

Register in Czech: Designing an MDA-based experimental study*

Anna Marklová, Olga Buchmüller, Christoph Demian, Roland Meyer
and Luka Szucsich

Humboldt University, Berlin, Germany

Abstract

There are no conventionalized ways to investigate the results of multidimensional analysis (MDA) from the perceptual perspective in an experimental setting. An MDA of the Czech corpus Koditex by Cvrček et al. (2020) established eight dimensions of variation based on 122 linguistic features. The first two dimensions, which explain the largest proportion of shared variance, are labeled as 1. dynamic (+)/static (-) and 2. spontaneous (+)/prepared (-). In our study, we investigated if some situational contexts of language use evoke stronger associations with the poles of the two dimensions than others. Furthermore, we aimed to explore the impact of the mode of language use and the properties of the interlocutor on the ratings.

Czech native speakers (n=107) rated various situational contexts on 7-point Likert-like scales representing the MDA-based dimensions. The items were balanced in the formality of the interlocutor's name (Mr. or Mrs. in connection to surname/first name), the interlocutor's gender, and the mode of language use (spoken/written). The statistical analysis uncovered a significant effect of the formality of the interlocutor's name and the mode of language use on the ratings. Using first names and spoken mode resulted in ratings closer to the positive poles of the dimensions. The comparison of individual items showed that some situations, mainly those representing the negative poles of the dimensions, are rated more consistently than others. The results of our study offer insight into how native speakers evaluate situations of language use on the scales of preparedness, subjectivity, and interactivity.

Keywords: register; Czech; language variation; situational context

1. Introduction

Testing findings of data-driven research, such as analysis of language corpora, brings challenges. Results that have been interpreted once can serve as a base for an experiment, but they must be carefully translated into a form approachable to testing subjects.

We aim to test the interpretation of the results of the multidimensional analysis (MDA) of Czech by experimental means. The study we present in this paper is the first step toward large-scale research, which will shed light on whether we can support the results of the MDA of

* Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1412, 416591334.

corpora by using experimental methods. We want to investigate whether speakers of Czech show a preference for utterances with linguistic features which correlate with particular situations in terms of the MDA results. To this end, a pre-study needs to be conducted. Before introducing linguistic features in situations representing, for example, spontaneous speech, we have to predetermine such situations. The present study investigates how Czech speakers evaluate different language usage situations regarding objectivity, subjectivity, preparedness, and spontaneity. Specifically, we test how the properties of interlocutors (gender, formal and informal naming) and the written vs spoken mode of language use influence the perception of language usage situations. We draw conclusions about the influence these factors had on the subjects' ratings of the situations presented in writing in the items. These results may be useful in constructing appropriate situational contexts for language use.

Figure 1 illustrates the goals of the present study and its place in our broader research. The black font, arrows, and borders indicate the portions discussed in the paper, while the grey font, arrows, and borders represent the parts currently being prepared. The investigation of situational context is a crucial initial step in our efforts to understand native speakers' intuitions about linguistic features. For example, in order to study the correlation between 'dynamic' features and 'dynamic' situations, it is necessary to first determine which situations are perceived by native speakers as 'dynamic.' Since we plan a forced-choice study, we cannot base it only on assumptions and introspection about situational contexts of language use.

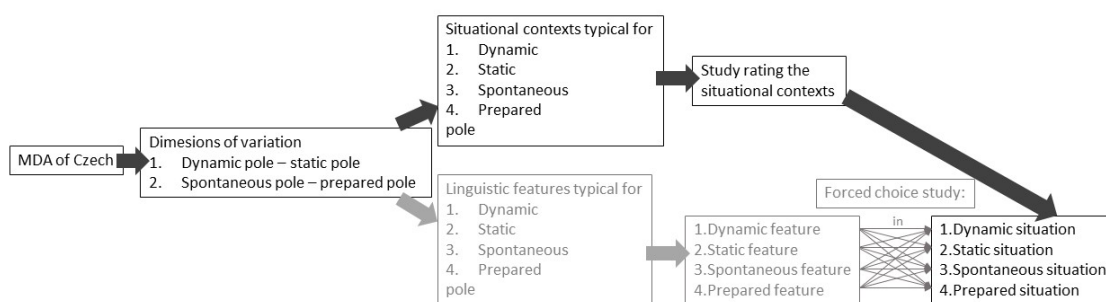


Figure 1: Scheme of the workflow

The linguistic variability which functionally contributes to the text composition is our broader focus. It has been a center of attention of the methodology developed by Douglas Biber (1988), which aims to interpret the variability according to several dimensions of variation, which then point out clusters of texts that are similar in those characteristics. Such clusters of texts are called registers. Any given register can be more or less well-defined concerning the dimensions (Biber et al. 2020).

According to Biber's line of research in the text-linguistic framework, a register is defined as a named, culturally recognized category of texts (Biber 2019). Registers are typically given specific names, such as conversation, classroom teaching, email messages, or newspaper articles. The register perspective focuses on identifying linguistic features that are commonly and extensively found in texts (Biber et al. 2021b:22). Goulart et al. (2020) describe registers as language variations associated with specific combinations of situational characteristics and communicative purposes. These variations often exhibit linguistic similarities (2020:436). We follow the definition by Lüdeling et al. (2022), stating that registers are those aspects of socially-

recurring intra-individual linguistic variation that are influenced by situational and functional parameters. One can analyze registers on different levels of granularity; for example, we can recognize “conversation” as a very general register with few broad characteristics (spoken mode, two or more participants). On the other hand, “a chemistry research article” can be an example of a very specific register. Even that can be more fine-grained by distinguishing the methodology part from the introduction or discussion section since they exhibit different properties. There is no “right” level of granularity for register analysis (Goulart et al. 2020).

