.261      | 1.345      |

| SEHW7 Pain | SEHW8 Joy | SEHW9 Sadness | SEHW10 Happiness | SEHW_N Neg. emo | SEHW_P Pos. emo |
| N           | 2552        | 2552          | 2552          | 2552          | 2552       |
| M           | 2.522       | 3.624         | 2.903         | 2.431         | 2.379      | 3.468      |
| Mdn         | 2           | 4             | 3             | 3             | 2.3        | 3.5        |
| Mode        | 1           | 3             | 3             | 3             | 2.3        | 3          |
| Std. Deviation | 1.233     | 1.073         | 1.275         | 1.147         | 0.783      | 0.958      |

| SEHW questionnaire: High and low scores |
|----------------------------------------|
| Gender | SEHW_N | Low | High | Total |
| Female | 1046 | 397 | 1443 |
| Male   | 969  | 140 | 1109 |
| Total  | 2015 | 537 | 2552 |
| Low    | 1210 | 299 | 1509 |
| High   | 805  | 238 | 1043 |
| Total  | 2015 | 537 | 2552 |
The Polarity of Respondents’ Emotional Experience in Relation to Usage of Linguistic Categories

Another set of analyses focused on the differences in linguistic categories usage between respondents who scored either high or low on the SEHW_N (negative emotions) and SEHW_P (positive emotions) scales (Table 7). A Mann–Whitney U test was run to determine whether there were significant differences in the relative frequencies of linguistic categories between these groups for both \( Q1 \) and \( Q2 \). However, the effect sizes (Cohen’s \( d \)) of the presented results are within the range that Cohen (1988) reports as a very small effect (0.1 on average), as given in Tables 12 and 13.

Although the differences between the groups are minor, people reaching a higher score in SEHW_N (negative
emotions) exhibit a significantly higher usage of adjectives (POS–A) and future tense (Ten–F) in both texts (Q1 and Q2), and, contrastingly, lower usage of proverbs (POS–D) (Table 12). Regarding SEHW_P (positive emotions), the differences are again very minor; however, there is an obvious difference between the groups in the category of Deg–2 (second degree, comparative), more often used by people with a higher ratio of positive emotions (in both texts; Table 13).
A Spearman’s rank-order correlation was run to assess the relationship between 27 linguistic categories (relative frequencies, Table 4) and 12 SEHW (SEHW results, Table 6) for Q1 and Q2. The test was performed on nine different sample groups altogether: on the whole sample ($N = 2552$), on the six cohorts (Table 3), and on men and women (Table 2).

A high number of small but significant correlations ($p < 0.05$) were found within all groups, usually with the value of $r_s$ (< 0.1 and > − 0.1). Three hundred and fifty-one significant correlations were found in Q1, and 336 significant correlations in Q2. Ninety-four correlations thereof were significant in both texts, and 93 of these demonstrated the same correlation direction (see Supplement 1).

After making Šidák’s adjustment (Šidák, 1967) to the level $p_{adj} < 0.0001583$, 48 significant relationships were found in Q1 across all groups and 10 significant relationships in Q2. Six relationships thereof fulfilled the $p$-adjustment conditions in both Q1 and Q2 (Table 10). The relationships proven by Šidák’s statistical correction are linked primarily to the linguistic category Em2.- (emotionally negatively loaded words), which positively correlates with the scales SEHW_6 (fear) and SEHW_N (negative emotions mean). It is therefore apparent that the respondents experiencing negative emotions use a higher number of negative words. The only confirmed morphological category was Ten-F (future tense), which showed positive correlation with the scale SEHW_6 (fear) within the whole sample ($N = 2552$). People experiencing fear

![Fig. 2 Q1: Most distinctive words (word cloud)](image1)

![Fig. 3 Q2: Most distinctive words (word cloud)](image2)

Table 10 Gender and linguistic categories for both Q1 and Q2 (Mann–Whitney U test)

| Gender     | $n$  | Q1_POS–V | Q2_POS–V | Q1_POS–R | Q2_POS–R | Q1_Em2.- | Q2_Em2.- |
|------------|------|----------|----------|----------|----------|----------|----------|
| Mdn (female) | 2015 | 0.200    | 0.205    | 0.107    | 0.100    | 0.000    | 0.020    |
| Mdn (male)   | 537  | 0.184    | 0.200    | 0.091    | 0.091    | 0.000    | 0.000    |
| M (female)   | 2015 | 0.206    | 0.205    | 0.107    | 0.101    | 0.031    | 0.032    |
| M (male)     | 537  | 0.188    | 0.195    | 0.092    | 0.094    | 0.021    | 0.029    |
| U           | 493,827 | 507,898.5 | 474,822 | 497,008 | 505,897 | 497,182.5 |
| $z$         | − 3.116 | − 2.184  | − 4.380  | − 2.904  | − 2.755  | − 3.104  |
| $p$         | 0.002  | 0.029    | 0.000    | 0.004    | 0.006    | 0.002    |
| $d$         | − 0.135 | − 0.107  | − 0.196  | − 0.119  | − 0.122  | − 0.063  |
| CI (95%)    | − 0.231 | − 0.04   | − 0.203  | − 0.012  | − 0.291  | − 0.1     | − 0.214  | − 0.024  | − 0.217  | − 0.026  | − 0.159  | − 0.032  |

