Slovak '*no*' and its pragmatic meanings and functions in relation to prosody

Lucia Mareková, Štefan Beňuš[*]
Constantine the Philosopher University in Nitra, Slovakia

Abstract

This paper examines the contributions of lexical context and prosody on the perception of the Slovak particle '*no*', [nɔ] in IPA. The functional meanings of this discourse marker are similar to those of 'okay' in English. Based on a literature review, we expected that the presence of the prosodic cues is not sufficient for decoding the functional meanings. We also explored how biological sex and age affected the perception of '*no*'. We found that both - context and prosody - to a great extent influence the disambiguation of '*no*', but context provides better information about the function of Slovak '*no*'. Additionally, females and younger adults, compared with males and middle-aged adults, were more sensitive to the cues provided by context than by prosody alone.

Key words

context, prosody, discourse markers, cue phrases, Slovak '*no*'

1. Introduction

This work experimentally analyses the functional perception of the Slovak particle '*no*', [nɔ] in IPA and examines the influence of the lexical context and prosody on this perception.

Slovak '*no*' is a discourse marker with various functional meanings that are similar to those of the English discourse markers 'okay', 'well' and 'now'. Discourse markers signal the speaker's attitudes towards the discussed topic; they signal the speaker's assumptions about the listener's knowledge; they organize, manage or connect what we say; they help in the organization of spoken interactions; e.g. by helping managing turns; they may also signal politeness between speakers and many other communicative functions.

Understanding the nature of the process underlying the coding and decoding of the functional meanings of the discourse marker '*no*' could help us to understand the nature of human dialogue better; it could also bring us closer to the possibility of developing more natural communication in spoken interactions between humans and machines.

Even though '*no*' has multiple prosodic realizations which imply different meanings and functions, people are able to decode its meaning in everyday face-to-face communication with ease. To distinguish between these various meanings of '*no*', people use prosodic and contextual cues. Using particular contextual and prosodic elements, the speaker is able to modify the information and intention s/he wants to convey to his/her communication partner. These prosodic and contextual cues are very important elements of natural human communication not only on the side of the speaker but also on the side of the listener and provide insights into the cognitive system underlying spoken communication.

---

[*] Address for correspondence: Štefan Beňuš, Constantine the Philosopher University, Faculty of Arts, Department of English and American Studies, Štefánikova 67, 949 74 Nitra, Slovakia. E-mail: sbenus@ukf.sk

Understanding the relationship between prosodic and contextual cues thus helps us understand the system of natural spoken communication better. For that purpose, we tested how people perceive the function of 'no' when having access to lexical context, in comparison with how the same people perceive this function when having only prosodic features available. Through these observations we wanted to get closer to the nature of the relationship between contextual and prosodic features of language and explore how these features influence human understanding of the information that speakers want to express by using the particle 'no'. Although context was a better cue than prosody, our results also show that prosody and its aspects are deeply ingrained in our cognitive system and that people are able to identify the meaning of spoken communication also without any lexical context. Regarding biological sex, even though women were not better than men in the interpretation of 'no' according to prosodic cues, we did observe a difference between the sexes in how they used contextual cues. Younger adults showed a similar pattern to females in our results.

The paper is organized as follows: section 2 presents a theoretical overview of discourse markers and cue phrases, their functions and features; it introduces the Slovak particle 'no', its pragmatic meanings and its relation to English discourse markers. Also, in this section we present a theoretical overview of previous studies on context and prosody. Section 3 describes the methodology of the experiment, in which we were interested in the subjects' ability to distinguish between various functions of 'no' when having either only prosodic cues or lexical context available. Section 4 presents the results, and section 5 discusses the findings in relation to the current understanding of the prosody-context relationship. Section 6 lists the limitations of the study.

2. Context and prosody in cue phrases

Natural human communication involves not only separate utterances but also a set of elements signalling relationships among these utterances. These relationships might be cues by various means. For example, an utterance with a lower pitch range than the preceding utterance is likely to be meaningfully linked to that preceding utterance. On the contrary, an utterance with a greater pitch range is likely to indicate some meaning separation from the preceding utterance with a lower pitch range. Cue phrases represent another system of basic elements that provide meaningful relations among utterances.