According to Biber (1988:20), “Given that the linguistic variation among texts comprises several dimensions, it is no surprise that the relations among texts must be conceptualized in terms of a multi-dimensional space.” In other words, a single dimension is insufficient to capture the full range of language variation, and a multidimensional approach is necessary to examine the various scales of language feature usage exhibited in speech and text. These dimensions are characterized by continuity, i.e., they are not discrete entities but rather scales with two opposite poles. The co-occurrence patterns that underlie these dimensions are identified empirically rather than being based on a priori functional assumptions.

Our focus in this study is on the Czech application of the multidimensional approach. Table 1 presents the dimensions of variation identified by Cvrček et al. (2020) for the Czech language.

Table 1: *Czech dimensions of variation*

Dimension	Positive pole	Negative pole
1	dynamic	static
2	spontaneous	prepared
3	higher level of cohesion	lower level of cohesion
4	polythematic	monothematic
5	higher amount of addressee coding	lower level of addressee coding
6	general/intension	particular/extension
7	prospective	retrospective
8	attitudinal	factual

In this study, we focused on the first two dimensions, as they explain the largest portion of variation in the Koditex corpus (Zasina et al. 2018) on which the MDA was conducted. The labels of the dimension poles are based on an interpretation of the features that are most strongly associated with these poles and the text types that tend to cluster around them. Our first study is grounded in this interpretation.

How can we transfer the concept of the MDA-based dimensions into items, ratable in an experiment? Since the dimensions represent scales, interpreted on the basis of which registers and linguistic features cumulate on their opposite poles, we have translated the poles into situations typical for the text types represented there. For example, the spontaneous pole of the second dimension is represented mainly by interactive spoken communication, private correspondence, and interactive web communication. In contrast, the prepared pole is occupied by administrative texts and scientific literature (Cvrček et al. 2020:95). We used these findings for simulating situations that exhibit the property dominant for the particular pole, e.g., for the spontaneous pole, we presented a situation of two flatmates chatting in the living room, while for the prepared pole, we introduced a situation of a person writing a Wikipedia article.

The first dimension was challenging to translate into ratable scales since the piloting of our initial design had uncovered that the native speakers did not understand what was meant by “Is this situation of language use rather dynamic or static?”. In order to transform the poles of the first dimension into comprehensible questions, we examined their detailed interpretation in Cvrček et al. (2020). The preference for verb expressions over nominal expressions defines the dynamic pole. Text categories grouped at this pole are various narrative novels, private correspondence, or web forums (Cvrček et al. 2020). They have narrative and reflective characteristics; verbs of thinking are common in web forums and correspondence (Cvrček et al. 2021). The preparedness of the text does not seem to play a role, while subjectivity and interactivity seem connected to the positive pole. On the other hand, nouns and adjectives crowd the static pole in various functions, reflecting the strategy of the speaker or writer to elaborate the clause members. Official documents, science papers, encyclopedias, and other official documents dominate it. The objective perspective and text types that appear without interaction seem to be associated with this pole (Cvrček et al. 2021). Based on the description of the poles of the first dimension, we created two rating scales. We asked subjects to rate the situations in terms of subjectivity/objectivity and interactivity, as these properties accurately describe the text types that are gathered at the poles of this dimension.

The second dimension of variation was easier to incorporate into our experimental design, as its poles (spontaneous versus prepared) are more comprehensible to participants. The spontaneous pole is associated with online production under time constraints, as reflected in the presence of contact expressions, fillers, non-dropped pronouns, and Common Czech (cf. Sgall et al. 1992) forms. The text category most strongly represented at this pole is spoken interaction, followed by private correspondence and web forums. The prepared pole is characterized by features such as a high inventory of prepositions, adverbs, lexical richness, or longer words. It is occupied by text categories such as prepared monologues, administrative texts, economic news, and Wikipedia articles (Cvrček et al. 2021). Therefore, we asked the following question: "Is this language use situation more spontaneous or more prepared?"

We constructed all items uniformly, showing the same structure since we expected that the mode and the interlocutor's name (Mr./Mrs. in connection to surname, we call it a formal name, and first name, which we call informal name) would have an impact on the ratings. We also balanced the items in the interlocutor's gender (see Stimuli). We formulated the following research question: Do the mode of language use and the name and gender of the interlocutor affect the ratings of situations of language use?

We anticipated that the mode of the language use and the interlocutor's name would play a role as an influential factor, while the gender of the interlocutor would not have an influence. We did not have any reason to assume that the gender of the interlocutor would affect evaluations of the situations. Additionally, we were interested if the participants' ratings fit our classification of the items into categories dynamic, static, spontaneous, and prepared.

This paper first presents the theoretical framework for situational contexts of language use, then explains our methodology. The experimental procedure is described in detail. Then, the results are presented. In the discussion, we summarize our main findings and draw implications for further research.

2. Situations of language use

Variability is an inherent property of language. We choose linguistic means appropriately for situations we engage in, and these means can differ significantly. A person speaks differently towards a friend, a teacher, or an employer and formulates a formal letter, a postcard, or a scientific article differently. The variability is inherent to all layers of language. An example of a phonetic feature that correlates with particular language situations is the Czech suffix ‘-ý’ in the masculine singular adjectives (*mladý muž* ‘young man’, *starý muž* ‘old man’). It competes with its variant *-ej* (*mladej muž* ‘young man’, *starej muž* ‘old man’). While the former variant appears in written texts and formal speech, we expect the latter variant to be more common in informal spoken situations. Examples of a syntactic feature asymmetrically distributed in language situations are clusters of two or more adjectives, which are typical for prepared texts and written language rather than for spontaneous, interactive speech (Cvrček et al. 2020).