For explanatory note on linguistic categories, see Table 4
Table 11 Cohort groups and linguistic categories for both Q1 and Q2 (Kruskal–Wallis H test for independent samples)

| Groups | Q1_POS–N | Q1_POS–P | Q1_POS–V | Q1_POS–Z | Q1_Per–1 | Q1_Ten–F | Q1_Ten–P |
|--------|---------|---------|---------|---------|---------|---------|---------|
| n      | 880     | 880     | 880     | 880     | 880     | 880     | 880     |
| Mdn    | 0.228   | 0.118   | 0.214   | 0.167   | 0.136   | 0.003   | 0.125   |
| M      | 0.241   | 0.122   | 0.219   | 0.177   | 0.135   | 0.003   | 0.125   |
| $\chi^2$ (5) | 20.397 | 32.403  | 51.863  | 27.239  | 50.718  | 23.225  | 36.100  |
| $p$    | 0.001   | 0.000   | 0.000   | 0.000   | 0.000   | 0.000   | 0.000   |
| $d$    | 0.156   | 0.209   | 0.274   | 0.188   | 0.27    | 0.17    | 0.17    |
Table 11 continued

| Groups | 1 | 2 | 3 | 4 | 5 | 6 | 1–6 |
|--------|---|---|---|---|---|---|-----|
| d | 0.222 |
| **Ling. cat.** | Q1_Vneg | | | | | | |
| n | 880 | 537 | 598 | 197 | 161 | 179 | 2552 |
| Mdn | 0.024 | 0 | 0 | 0.023 | 0 | 0 | 0.008 |
| M | 0.048 | 0.035 | 0.042 | 0.047 | 0.042 | 0.05 | 0.044 |
| $\chi^2$ (5) | | | | | | | |
| p | | | | | | | |
| d | 0.221 |
| **Ling. cat.** | Q1_Em2.* | | | | | | |
| n | 880 | 537 | 598 | 197 | 161 | 179 | 2552 |
| Mdn | 0 | 0 | 0 | 0 | 0 | 0 | 0.000 |
| M | 0.026 | 0.026 | 0.025 | 0.025 | 0.029 | 0.027 | 0.026 |
| $\chi^2$ (5) | | | | | | | |
| p | | | | | | | |
| d | 0.219 |
| **Ling. cat.** | Q2_POS–N | | | | | | |
| n | 880 | 537 | 598 | 197 | 161 | 179 | 2552 |
| Mdn | 0.228 | 0.25 | 0.225 | 0.211 | 0.235 | 0.214 | 0.227 |
| M | 0.2353 | 0.2635 | 0.2675 | 0.2371 | 0.287 | 0.2675 | 0.26 |
| $\chi^2$ (5) | | | | | | | |
| p | | | | | | | |
| d | 0.126 |
| **Ling. cat.** | Q2_POS–P | | | | | | |
| n | 880 | 537 | 598 | 197 | 161 | 179 | 2552 |
| Mdn | 0.149 | 0.125 | 0.13 | 0.143 | 0.125 | 0.139 | 0.135 |
| M | 0.148 | 0.127 | 0.1327 | 0.1459 | 0.1208 | 0.1374 | 0.135 |
| $\chi^2$ (5) | | | | | | | |
| p | | | | | | | |
| d | 0.134 |
| **Ling. cat.** | Q2_POS–V | | | | | | |
| n | 880 | 537 | 598 | 197 | 161 | 179 | 2552 |
| Mdn | 0.215 | 0.196 | 0.2 | 0.2 | 0.196 | 0.194 | 0.2 |
| M | 0.2159 | 0.1922 | 0.1997 | 0.2047 | 0.1901 | 0.1894 | 0.199 |
| $\chi^2$ (5) | | | | | | | |
| p | | | | | | | |
| d | 0.229 |
| **Ling. cat.** | Q2_POS–Z | | | | | | |
| n | 880 | 537 | 598 | 197 | 161 | 179 | 2552 |
| Mdn | 0.176 | 0.179 | 0.194 | 0.182 | 0.194 | 0.19 | 0.186 |
| M | 0.1847 | 0.1897 | 0.2011 | 0.1921 | 0.2075 | 0.2139 | 0.198 |
| $\chi^2$ (5) | | | | | | | |
| p | | | | | | | |
| d | 0.134 |
| **Ling. cat.** | Q2_Per–1 | | | | | | |
| n | 880 | 537 | 598 | 197 | 161 | 179 | 2552 |
| Mdn | 0.122 | 0.111 | 0.128 | 0.125 | 0.1 | 0.118 | 0.117 |
| M | 0.1261 | 0.1065 | 0.1279 | 0.1238 | 0.1021 | 0.1186 | 0.118 |
Table 11 continued