2.1 General description of cue phrases

Cue phrases (also known as discourse markers or clue words) "function as explicit indicators of the structure of a discourse" (Hirschberg and Litman, 1993, p. 501). These linguistic elements shape natural spoken communication and convey information that a speaker wishes to pass to his/her communication partner. Andersen defines the way in which cue phrases work in communicative acts as follows: "(…) communicative acts involve ostensive behaviour. The utterance interpretation is an inferential process. Successful communication involves intentional behaviour on the part of communicator and it involves recognition of this intentional behaviour on the part of another communicator" (Andersen, 2001, p.14). Cue phrases are one of the means of signalling this intentional behaviour - they are a code standing for the communicator's attitude towards the discussed topics.

Andersen questions the term "discourse markers" and proposes, in his opinion, a more appropriate term- "pragmatic markers": "(…) the label 'pragmatic' is meant to suggest a relatively low degree of lexical specificity and a high degree of context-sensitivity" (Andersen, 2001, p. 40). He argues that cue phrases are "pragmatic" because they contribute to inferential processes (such as enrichment, loosening - very typical functions of the discourse marker 'like') and they also control the choice of the contextual background. For example, the discourse marker 'you know' expresses that the information the speaker is about to say to the addressee might come as something new/surprising to him. Hence, the hearer (addressee) may be forced to reorganize his/her contextual background (Andersen, 2001, p. 73).

2.1.1 Characteristics of cue phrases

Jucker and Ziv introduce Hölker´s (1988) four features of discourse markers: "(1) they do not affect the truth conditions of an utterance; (2) they do not add anything to the propositional content of an utterance; (3) they are related to the speech situation and not to the situation talked about; and (4) they have an emotive, expressive function rather than a referential, denotative, or cognitive function" (Jucker and Ziv, 1998, p.3). Jucker and Ziv (1998, p.3) also introduce Brinton's (1996) list of features of cue phrases.

This definition of cue phrases' features is much more complex and is constructed according to different points of view; Brinton defines:

- **phonological and lexical features**
  - ➢ Cue phrases are short and phonologically reduced;
  - ➢ they form a separate tone group;
  - ➢ they are marginal forms, which are difficult to place within a traditional word class.
- **syntactic features**
  - ➢ Cue phrases are restricted to the sentence-initial position;
  - ➢ they occur outside the syntactic structure;
  - ➢ they are optional.
- **semantic feature**
  - ➢ Cue phrases have little or no propositional meaning.
- **functional feature**
  - ➢ Cue phrases are multifunctional.
- **sociolinguistic and stylistic features**
  - ➢ Cue phrases are a feature of oral rather than written discourse and are associated with informality;
  - ➢ they appear with high frequency;
  - ➢ they are stylistically stigmatized;
  - ➢ they are biological sex specific, more typical of women's speech.

According to these definitions, cue phrases are considered linguistic elements with limited lexical meaning that are difficult to place within a word class and are difficult to translate. However, there are linguistic analyses that object to some aspects of these definitions. We introduce for example Andersen's opinion which is related to Brinton's semantic feature definition (cue phrases have little or no propositional meaning). Andersen argues that "non-propositionality is only partly a valid criterion, because some pragmatic markers can be seen to have truth-conditional implications: "rather than taking non-propositionality for granted, I argue that a more precise description of pragmatic markers is that they guide the hearer in utterance interpretation and constrain the identification of the intended explicit and implicit meaning of an utterance" (Andersen, 2001, pp. 40 - 41).

A single cue phrase might convey several different meanings. This quality is called ambiguity. The majority of existing analyses focus on distinguishing between the sentential and discourse uses of cue phrases. Hirschberg and Litman define this difference using the word 'incidentally'. While this word "may be used sententially as an adverbial, for example, the discourse use initiates a digression" (Hirschberg and Litman, 1993, p. 501). More recent analyses also investigate the disambiguation among several discourse meanings of cue phrases. (Hirschberg and Litman, 1993). In the following example we present the disambiguation among two discourse meanings of 'okay' - acknowledgment (1) and cue beginning (2):

(1) a. Just tell father that I would like to talk to him.
    b. Okay.
(2) a. Okay, it's time to talk about your behaviour.