To select appropriate linguistic features, it is necessary for a speaker to accurately assess the specific context in which language is being used. Studies have been conducted on various situational factors, such as formality, the familiarity between speakers, and the mode of language form, to determine which variables are crucial in identifying language situations (see, e.g., Agha 2006; Biber et al. 2021a; Egbert and Gracheva 2022; Pescuma et al. 2023; Sharoff 2018; Wiese 2020). It is worth mentioning that the approaches toward the relationship between registers and language situations differ. While Egbert and Gracheva (2022:4) state that “it is well-established that situational characteristics vary between texts from different registers,” Biber et al. (2021a) criticize the lack of a methodological approach toward situational categorization, which has been primarily binary instead of continuous, and it expects that one register fits one situational context.

The text-linguistic framework focuses on the situational and lexico-grammatical characteristics of language use when describing registers (Biber et al. 2020). There is a long tradition of recognizing the importance of both situational and linguistic characteristics in accurately describing text categories (De Beaugrande and Dressler 1981; Halliday and Hasan 1976; Hymes 1974). The communicative function is crucial in understanding the correlations between situational and linguistic elements, particularly in sociolinguistic research (Hymes 1974). In the text-linguistic approach, there is a three-way relationship between situational context, communicative function, and linguistic form. “Text-linguistic register analysis begins with analysis of the situational characteristics of the register, including consideration of participant identities, relations among participants, channel, production circumstances, setting, and communicative purposes” (Biber et al. 2020:583). External indicators, such as the physical context (time and place) and other considerations (Biber and Conrad 2009), can signal the presence of certain register categories. For example, lectures are often indicated by the location (an auditorium or lecture hall) and the speaker standing in front of the audience. Political rallies can be indicated by the speaker's formal clothing and use of a microphone, standing on a stage with political banners behind them. The medium in which written texts are produced can also indicate registers, such as magazines printed on glossy paper versus academic journals on thicker paper with a table of contents. Additionally, texts within the same register are expected to share “other situational characteristics relating to interactivity, personal involvement, production

circumstances, and the relations among participants" (Biber et al. 2020: 584). Situational characteristics are considered more fundamental than linguistic characteristics (Biber and Conrad 2009) and their identification precedes the selection of linguistic features. It is not possible to infer the situational context in which language is used solely from linguistic phenomena. Linguistic features are argued to be functional, occurring in a register because they suit the purposes and situation (Biber and Conrad 2009). In a conversational situation, for example, one begins to use language with the appropriate linguistic features for that type of conversation.

It is worth noting that the differentiation of registers is not culturally universal; different cultures may have different ways of dividing the range of language use situations. Some cultures may view minor differences as indicative of distinct registers, while others may consider the same range of communicative events to belong to a single register (Biber and Conrad 2009). Biber and Conrad (2009) thoroughly examined situational factors relevant to the description of registers. They introduced seven categories of situational characteristics: participants, relationships among participants, channel, production circumstances, setting, communicative purposes, and topic. Our condensed description of language use situations aims to encompass most of these characteristics. For example, the characteristics of the situation *Tereza mluví se spolubydlicí v obývacím pokoji*. 'Tereza is talking with a roommate in the living room' are shown in Table 2.

Table 2: Situational characteristics for the situation *Tereza mluví se spolubydlicí v obývacím pokoji*.

Participants:	Single addressor; Single addressee; No explicit on-lookers
Relations among participants:	Interactivity; Implicit power equality; Familiarity
Channel	Spoken mode; Face-to-face transient speech
Production circumstances:	Real-time
Setting:	Time and place of communication shared by participants; Relatively private place of communication
Communicative purposes:	NA
Topic:	NA

We left the topic vague because the purpose of the items is to serve as a backdrop for various linguistic features. As such, it is essential to avoid constraining the context too narrowly.

Biber et al. (2020) developed an analysis of underlying situational dimensions of variation. They modified the method of MDA of linguistic features to measure situational variation. Their goal was to identify the co-occurrence patterns among situational variables. They used a 2-factor solution since it accounts for 36% of the shared variance in their data. The two dimensions resulting from the analysis share similarities with the first two dimensions identified by the Czech MDA regarding the text types occurring on their poles. The two first dimensions of variation, and especially the first one, seem to share similarities across the MDA performed on different languages (Biber 1988 for English; Biber and Hared 1992 for Somali; Cvrček et al. 2020 for Czech; Katinskaya and Sharoff 2015 for Russian web corpora; Kim 1994 for Korean). No dimension in these languages defines an absolute dichotomy between speech and writing. Multidimensional studies have repeatedly shown that physical mode is inadequate in itself to account for the relations among registers in a language. However, each language has dimensions closely associated with speech and writing. These dimensions typically isolate

spoken registers at one extreme and written registers at the other extreme, with registers from both modes overlapping in the middle (Biber 2009).

Table 3: Co-occurring situational variables on Biber's dimensions of situational variation (Biber et al. 2020: 594)

	Dimension 1		Dimension 2	
	+	-	+	-
	Personal opinionated discourse	Technical information supported with evidence	Narrative, entertaining discourse	Other communicative purposes (explanatory, advice, or procedural discourse)
The text is:	interactive	pre-planned and edited	a spoken transcript; lyrical or artistic	
The author:	focuses on self; assumes personal knowledge about self	is an expert; assumes technical background knowledge	assumes cultural social knowledge	
The purpose:	persuade the reader; entertain the reader; give advice or recommendations; express opinion	explain information	narrate past events; entertain the reader	explain information; give advice or recommendations; provide how-to instructions
The basis of information:	common knowledge; opinion; personal experience	factual scientific evidence	direct quotes	

Different social cues can lead to the assessment of a situation as either more or less formal. For example, visual cues can influence our perception of a speaker (Rutter 1984). Morand (1995) discusses the behavioral and contextual codes of formality and informality, including the use of honorifics and first names as linguistic elements signaling formality or informality in speech. The formality of the situation plays a role in the choice of linguistic features and, as we assume, in assessing the suitability of the features for a particular situation. Formality is defined by setting (private vs. public), topic and communicative purpose, and the relationship between participants (Szmrecsanyi and Engel 2022). We assumed that formal names imply different characteristics of the speaker and, therefore, of some situational characteristics than the informal name. For example, *Tereza* might evoke other social characteristics, such as younger age than *Paní Novotná* 'Mrs. Novotná'. It can also influence the perception of relations among participants (friends, colleagues, familiar with each other, etc.) and settings (formal name might suggest a less private setting).