| Groups | 1 | 2 | 3 | 4 | 5 | 6 | 1–6 |
|--------|---|---|---|---|---|---|-----|
| \(\chi^2\) (5) | | | | | | | 34.178 |
| \(p\) | | | | | | | 0.000 |
| \(d\) | | | | | | | 0.215 |
| Ling. cat. | \(Q2\_Ten–F\) | | | | | | |
| \(n\) | 880 | 537 | 598 | 197 | 161 | 179 | 2552 |
| Mdn | 0 | 0 | 0 | 0 | 0 | 0 | |
| M | 0.004975 | 0.003669 | 0.003174 | 0.003162 | 0.002447 | 0.003587 | |
| \(\chi^2\) (5) | 880 | 537 | 598 | 197 | 161 | 179 | 12.059 |
| \(p\) | | | | | | | 0.034 |
| \(d\) | | | | | | | 0.105 |
| Ling. cat. | \(Q2\_Ten–P\) | | | | | | |
| \(n\) | 880 | 537 | 598 | 197 | 161 | 179 | 2552 |
| Mdn | 0.15 | 0.143 | 0.148 | 0.143 | 0.143 | 0.143 | 0.145 |
| M | 0.1523 | 0.1349 | 0.148 | 0.1442 | 0.145 | 0.1466 | |
| \(\chi^2\) (5) | | | | | | | 16.573 |
| \(p\) | | | | | | | 0.005 |
| \(d\) | | | | | | | 0.135 |
| Ling. cat. | \(Q2\_Vneg\) | | | | | | |
| \(n\) | 880 | 537 | 598 | 197 | 161 | 179 | 2552 |
| Mdn | 0.022 | 0 | 0 | 0.022 | 0 | 0 | |
| M | 0.03125 | 0.02694 | 0.02884 | 0.03056 | 0.0282 | 0.02678 | 0.029 |
| \(\chi^2\) (5) | | | | | | | 16.189 |
| \(p\) | | | | | | | 0.006 |
| \(d\) | | | | | | | 0.133 |
| Ling. cat. | \(Q2\_Em2.*\) | | | | | | |
| \(n\) | 880 | 537 | 598 | 197 | 161 | 179 | 2552 |
| Mdn | 0.038 | 0.026 | 0.025 | 0.031 | 0.033 | 0.026 | 0.03 |
| M | 0.04357 | 0.03569 | 0.0362 | 0.04425 | 0.04573 | 0.03559 | 0.04 |
| \(\chi^2\) (5) | | | | | | | 13.066 |
| \(p\) | | | | | | | 0.023 |
| \(d\) | | | | | | | 0.113 |

Groups: 1 = females at the age of 18–25 years; 2 = females 26–44 years; 3 = females 45+ years; 4 = males 18–25 years; 5 = males 26–44 years; 6 = males 45+ years. For explanatory note on linguistic categories, see Table 4

Table 12 SEHW_N score (negative emotions) and linguistic categories for both Q1 and Q2 (Mann–Whitney U test)

| SEHW_N | \(n\) | Q1_POS–A | Q2_POS–A | Q1_POS–D | Q2_POS–D | Q1_Ten–F | Q2_Ten–F |
|--------|-----|----------|----------|----------|----------|----------|----------|
| Mdn (SEHW_N low score) | 1443 | 0.061 | 0.083 | 0.095 | 0.089 | 0.000 | 0.000 |
| Mdn (SEHW_N high score) | 1109 | 0.068 | 0.091 | 0.087 | 0.080 | 0.000 | 0.000 |
| M (SEHW_N low score) | 1443 | 0.080 | 0.092 | 0.117 | 0.097 | 0.002 | 0.003 |
| M (SEHW_N high score) | 1109 | 0.082 | 0.099 | 0.103 | 0.085 | 0.003 | 0.005 |
| U | 838,716 | 856,402 | 763,062 | 748,819 | 816,748.5 | 828,066 |
| z | 2.121 | 3.054 | -2.027 | -2.796 | 2.328 | 3.011 |
| \(p\) | | | | | | | 0.028 |
| \(d\) | | | | | | | 0.123 |
| CI (95%) | | | | | | | |

For explanatory note on linguistic categories, see Table 4

© Springer
therefore use a higher number of words related to the future (Table 14).

In terms of relationships concerning solely morphological categories, Table 15 features an overview of 11 significant relationships that are related to both Q1 and Q2 and scales SEHW_N (negative emotions mean) and SEHW_P (positive emotions mean). The complete correlation matrix is included in Supplement 2.

### Discussion

The previous text introduced the results of a study focusing on word usage in a reflection of the state of emergency, COVID-19 epidemic in the Czech Republic, and the connections these words have to the emotional experience in 2552 respondents. The importance of the study in this regard lies in two aspects—first, it describes the specifics of thematically focused utterances and their linguistic parameters in different respondent groups, and, second, it documents those linguistic features that refer to the respondents’ emotional experience.