2.1.2 Functions of cue phrases

The system of meanings associated with cue phrases is very complex. Some functions might be signalled with several cue phrases and some might be particular for a unique cue phrase. For example, Gravano et al. (2007) identify at least 10 functions of 'okay' in American English, such as acknowledgment, agreement, backchannel, cue beginning, pivot, etc. But backchannelling might be also signalled with 'alright' and cue beginning with 'so'. And these two, in turn, might signal functions that 'okay' does not signal. In an attempt to provide a general description of these functions, Brinton (1996) firstly introduced global functions of cue phrases on the textual level: "(a) to mark various kinds of boundaries (to initiate or end a discourse unit or to effect a shift to topic), and (b) to assist in turn-taking in oral discourse or "chunking" (marking of episode or paragraph) in written discourse." Secondly, on the

interpersonal level, Brinton introduced the following two functions: "(a) subjectively, to express speaker attitude, and (b) interactively, to achieve intimacy between speaker and addressee (for example, by appealing to the addressee or expressing shared or common knowledge)" (Brinton, 1996, p. 6).

2.2 Slovak particle *'no'*

Slovak linguistic tradition (e.g. Findra, 2004) defines particles as expressive-evaluative words which are used by a language user to subjectively express his/her relationship to the content s/he is talking about. Mistrík characterizes particles as "par excellence expressive, personal words" (Mistrík, 1997, p.83). He suggests that their usage is more appropriate in subjectively nuanced texts and they are most often used in oral informal communication.

In English, cue phrases are represented by highly used particles like *'okay'*, *'well'*, or *'now'*. In Slovak, *'no'* has very similar functions and shares many pragmatic meanings and functions that are typical for English discourse markers (discussed in 2.1.1 and 2.1.2). Slovak *'no'* is a short form of *'áno'* meaning 'yes', but it also has many other meanings. Beňuš (2012) introduces the synopsis of various *no*-meanings. For example, *'no'* can convey the conjunctive meaning of 'but'; it can also function as a backchannel, as a pivot from one topic to another, or it can signal hesitation.

We present two major characteristics of *'no'* using this example from our corpus:

(3) a. *Čo   je to? Sú   to   nohy?*
       What   is it? Are   it   legs?
       'What is it? Are they legs?'
    b. *No,   vyzerá to   ako   chodidlá.*
       Well, looks   it   like   feet.
       '*Well*, it looks like feet.'

In this example, *'no'* does not affect the truth conditions of an utterance and neither does it add anything to the propositional content of the utterance. In that, 'no' has characteristics introduced in Section 2.1.1 above. However, although these characteristics might be appropriate for written communication, spoken dialogues might be different. Even though *'no'* in this example does not affect the main information that something looks like feet, to say that it does not add anything to the propositional content would be in case of oral communication questionable and subjective. This discourse marker can convey a lot of different meanings - hesitation, uncertainty, agreement - and all of these modify the information that speaker B wants to convey to speaker A. These meanings are channelled through various aspects of prosody or lexical context. If a speaker uses plateau intonation evoking hesitation, that melody modifies the intended meaning. In the case of our example, B's utterance could convey a meaning that the objects on the screen might be feet but it is possible that it is not - the speaker is not sure. This influence of the prosodic form and lexical context will be discussed in the next section.

In our view it is also questionable whether discourse markers like *'no'* should be restricted to the sentence-initial position (the syntactic feature from 2.1.1). Even though this particle is *usually* used in this position, we would like to argue that in some cases of oral informal communication speakers use these particles even in the middle of the sentence. This is rather common especially in communicational situations in which the speaker hesitates about what s/he is going to say. For example:

(4) a. *Zajtra   by   som si   chcela íst' vyzdvihnúť ten,   no,   ten papier.*
       Tomorrow   would I myself   want   go pick up   that, you know,   that paper
       'I would like to go and pick up that, *you know*, that paper tomorrow.'