Based on the findings described above, it was hypothesized that the formality of the interlocutor's name and the mode of communication would impact the ratings. We expected that written texts would be perceived as more prepared and less interactive than spoken forms and that the use of the formal name would be associated with a more objective, prepared, and less interactive environment than the informal name. Furthermore, we examined if the gender of the interlocutor impacts the ratings. Finally, we explored the ratings of individual items to identify the most consistently rated ones, which would serve as the situational context in the subsequent forced choice study.

3. Methodology

Since the MDA dimensions are labeled by two terms, which can be located on two opposite ends of a scale, e.g., dynamic-static, we have created descriptions of various situations assigned to the respective poles of those scales. The participants rated the situations on 7-point Likert-like scales representing the MDA-based dimensions. The experiment was conducted online, using the PCIBex Farm's interface (Zehr and Schwarz 2022), and it was 15-25 minutes long, depending on personal progress.

3.1. Participants

The sample for this study consisted of 107 native Czech speakers who were recruited through an online platform. Their ages ranged from 18 to 45. The participants were compensated with a financial reward of 5.5 euros for their participation, which took place between May and September 2022.

3.2. Stimuli

The battery of stimuli contained four kinds of situations for evaluation. The categories follow two poles of the first two dimensions of variation of Czech.

1. Dynamic situations (n=12): situations we expected to be rated as interactive and subjective rather than without interaction and objective. It contained situations as *Pan Svoboda si povídá s přítelkyní v útulné kavárně* 'Mr. Svoboda is talking with (his) girlfriend in a cozy cafeteria', or *Tereza vypráví kamarádce veselou příhodu* 'Tereza is narrating to (a) girlfriend a funny story'.
2. Static situations (n=12): situations we expected to be rated as without interaction and objective than as interactive and subjective. An example of such a situation is *Pan Novotný prezentuje šéfové měsíční progres* 'Mr. Novotný is presenting to (his) boss monthly progress', or *Petra píše kolegovi pracovní postup* 'Petra is writing to (her) colleague a work procedure'.
3. Spontaneous situations (n=12): situations we expected to be rated as rather spontaneous than prepared. This category contains situations as for example *Petr čte esemesku od své přítelkyně* 'Petr is reading an SMS from his girlfriend' or *Paní Kučerová píše zprávu do skupinového chatu* 'Mrs. Kučerová is writing a message to the group chat'.
4. Prepared situations (n=12): situations which we expected to be rated rather as prepared than spontaneous. An example is *Paní Dvořáková hlásí zprávy v dopravním zpravodajství* 'Mrs. Dvořáková is announcing news in the traffic news' or *Jitka vypráví návštěvníkům historii hradu* 'Jitka is narrating to visitors the history of the castle'.

The items followed the same syntactic structure and were balanced in terms of mode, gender, and name of the interlocutor. Table 4 presents the structure, and Table 5 lists these elements in the four categories. Each category contained six written and six spoken situations of language use, six formal and informal forms of the interlocutor name, and six male and six female interlocutors. First names and surnames were chosen from the 50 most frequent Czech names

separately for men and for women (Samek and Malačka 2011). We chose such names, which we assessed as neutral; Czech has a rich derivation of diminutives, and it is common to alternate the official form into a hypocoristic. Some widely used hypocoristics have lost their expressive function, and they compete with the official name variant (Pleskalová 2017). It is connected with the loss of emotional expressiveness of some suffixes, for example, *-ek* for male names and *-ka* for female names (Knappová 2017). E.g., *Hana* is alternated with *Hanka*, and the form *Hanka* dominates spoken registers (Cvrček and Vondříčka 2011). Therefore, we chose names that do not sound expressively marked in their official form, such as *Eva*, *Michal*, *Patrik*, etc.

Table 4: *The structure of the experimental items*

interlocutor	action	object	connectors	specifier	localization
<i>Pan Novák</i> 'Mr. Novák	<i>vypráví</i> is narrating	<i>dětem</i> (to) children		<i>napínavý</i> (a) thrilling	<i>příběh</i> story'
<i>Jan</i> 'Jan	<i>mluví</i> is talking	<i>se sestřenicí</i> with cousin	<i>na</i> at	<i>rodinné</i> (the) family	<i>oslavě</i> celebration'
<i>Michal</i> 'Michal	<i>píše</i> is writing	<i>komentář</i> (a) comment	<i>pod</i> under	<i>facebookovým</i> (a) Facebook	<i>příspěvek</i> post'

Table 5: *Distribution of the mode, interlocutor's gender, and name of the interlocutor (f=formal, i=informal) throughout our stimuli battery categories*

	dynamic		static		spontaneous		prepared	
	female	male	female	male	female	male	female	male
written	2f 1i	1f 2i	1f 2i	2f 1i	2f 1i	1f 2i	1f 2i	2f 1i
spoken	1f 2i	2f 1i	2f 1i	1f 2i	1f 2i	2f 1i	2f 1i	1f 2i

For written mode, we have two versions of items, one version with the action verb *čte* 'read' and one with the action verb *píše* 'write'. For spoken mode, we used various verbs of speaking, as *povídá* 'tells', *mluví* 'says', *říká* 'says', *vypráví* 'narrates', etc.