Before interpreting the results as such, it is necessary to point out to the specifics of the research sample, the specifics of the time framework of the data collection, and the specifics of the Czech language. As mentioned above, the sample features a majority of women and young people, predominantly students. It is, therefore, necessary to consider the extent of the influence of selection bias on the results. Nevertheless, we appreciate, in comparison with other COVID-19-themed researches (see, e.g. Özdin & Bayrak Öydin, 2020; Rodriguez-Rey et al., 2020), the relatively large representation of older people. This representation was achieved also because the older respondent group was often questioned via assistant interviewers (see above), without relying on contacting respondents solely via social media. The time frame selected for the research covered 68 days. It is therefore apparent that the respondents’ utterances might have been (and undoubtedly partially were) influenced by the situation at that time. From 18 March to 5 April, the situation was at its most serious in the Czech Republic (adoptions of major emergency measures, e.g. closure of schools, restaurants, shops, imposed face masks, restrictions of free movement, etc.).

### Table 13 SEHW_P score (positive emotions) and linguistic categories for both Q1 and Q2 (Mann–Whitney U test)

| SEHW_P        | n   | Q1_Deg–2 | Q2_Deg–2 |
|---------------|-----|----------|----------|
| Mdn (SEHW_P low score) | 1509 | 0.000    | 0.000    |
| Mdn (SEHW_P high score) | 1043 | 0.000    | 0.000    |
| M (SEHW_P low score) | 1509 | 0.017    | 0.007    |
| M (SEHW_P high score) | 1043 | 0.024    | 0.009    |
| U              | 834,059 | 820,086.5|          |
| z              | 3.352 | 2.693    |          |
| p              | 0.001 | 0.007    |          |
| d              | 0.124 | 0.085    |          |
| CI (95%)       | 0.045–0.203 | 0.006–0.164|          |

For explanatory note on linguistic categories, see Table 4.

### Table 14 SEHW and linguistic categories for Q1 and Q2 (Spearman’s rank-order correlation; Šídák’s adjustment)

| Group           | SEHW        | Ling. cat | Q1 rs* | Q2 rs* |
|-----------------|-------------|-----------|--------|--------|
| Cohort_1        | SEHW6_Fear  | Em2.-     | 0.171  | 0.144  |
| N = 2552        | SEHW6_Fear  | Em2.-     | 0.220  | 0.137  |
|                 | SEHW6_Fear  | Ten–F     | 0.098  | 0.089  |
|                 | SEHW6_Fear  | Em2.-     | 0.199  | 0.100  |
| Female          | SEHW6_Fear  | Em2.-     | 0.211  | 0.117  |
|                 | SEHW_N      | Em2.-     | 0.196  | 0.085  |

*p < 0.00001
afterwards, the restrictions were softening, and in late May, the situation in the society was relatively optimistic (albeit with milder restrictions still applicable). In this aspect, it was difficult to set a clearer limit than the one relating to the adoption of government measures (see Vlada, 2020). The use of Czech language analysis also resulted in certain compromises, connected predominantly to the necessity of key word translation (including presentations of relevant examples and tracing relationships with the English expressions) and the selection of suitable linguistic categories (which were selected especially regarding their compatibility with English). The central point in this regard was the transparency of the whole process while also striving for a maximum transferability of the results to other languages.

If we focus on the first research question (What words resonate the most in the thematic utterances, i.e. which lexical–semantic basis do people use to describe the current situation and their emotional experience thereof?), it is not surprising that across all utterances, the words that resonated most were words connected with the social situation and with negative connotations. Words related to anxiety and inhibition and references to social environment and work are prevalent. However, in the second utterance (focused, among else, on coping), words suggesting activities perceived positively appeared as well (e.g. “calmness, nature, walk, chill”). It is also not surprising that the highest ranking positions of lexically unique words are occupied by such words as “face mask, lockdown, infected”, which were omnipresent in the media in the Czech Republic at that time as well (see, e.g. Trait, 2020). The adverb “as” (“jako” in Czech) is an interesting phenomenon, because it appeared in both utterances 390 times more than in regular communication. This word may have fulfilled several roles in the utterances—the common usage (e.g. She works “as” a teacher), to express similarity (e.g. He behaves “like” a mad man), to connect (e.g. In winter “as well” as in summer), to present an example (e.g. Some people, “such as” old persons), and colloquially to express aloofness (e.g. So what?). The word may thus indicate a tendency to refer to another fact or parallel, or the inability to specify the content of the communication. The transition towards unspecified cognitive categories and metaphorical language might mean that the situation is cognitively more complex than is common, or that it is not sufficiently cognitively processed by the respondent (which manifests also on the verbalization level; see e.g. Lupyan & Casasanto, 2015).

In terms of the second research question (What are the specifics of the utterances on the lexical–morphological level in terms of respondents’ gender and various age cohorts?), the analysis of differences in linguistic categories usage in the utterances among various respondent groups proved several significant results, albeit with a relatively low effect size. In their utterances on the perception of the COVID-19 situation, men used fewer verbs, fewer prepositions, and fewer emotionally negatively loaded words. However, these findings generally conform to the referential research focusing on a common text, i.e. communication outside of an exceptional situation (e.g. Biber, 1991; Newman et al., 2008). That the presented findings are more of a result of common gender differences is supported by comparing the results with studies on the Czech language carried out within the CPACT project (Kučera, 2018b), i.e. the use of verbs and the more frequent use of the first person can be generally considered as a

---

**Table 15** SEHW_N and SEHW_P scales and morphological categories for Q1 and Q2 (Spearman’s rank-order correlation)

| Linguistic category | Group | SEHW | Q1 rs | Q1 p  | Q2 rs | Q2 p  |
|---------------------|-------|------|--------|--------|--------|--------|
| POS–N               | Female| SEHW_P | -0.046 | 0.041 | -0.057 | 0.011 |
| POS–A N = 2552      | SEHW_P | -0.052 | 0.008 | -0.040 | 0.042 |
| POS–A               | Female| SEHW_P | -0.051 | 0.023 | -0.048 | 0.032 |
| POS–A N = 2552      | SEHW_N | 0.056 | 0.005 | 0.045 | 0.023 |
| POS–D               | Female| SEHW_P | 0.068 | 0.002 | 0.073 | 0.001 |
| POS–D               | SEHW_N | -0.046 | 0.039 | -0.048 | 0.031 |
| Ten–F N = 2552      | SEHW_N | 0.063 | 0.001 | 0.064 | 0.001 |
| Ten–F               | Female| SEHW_N | 0.051 | 0.023 | 0.054 | 0.016 |
| Deg–2 Cohort_1      | SEHW_P | 0.096 | 0.004 | 0.083 | 0.014 |
| Deg–2 N = 2552      | SEHW_P | 0.058 | 0.004 | 0.052 | 0.008 |
| Deg–2               | Female| SEHW_P | 0.081 | 0.000 | 0.053 | 0.017 |

For an explanatory note on linguistic categories, see Table 4.
relatively reliable gender indicator (Kučera, 2020, p. 84). In terms of comparing the six cohorts (based on gender and age), the distinctive differences include the parts of speech of prepositions, which were more frequently used by younger people (men and women) in comparison with middle-aged people, verbs, which were more frequently used by young people (especially women; in contrast, less frequently by older men), and first person, which was again used predominantly by young people, but also older women.

The results related to the third research question (In which manner is the lexical–morphological level of the utterances influenced by the respondents’ current emotional experience?), which concentrated on the relationship between the linguistic categories usage in the utterances and the score in the SEHW questionnaire (negative and positive emotions), confirm the premises presented in the referential research. Regarding the lexical–semantic basis of the utterances, it is apparent that negative emotional experience positively correlates with the usage of emotionally negatively loaded words (see Bernard et al., 2016). This relationship is even more distinctive in the sixth item of the SEHW questionnaire, which asks, within the negative emotions group, directly about the experience of fear, standing out in the group of younger women in particular. It is necessary to mention that especially the female group, namely younger women, generally scored highest in the negative emotions scales in comparison with other groups (although the negative emotions scores generally fell within the SEHW test norms). The congruence between words and emotions is therefore the most pronounced here.

The importance of the aforementioned relationships is supported also by the comparison with the results of studies carried out within the CPACT project (see above), which attest to a significantly lower occurrence of emotionally negatively loaded words in a common text in contrast to their occurrence in the studied thematized utterances. The evidence of higher scores of negative emotional experiences in women (especially in younger women) is well documented in many cross-cultural studies (see, e.g. De Bolle et al., 2015; Klimstra et al., 2009). It could therefore be assumed that even in the state of emergency, the general trend is similar.

Regarding purely morphological variables, 93 significant correlations (without performing statistical correction) were found appearing in both utterances in the same manner, with six significant correlations thereof after performing Šidák’s p-adjustment. One of the interesting findings is, for instance, the higher usage of future tense in persons who describe more negative emotions, and, conversely, a higher usage of comparatives in persons who express more positive emotions. Let us add that both these morphological categories appear in the respondents’ utterances to a degree significantly different from common Czech text—they might therefore present a potentially interesting psychological indicator. A higher degree of future tense usage was documented, for example, in research focused on observing respondents’ confusion (see D’Mello & Graesser, 2012). Simultaneously, this relationship may be supported by the reasoning that the worries of respondents scoring higher in the negative emotions scale will be directed primarily towards the future (e.g. anticipatory anxiety, Butler & Mathews, 1987), and therefore refer to future. In terms of the use of comparative, the use of the second degree (comparison or gradation of the meaning) may be an expression of a certain aloofness of the communicator, related to experiencing the situation in a more positive manner. Nevertheless, a more precise interpretation must be verified by further research. It is pertinent to add that in contrast to anglophone research, no relationship of higher significance between the respondents’ characteristics and pronouns usage was detected. However, it is probably a specific characteristic of the Czech language, which does not require the use of a pronoun in a sentence. (The pronoun can be implicitly expressed by the verb form.) Additionally, the low frequency of these relationships is confirmed by previous research on Czech texts (see, e.g. Kučera, 2018a).