3.3. Procedure

The participants were instructed to conduct the experiment on a laptop or computer, not on a phone or a smartphone. The scheme of the experimental procedure is illustrated in Figure 2. After confirming an informed consent, an elaborated description instructed the participants about the terms which would appear in the study. They had an example item for training. There were also three control questions throughout the experiment which served as a check for the participants and for the examiners to assess if the task was understood correctly. We included the questions since the piloting suggested that the participants become confused about what exactly they should rate in the experiment. The questions asked about the understanding of *užití jazyka* 'language use' and always revealed the correct answer after the participant has chosen, as in the following example:

- (1) *Mr. Soukup is talking with an accountant in the city bank.*
 The “usage of language” is, in this case:
 (a) *Conversation between Mr. Soukup and the accountant*
 (b) *The thoughts of Mr. Soukup about the conversation*

The control questions proved to be an excellent tool for avoiding misunderstandings.

The following instruction explaining the terminology, translated from Czech into English, was provided in written form at the beginning of the experiment: “Your task will be to read sentences describing different language use situations and answer questions about these sentences. The questions will ask for your opinion, and there are no wrong or correct answers. Please do the study in a calm environment, read carefully and answer questions quickly, without much contemplation. It is about your intuition and first impression.”

After the instruction, an elaborated explanation of the terms ‘language use’, ‘subjective’, ‘objective’, ‘spontaneous’, ‘prepared’, ‘interactive’, and ‘without interaction’ followed.

The items were presented in randomized order.

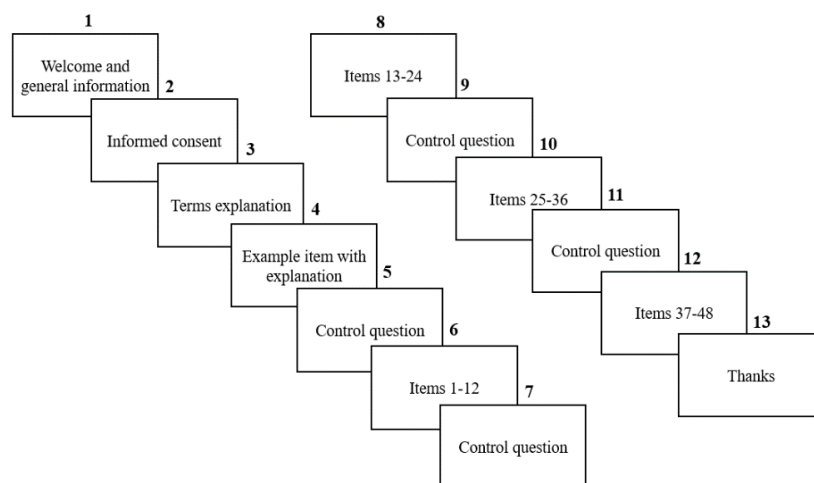


Figure 2: Chronological order of the experimental procedure

3.4. Results

Three independent variables were statistically measured: the effect of the interlocutor’s name, the interlocutor’s gender, and the mode of language use.

3.4.1. Effect of properties of the interlocutor

Two groups of stimuli were established: a formal group, which consisted of descriptions with the surname form of the interlocutor’s name, and an informal group, which consisted of descriptions with the first name form of the interlocutor’s name. The mean rating on the scales in the formal group was 4.419132, and in the informal group, 4.008462. A linear regression model with a single predictor variable (R Core Team 2022) was used to assess the effect of interlocutor name on the ratings. The model aimed to investigate the relationship between the response variable “value of the rating” and the predictor variable “interlocutor’s name.” The results of the linear regression analysis showed that the predictor variable “interlocutor’s name”

was significantly related to the response variable "value of the rating" ($p < 0.001$). The results of the linear model are displayed in Table 6. The results suggest that "interlocutor's name" is a significant predictor of "value of the rating." However, further investigations are needed to establish other factors influencing the "value of the rating."

Table 6: Linear regression model assessing the role of the name of the interlocutor as a predictor of rating

Residuals:				
Min	1Q	Median	3Q	Max
-3.4191	-2.0085	-0.0085	1.9915	2.9915
Coefficients:				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.41913	0.02638	167.5	<2e-16 ***
Informal interlocutor	-0.41067	0.03763	-10.91	<2e-16 ***
Residual standard error: 2.103 on 12499 degrees of freedom				
Multiple R-squared: 0.009439, Adjusted R-squared: 0.009359				
F-statistic: 119.1 on 1 and 12499 DF, p-value: < 2.2e-16				

A linear regression analysis was conducted to examine the effect of interlocutor's gender on the ratings. The study used Czech first names and surnames that marked gender, and the analysis included two groups: a group of male interlocutors and a group of female interlocutors. The mean rating on the scales was 4.229769 for male interlocutors and 4.204846 for female interlocutors. The linear regression model did not find interlocutor's gender to be a significant predictor of the value of the ratings.

3.4.2. Effect of mode

The linear regression model was used for assessing the effect of the mode of the language use on the ratings. The Spoken group consisted of the descriptions containing the verb of speaking (*říká* '(s/he) is saying', *povídá* '(s/he) is saying', *mluví* '(s/he) is talking', etc.) and the Written group contained descriptions with verbs 'write' or 'read' (*píše* '(s/he) is writing', *čte* '(s/he) is reading'). The model aimed to investigate the relationship between the response variable "value of the ratings" and the predictor variable "mode." It found a significant relation between the predicative variable "mode" and the response variable "value". The results of the analysis are presented in Table 7.

Table 7: Linear regression model assessing the role of mode as a predictor of rating

Residuals:				
Min	1Q	Median	3Q	Max
-3.5524	-1.8788	0.1212	2.1212	3.1212
Coefficients:				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.41913	0.02638	167.5	<2e-16 ***
Written mode	0.67361	0.03732	18.05	<2e-16 ***
Residual standard error: 2.086 on 12499 degrees of freedom				
Multiple R-squared: 0.0254, Adjusted R-squared: 0.02532				
F-statistic: 325.8 on 1 and 12499 DF, p-value: < 2.2e-16				

3.4.3. Ratings of individual items

Figure 3 visualizes items that we predicted as spontaneous (A) and prepared (B) on the spontaneous-prepared scale, where one equals the spontaneous pole and seven the prepared pole. Figure 4 displays predicted dynamic (A; C) and static (B; D) items on the two scales representing dynamic and static poles of the first dimension: subjective-objective scale (A; B) and interactive-non-interactive scale (C; D). Subjective and interactive poles equal one, and objective and non-interactive poles equal seven. We searched for items consistently rated by values near the poles of the dimensions.

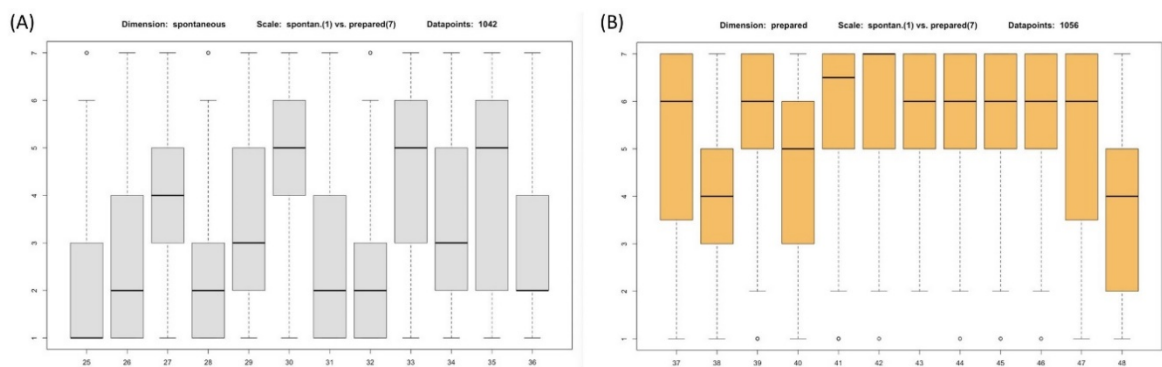


Figure 3: Results of the second-dimension scale: (A) Spontaneous items displayed on the scale spontaneous (1) – prepared (7); (B) prepared items displayed on the same scale. The y axis displays the rating scale (1-7); each number on the x axis refers to one experimental item.

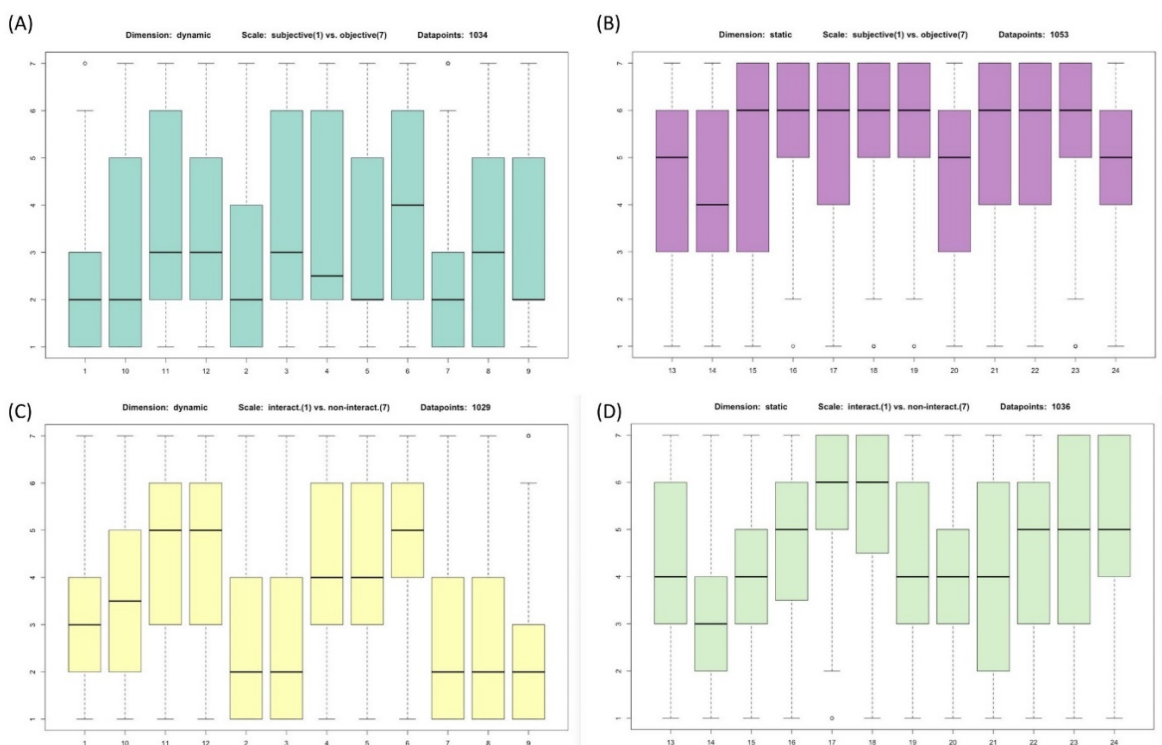


Figure 4: Results of the second-dimension scale: (A) Dynamic items displayed on the scale subjective (1) – objective (7); (B) Static items displayed on the same scale; (C) Dynamic items displayed on the scale interactive (1) – without interaction (7); (D) Static items displayed on the same scale. The y axis displays the rating scale (1-7); each number on the x axis refers to one experimental item.

The visual inspection uncovered several patterns: objective and prepared stimuli groups were rated most consistently, following our predictions about the items. In general, the ratings were more consistent on negative poles (non-interactive, objective, and prepared) than on positive poles (interactive, subjective, and non-interactive). In Figure 3, we present each group displayed on the respective scale (e.g., prepared group on the spontaneous-prepared scale, not on subjective-objective scale), but the participants rated all items at all three scales. We observed that when an item is placed on the negative pole of one scale, it tends to be rated on the negative pole also on the other scales (i.e., items rated as prepared tend to be also rated as objective and non-interactive). The same holds for the positive pole.