Several key findings arise from the presented study. The situation related to COVID-19 modified the respondents’ (personal) vocabularies in connection with the description of their emotional experience. The respondents presumably adapted to the general discourse, and a distinctive preference for words with a negative connotation appeared. Negatively emotionally loaded words occurred more frequently in women’s utterances and positively correlated with experienced negative emotions, especially with fear. This relationship was also confirmed within the whole sample. The experience of fear also positively correlated with the morphological category of future tense, where the highest scores were detected in the younger age category of 18–25 years.

The benefits of this study in comparison with big-data analyses of online communication (e.g. Yu et al., 2020; Madria & Kabir, 2020) lie in the emphasis on the psychological level of communication and the usage of standardized psychological measures, and the more precise thematic specification of the analysed texts (the relationship to subjective interpretation, emotional experience, and the respondents’ coping with the situation). Owing to the data collection procedure, it was also possible to ensure a higher representation of older people, who are especially important with regard to the topic of the study. Another valuable aspect of the presented research lies in the use of the combination of lexical–semantic and morphological
analyses of the texts, in contrast to, for example, stand-alone sentiment analysis (e.g. Liu, 2015).

Authors' Contributions Dalibor Kucera was involved in data collection, data processing, and publication outcomes.

Funding This research study was not funded by a grant or subvention.

Availability of Data and Materials All presented data and results are available to other researchers by request.

Declarations

Conflict of Interest The author of the paper and the paper content have no conflict of interest.

Ethics Approval Ethics Committee of the Faculty of Education, University of South Bohemia, decided that the methods of implementation of the assessed project JUPSYCOR comply with applicable principles, regulations, and international standards for conducting research involving human participants and declared oral consent for anonymous collection and processing of data.

Consent to Participate The research was carried out in accordance with the recommendations of the ethical code of the University of South Bohemia in Ceske Budejovice. All participants agreed to participate voluntarily and on their own free will. Participants provided consent to participate through a web interface. For further details, see http://jupycor.cz/instrukce-k-rozhovoru/#souhas.

References

Anderson, B., Goldin, P. R., Kurita, K., & Gross, J. J. (2008). Self-representation in social anxiety disorder: Linguistic analysis of autobiographical narratives. Behaviour Research and Therapy, 46(10), 1119–1125.

Arntz, A., Hawke, L. D., Bamelis, L., Spinhoff, P., & Molendijk, M. L. (2012). Changes in natural language use as an indicator of psychotherapeutic change in personality disorders. Behaviour Research and Therapy, 50(3), 191–202.

Bernard, J. D., Baddeley, J. L., Rodriguez, B. F., & Burke, P. A. (2016). Depression, language, and affect: An examination of the influence of baseline depression and affect induction on language. Journal of Language and Social Psychology, 35(3), 317–326.

Biber, D. (1991). Variation across speech and writing. Cambridge University Press.

Biber, D. (1993). Using register-diversified corpora for general language studies. Computational Linguistics, 19(2), 219–241.

Brewer, M. B., & Gardner, W. (1996). Who is this “we”? Levels of collective identity and self-representations. Journal of Personality & Social Psychology, 71, 83–93.

Butler, G., & Mathews, A. (1987). Anticipatory anxiety and risk perception. Cognitive Therapy and Research, 11(5), 551–565.

Canter, D., & Young, D. (2009). Investigative Psychology: Offender Profiling and the Analysis of Criminal Action. John Willey and Sons.

Chen, S. X., & Bond, M. H. (2010). Two languages, two personalities? Examining language effects on the expression of personality in a bilingual context. Personality and Social Psychology Bulletin, 36(11), 1514–1528.

Chung, C., & Pennebaker, J. W. (2007). The psychological functions of function words. Social Communication, 1, 343–359.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Erlbaum.

Crystal, D. (1987). Towards a ‘bucket’ theory of language disability: Taking account of interaction between linguistic levels. Clinical Linguistics & Phonetics, 1(1), 7–22.

CSU: Czech Statistical Office (2020). Statistics—Population. Retrieved from: https://www.czso.cz/csu/czso/population

D’Mello, S. K., & Graesser, A. (2012). Language and discourse are powerful signals of student emotions during tutoring. IEEE Transactions on Learning Technologies, 5(4), 304–317.

De Bolle, M., De Fruyt, F., McCrae, R. R., Löckenhoff, C. E., Costa, P. T., Jr., Aguilar-Vafiae, M. E., & Terracciano, A. (2015). The emergence of sex differences in personality traits in early adolescence: A cross-sectional, cross-cultural study. Journal of Personality and Social Psychology, 108(1), 171.

Demjén, Z. (2014). Drowning in negativism, self-hate, doubt, madness: Linguistic insights. Communication & Medicine, 11(1), 41.

Diener, E. (1994). Assessing subjective well-being: Progress and opportunities. Social Indicators Research, 31(2), 103–157.

Džuka, J. (2015). Well-Being in Slovakia. In W. Glatzer, L. Camfield, V. Moller, & M. Rojas (Eds.), Global Handbook of Quality of Life (pp. 663–684). Springer.