4. Discussion

This study aimed to uncover factors influencing people's evaluation of various language use situations represented by written descriptions. The findings indicate that using an in/formal form of the interlocutor's name performing a linguistic action affects the evaluation of the situation. Using a surname with the specifier Mr. or Mrs. leads to evaluating situations as more objective, prepared, and without interaction than using a first name, which we found statistically significant. On the other hand, the gender of the subject did not show a significant observable effect. We can connect this finding to the previous knowledge about Czech names: first names are used chiefly in informal situations. At the same time, surnames are associated with more formal settings. This association seems activated when the name is used to label an interlocutor in our situation descriptions. Given the considerable variability within forms of Czech names, it would be interesting to investigate if similar tendencies occur in the context of other languages.

Mode of language use plays a role as well. Spoken mode predicts ratings on the positive pole of the scales; thus, spoken situations tend to be rated as more dynamic and spontaneous than written situations, which are connected to static and prepared interpretations. There was a significant difference found between the spoken mode and written mode in ratings. It suggests that even though there are some fairly unprepared written situations, such as chatting on WhatsApp or commenting on a Facebook post, the spoken situations are more prototypical examples, and they evoke spontaneity and dynamicity to a higher extent. Similarly, prepared and static situations can occur in spoken modalities, such as a presentation at a company meeting. However, the prototypicality is connected to the written situations.

It is important to note that although the regression models for mode and interlocutor's name yielded significant results, the effect size observed was relatively small. Therefore, further research is necessary to validate the identified influences and examine other factors that might impact the ratings. The results of our analysis suggest that some items successfully simulated situations of language use as we predicted, while others did not. The data only partially supported our predictions about the nature of the items. This outcome is unsurprising, given that we aimed to balance the representation of different modalities, name forms, and gender in each group. It is worth noting that the situations were rated more consistently in terms of being prepared, non-interactive, and objective compared to spontaneous, interactive, and subjective.

Another pattern that emerged was that the poles of static-prepared and dynamic-spontaneous overlap, with items rated higher on one scale also tending to be rated higher on the others.

Our central objective is to utilize experimental tools to examine the perception of situations determined by the interpretation of dimensions that resulted from the Czech MDA conducted by Cvrček et al. (2020). The chosen labels employed to interpret the findings of the Czech MDA (e.g., static or dynamic dimension pole) are inherently simplifying in nature. That is inevitable in order to encapsulate the extensive array of linguistic features and cumulative text types on opposite ends of each dimension within a single word. In our preliminary study, our specific focus was on investigating the poles of the two Czech dimensions of variation. We established the assessment scales based on a detailed description of the dimension poles and the text types associated with them (Cvrček et al. 2020). This approach enables us to present particular linguistic features in the situational context with higher certainty that the contexts reflect the outcomes of the Czech MDA. However, we must be careful when applying the same research design to other languages.

Overall, the results of this study provide insights into situations of language use and how different characteristics, such as the mode of language production or attributes of an interlocutor, can affect perceptions of these situations. These findings will be used in a forced-choice study where the appropriateness of particular linguistic features relative to situations will be investigated. They can serve as a resource while constructing situational contexts in empirical research.

It is important to acknowledge that our study can only reveal the perception of the provided descriptions of the situations. Due to the inherent limitations and artificiality of online experiments, it is not possible to fully replicate real-life situations as participants would experience them. However, for future research that aims to simulate real-life situations, it is crucial to have consistent assessments of the descriptions from native speakers. Our study offers a method of obtaining such situations and provides insights into the factors that influence these assessments.

It should be noted that it is impossible to examine all potential factors that may influence the assessments in one study. Specifically, the topic of conversation may also impact the perception of the appropriateness of linguistic features used by a speaker, which should be taken into consideration in subsequent studies.

References

- Agha, A. 2006. *Language and Social Relations*. Cambridge: Cambridge University Press.
- Biber, D. 1988. *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- Biber, D. 2009. *Dimensions of Register Variation*. Cambridge, GBR: Cambridge University Press.
- Biber, D. 2019. Text-Linguistic Approaches to Register Variation. *Register Studies* 1(1):42–75. doi: 10.1075/rs.18007.bib.
- Biber, D., and S. Conrad. 2009. *Register, Genre, and Style*. Cambridge: Cambridge University Press.
- Biber, D., J. Egbert, and D. Keller. 2020. Reconceptualizing Register in a Continuous Situational Space. *Corpus Linguistics and Linguistic Theory* 16(3):581–616. doi: 10.1515/cllt-2018-0086.
- Biber, D., J. Egbert, D. Keller, and S. Wizner. 2021a. Chapter 2. Extending Text-Linguistic Studies of Register Variation to a Continuous Situational Space: Case Studies from the Web and Natural Conversation. In E.