Džuka, J., & Dalbert, C. (2002). Vývoj a overené validity škál emocionálnej habituálnej subjektívnej pohody (sehp) [Elaboration and verification of emotional habitual subjective well-being scales (SEHP)]. Československá Psychologie: Časopis pro Psychology & Policy, 3(2), 95.

Džuka, J., & Dalbert, C. (2006). The belief in a just world and subjective well-being in old age. Aging and Mental Health, 10(5), 439–444.

Fiehler, R. (2002). How to do emotions with words: Emotionality in conversations. The verbal communication of emotions, 79–106.

Gray, P. (1991). Psychology. Worth.

Gurková, E., Džuka, J., Soósová, M. S., Žiaková, K., Haroková, S., & Seraflová, R. (2012). Measuring subjective quality of life in Czech and Slovak nurses: Validity of the Czech and Slovak versions of personal well-being index. Journal of Social Research & Policy, 3(2), 95.

Gurková, E., Haroková, S., Džuka, J., & Žiaková, K. (2014). Job satisfaction and subjective well-being among Czech nurses. International Journal of Nursing Practice, 20(2), 194–203.

Hajíc, J. (2001). Disambiguation of Rich Infection. Praha: Karolinum.

Havigerová, J. M., Haviger, J., & Franková (2018). Odráz osobnosti v textu - rysy modelu Big Five [Reflection of a personality in texts—Big Five traits]. In D. Kucera, J. M. Havigerová, J. Haviger, V. Čvrček, Z. Komrsková, D. Lukš, T. Jelíněk, T. Urbánek, J. Franková (Eds.), CPACT Research: Computational psycholinguistic analysis of Czech text (pp. 154–173). České Budějovice: PF JU.

Havigerová, J. M., Haviger, J., Kucera, D., & Hoffmannová, P. (2019). Text-based detection of the risk of depression. Frontiers in Psychology. https://doi.org/10.3389/fpsyg.2019.00513

Holgraves, T. M. (Ed.). (2014). The Oxford handbook of language and social psychology. Oxford Library of Psychology (p. 211).

Holzman, N. S. (2017). A meta-analysis of correlations between depression and first person singular pronoun use. Journal of Research in Personality, 68, 63–68.

Ireland, M. E., & Mehl, M. R. (2014). Natural language use as a marker. In T. M. Holgraves (Ed.), The Oxford Handbook of Language and Social Psychology (pp. 201–237). Oxford University Press.

Jelíněk, T. (2018). Linguistic analysis of texts. In D. Kucera, J. M. Havigerová, J. Haviger, V. Čvrček, Z. Komrsková, D. Lukeš,
T. Jelíněk, T. Urbánk, I. Franková (Eds.), *CPACT Research: Computational psycholinguistic analysis of Czech text* (pp. 89–110). České Budějovice: PF JU.

Klimstra, T. A., Hale, W. W., Raajmakers, A. W., Branje, S. J., & Meues, W. H. (2009). Maturation of personality in adolescence. *Journal of Personality and Social Psychology, 96*(4), 898.

Křen, M., Čvrček, V., Čapka T., Čermáková A., Hnátková M., Chlumská L., Jelíněk T., Kováříková D., Petkevič P., Procházková P., Skoumalová H., Škrabál M., Truneček P., & Vondřička P. (2016). SYN2015: Representative Corpus of Contemporary Written Czech. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16). Portoroz, ELRA, (pp. 2522–2528).

Kučera, D. (2018a). Komputační lingvistika v kontextu psychologického výzkumu: Současná témata a vybrané přístupy [Computational linguistics in the context of psychological research: Current issues and selected approaches]. In D. Kučera, J. M. Havigerová, J. Haviger, V. Čvrček, Z. Komrsková, D. Lukeš, T. Jelíněk, T. Urbánk, J. Franková (Eds.), *CPACT Research: Computational psycholinguistic analysis of Czech text, České Budějovice: PF JU*, (pp. 19–30).

Kučera, D. (2018b). Úvod do výzkumu CPACT [Introduction to the CPACT research]. In D. Kučera, J. M. Havigerová, J. Haviger, V. Čvrček, Z. Komrsková, D. Lukeš, T. Jelíněk, T. Urbánk, J. Franková (Eds.), *CPACT Research: Computational psycholinguistic analysis of Czech text, České Budějovice: PF JU*, (pp. 44–58).

Osobnostní markery v textu: Aplikace kvantitativní psycholinguistické analýzy písemného projevu při popisu osobnosti [Personality markers in text: Application of quantitative psychological-linguistic analysis of written text in personality description] Jihočeská univerzita v Českých Budějovicích

Kučera, D., & Haviger, J. (2019). Analysis of formal characteristics of text in the CPACT Research: Enhancing the LIWC linguistic processing for the Czech language. *Journal of Advanced Research in Social Sciences and Humanities, 4*(2), 57–65.

Kučera, D., Haviger, J., & Havigerová, J. M. (n.d.). Text-based Detection of Dominance and Arrogance in Different Types of Letters. Submitted manuscript preprint: https://osf.io/6wpjbb

Liu, B. (2015). Sentiment analysis: Mining opinions, sentiments, and emotions. Cambridge University Press.

Lupyan, G., & Casasanto, D. (2015). Meaningless words promote meaningful categorization. *Language and Cognition, 7*(2), 167–193.

Madria, S., & Kabir, Y. (2020). CoronaVis. Retrieved from: https://mykabar.github.io/coronavis

Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research, 30*, 457–500.

Matoušková, I. (2013). Aplikovaná forenzní psychologie [Applied Forensic Psychology]. Praha: Grada.

Mehl, M. R., & Pennebaker, J. W. (2003). The sounds of social life: A psychometric analysis of students’ daily social environments and natural conversations. *Journal of Personality & Social Psychology, 84*, 857–870.

NVF: National Education Fund (2015). Educational structure of the NVF: National Education Fund (2015). Educational structure of the Czech Education System (2015–2016). Praha: Nakladatelství Lidové noviny, (pp. 199–207).

Ozdín, S., & Bayrak Özdin, Ş. (2020). Levels and predictors of anxiety, depression and health anxiety during COVID-19 pandemic in Turkish society: The importance of gender. Advance online publication. https://doi.org/10.1177/002064020927051

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval, 2*(1–2), 1–135.

Pennebaker, J. W., & Stone, L. D. (2003). Words of wisdom: Language use over the lifespan. *Journal of Personality & Social Psychology, 85*, 291–301.

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). Linguistic inquiry and word count: LIWC 2005. Austin, TX: LIWC.

Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015). Linguistic Inquiry and Word Count: LIWC2015. Austin, TX: Pennebaker Conglomerates (www.LIWC.net).

Peters, K., Kashima, Y., & Clark, A. (2009). Talking about others: Emotionality and the dissemination of social information. *European Journal of Social Psychology, 39*(2), 207–222.

Ribeiro, F. N., Araújo, M., Gonçalves, P., André Gonçalves, M., & Benevenuto, F. (2016). SentiBench—a benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science, 5*(1), 1–29. https://doi.org/10.1140/epjds/s13688-016-0085-1

Rodriguez-Puente, P., Fanego, T., Gandón-Chapela, E., Riveiro-Oteiral, S. M., & Roca-Varela, M. L. (2016). *Current Research in Applied Linguistics: Issues on Language and Cognition*. CS Publishing.

Rodriguez-Rey, R., Garrido-Hernansaz, H., & Collado, S. (2020). Psychological impact and associated factors during the initial stage of the coronavirus (COVID-19) pandemic among the general population in Spain. *Frontiers in Psychology, 11*, 1540.

Rocki, A., Lativinova, T., Gudovskikh, D., Rybka, R., & Moloshoski, I. (2016). Machine learning models of text categorization by author gender using topic-independent features. *Procedia Computer Science, 101*, 135–142.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., & Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one, 8*(9), e73791.

Šidák, Z. K. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association, 62*(318), 626–633. https://doi.org/10.1080/01621459.1967.10482935

Skoumalová, H. (2011). Porovnání úspěšnosti tagování korpusu [Comparison of corpus tagging success rates]. In Petkevič V. & Renát A., Korpusová lingvistika Praha (Eds.) Č hranství a značkování korpusů. Praha: Nakladatelství Lidové noviny, (pp. 199–207).

Soukup & Rabušic. (2020). *Hodnoty ve světle COVID-19* [Values in the COVID-19 situation]. Retrieved from: https://siv.cuni.cz/sites/default/files/uploads/files/v%C3%BDsledky_v%C3%BDzku_Hodnoty_ve_sv%C3%A1tlu_COVID-19.pdf

Sun, J., Schwartz, H. A., Son, Y., Kern, M. L., & Vazire, S. (2019). The language of well-being: Tracking fluctuations in emotion experience through everyday speech. *Journal of Personality and Social Psychology, 123*(2), 154–167.

Twenge, J. M., & Joiner, T. E. (2020). Mental distress among U.S. adults during the COVID-19 pandemic. https://doi.org/10.31234/osf.io/wc8ud

Veselovská, K., & Bojar, O. (2013). *Czech SubLex 1.0, LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (UFAL).* Faculty of Mathematics and Physics, Charles University. Dostupné z http://hdl.handle.net/11858/00-097C-0000-0022-FF60-B

Springer
Vlada CR: Government of the Czech Republic (2020). Measures adopted by the Czech Government against the coronavirus. https://vlada.cz/en

W3Techs (2020, July 3). Usage of content languages for websites. W3Techs.com. https://w3techs.com/technologies/overview/content_language

Wang, C., Pan, R., Wan, X., Tan, Y., Xu, L., Ho, C. S., & Ho, R. C. (2020). Immediate psychological responses and associated factors during the initial stage of the 2019 coronavirus disease (COVID-19) epidemic among the general population in China. International Journal of Environmental Research and Public Health, 17(5), 1729.

Yu, M., Li, Z., Yu, Z., He, J., & Zhou, J. (2020). Communication related health crisis on social media: a case of COVID-19 outbreak. Current Issues in Tourism, 1–7.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.