- Seone, and D. Biber (eds.), *Studies in Corpus Linguistics*, 19–50. Amsterdam: John Benjamins Publishing Company.
- Biber, D., Jesse E., Daniel K., and S. Wizner. 2021b. Towards a Taxonomy of Conversational Discourse Types: An Empirical Corpus-Based Analysis. *Journal of Pragmatics* 171:20–35. doi: 10.1016/j.pragma.2020.09.018.
- Biber, D., and M. Hared. 1992. Dimensions of Register Variation in Somali. *Language Variation and Change* 4(1):41–75. doi: 10.1017/S095439450000065X.
- Cvrček, V., Z. Komrsková, D. Lukeš, P. Poukarová, A. Řehořková, and A. J. Zasina. 2021. From Extra- to Intratextual Characteristics: Charting the Space of Variation in Czech through MDA. *Corpus Linguistics and Linguistic Theory* 17(2):351–82. doi: 10.1515/cllt-2018-0020.
- Cvrček, V., Z. Laubeová, D. Lukeš, P. Poukarová, A. Řehořková, and A. J. Zasina. 2020. *Registry v češtině*. Praha: NLN.
- Cvrček, V., and P. Vondříčka. 2011. SyD – Korpusov Průzkum Variant. Praha: FF UK. URL: <http://syd.korpus.cz>.
- De Beaugrande, R., and W. Dressler. 1981. *Introduction to Text Linguistics*. London: Longman.
- Egbert, J., and M. Gracheva. 2022. Linguistic Variation within Registers: Granularity in Textual Units and Situational Parameters. *Corpus Linguistics and Linguistic Theory* 0(0). doi: 10.1515/cllt-2022-0034.
- Goulart, L., B. Gray, S. Staples, A. Black, A. Shelton, D. Biber, J. Egbert, and S. Wizner. 2020. Linguistic Perspectives on Register. *Annual Review of Linguistics* 6(1):435–55. doi: 10.1146/annurev-linguistics-011718-012644.
- Halliday, M. A. K., and R. Hasan. 1976. *Cohesion in English*. London: Longman.
- Hymes, D. 1974. *Foundations in Sociolinguistics: An Ethnographic Approach*. Philadelphia: University of Pennsylvania Press.
- Katinskaya, A., and S. Sharoff. 2015. Applying Multi-Dimensional Analysis to a Russian Webcorpus: Searching for Evidence of Genres. Hissar, Bulgaria.
- Kim, Y.-J. 1994. A Corpus-Based Analysis of Register Variation in Korean. In D. Biber, and E. Finegan (eds.), *Sociolinguistic Perspectives on Register*, 157–81. Oxford: Oxford University Press.
- Knappová, M. 2017. Rodné jméno. In P. Karlík, M. Nekula, and J. Pleskalová (eds.), *Nový Encyklopedický Slovník Češtiny*. URL: https://www.czechency.org/slovník/RODNÉ_JMÉNO.
- Lüdeling, A., A. Alexiadou, A. Adli, K. Donhauser, M. Dreyer, M. Egg, A. H. Feulner, N. Gagarina, W. Hock, S. Jannedy, F. Kammerzell, P. Knoeferle, T. Krause, M. Krifka, S. Kutscher, B. Lütke, T. McFadden, R. Meyer, C. Mooshammer, S. Müller, K. Maquate, M. Norde, U. Sauerland, S. Solt, L. Szucsich, E. Verhoeven, R. Waltereit, A. Wolfgruber, and L. E. Zeige. 2022. *Register: Language Users' Knowledge of Situational-Functional Variation*. doi: 10.18452/24901.
- Morand, D. A. 1995. The Role of Behavioral Formality and Informality in the Enactment of Bureaucratic versus Organic Organizations. *The Academy of Management Review* 20(4):831–72. doi: 10.2307/258958.
- Pescuma, V. N., D. Serova, J. Lukassek, A. Sauermann, R. Schäfer, A. Adli, F. Bildhauer, M. Egg, K. Hülk, A. Ito, S. Jannedy, V. Kordoni, M. Kuehnast, S. Kutscher, R. Lange, N. Lehmann, M. Liu, B. Lütke, K. Maquate, C. Mooshammer, V. Morteza pour, S. Müller, M. Norde, E. Pankratz, A. G. Patarroyo, A.-M. Pleşca, C. R. Ronderos, S. Rotter, U. Sauerland, G. Schnelle, B. Schulte, G. Schüppenhauer, B. M. Sell, S. Solt, M. Terada, D. Tsiapou, E. Verhoeven, M. Weirich, H. Wiese, K. Zaruba, L. E. Zeige, A. Lüdeling, and P. Knoeferle. 2023. Situating Language Register across the Ages, Languages, Modalities, and Cultural Aspects: Evidence from Complementary Methods. *Frontiers in Psychology* 13:964658. doi: 10.3389/fpsyg.2022.964658.
- Pleskalová, J. 2017. Hypokoristikum. In P. Karlík, M. Nekula, and J. Pleskalová (eds.), *Nový Encyklopedický Slovník Češtiny*. URL: <https://www.czechency.org/slovník/HYPOKORISTIKUM>.
- R Core Team. 2022. *A Language and Environment for Statistical Computing*. Vienna, Austria.
- Rutter, D. R. 1984. *Looking and Seeing: The Role of Visual Communication in Social Interaction*. Chichester; New York: Wiley.
- Samek, O., and J. Malačka. 2011. *KdeJsm.cz | Četnost příjmení nebo jména v České republice*. Retrieved December 30, 2022 (<https://www.kdejsme.cz/>).
- Sgall, P., J. Hronek, A. Stich, and J. Horecký. 1992. *Variation in Language: Code Switching in Czech as a Challenge for Sociolinguistics*. John Benjamins Publishing Company.
- Sharoff, S. 2018. Functional Text Dimensions for the Annotation of Web Corpora. *Corpora* 13(1):65–95. doi: 10.3366/cor.2018.0136.

- Szmrecsanyi, B., and A. Engel. 2022. A Variationist Perspective on the Comparative Complexity of Four Registers at the Intersection of Mode and Formality. *Corpus Linguistics and Linguistic Theory* 0(0). doi: 10.1515/cllt-2022-0031.
- Wiese, H. 2020. *Language Situations: A Method for Capturing Variation within Speakers' Repertoires*. in *Methods of Dialectology*, Vol. XVI. Frankfurt a. M.
- Zasina, A. J., D. Lukeš, Z. Komrsková, P. Poukarová, and A. Řehořková. 2018. *Koditex: Korpus Diverzifikovaných Textů*. Praha: Ústav Českého národního korpusu FF UK.
- Zehr, J., and F. Schwarz. 2022. *PennController for Internet Based Experiments (IBEX)*. doi: 10.17605/OSF.IO/MD832.