Therefore, for our purposes, we eliminate the expression 'restricted", which evokes a rule that has to be obeyed since this is not how it works in the spontaneous speech of communicative interactions with this particle.

Findra (2004) states that combining with various elements of contextual background, facial expressions, or gestures, one particle can convey a lot of different meanings. One of the main distinguishing elements between different meanings of discourse markers is the context in which they occur and the prosody with which they are realized. Both of these are helpful in the identification of the

meaning of the discourse marker in a particular communicative situation. The relationship between context and prosody is discussed in the following section.

2.3 The relationship between context and prosody
Speakers using particular contextual and prosodic elements of speech are able to code their intentions in the way in which they want their communication partner to decode them. Prosody and context are also very useful in the disambiguation process on the side of the addressees - they help them to decode the information they perceive and to distinguish between various functional meanings of a particular utterance.

2.3.1 The role of prosody
Prosody plays an important role in cuing the functions and pragmatic meanings of feedback expressions and discourse markers like the Slovak 'no', English 'okay', German 'ja ja', French 'bon', Czech 'jasně', and others (e.g. Gravano et al., 2007; Golato and Fagyal, 2008; Lee et al., 2019; Volín et al., 2015). Cole (2015) provides a comprehensive overview of different aspects of prosody and its influence on the discourse segments. She introduces the ability of prosody to signal discourse meaning as "one of the most salient aspects of prosody" (Cole, 2015, p. 15). Prosody is an important element that communicates the meaning of an utterance and therefore it communicates also the meaning of discourse markers. Based on Cole's study (2015), we assume that prosody can encode, among others:

- The illocutionary force;
- The affective meaning;
- The indication regarding which part of the utterance contains new information.

The term "illocutionary force" is related to the notion "illocutionary act". According to Austin (1975) every utterance performs an illocutionary act which is characterized by a particular illocutionary force - the speaker's intention in producing the speech (e.g. promising, requiring, advising, warning). There are various elements that help speakers to express their intentions (the illocutionary force) - one of them is prosody. However, according to Récanati (1987) "these indications are not always precise, and often the hearer must fall back on his knowledge of the context to determine the exact illocutionary force conveyed by the utterance" (Récanati, 1987, p. 20).

Prosody is also a very useful element of signalling the affective meaning, i.e. the speaker's attitude "related to the interpretation of the utterance" (Cole, 2015, p. 21). For example, Volin et al. (2016), studying the Czech discourse marker 'jasně', 'sure', found out that there are "multi-parametric differences between 'jasně' realizations in terms of their F0, timing and intensity patterns, which gave rise to generally consistent form-function mappings. For example, resignation seems to be realized with a falling intonation contour, relatively slow tempo, long word-initial consonant and a short word-final vowel" (Volín et al., 2016, p. 314).

However, results of studies focusing on the role of prosody in decoding the affective meaning presented by Cole suggest great variability and also inconsistencies regarding the prosody-meaning relationship. Take, for example, the influence of prosody on distinguishing between which information in the utterance is new and which was previously given in context. On the one hand, it seems that "listeners tend to interpret unaccented words as given or accessible based on the prior discourse, while accented words are interpreted as new information" (Cole, 2015, p.19). On the other hand, it seems that there are considerable mismatches which are represented by the possibility that "not all pitch accents mark focus or new information" (Cole, 2015, p. 19). These mismatches tend to be prevalent in spontaneous speech. Hence, the results are not conclusive regarding this prosody-meaning domain.

Considering the results from the above three areas dealing with the influence of prosody on discourse meaning, we can say that the role of prosody in the disambiguation of discourse is rather complex. The experiment of Sridhar et al. (2009) shows that with the addition of prosodic features to an observed linguistic content, the accuracy of discourse disambiguation improves by only a small margin (2.8%).

These observations, taken together, suggest that the prosody-meaning system is a complex organization that includes some salient and robust one-to-one links associating a single prosodic cue with a single pragmatic meaning, but also multiple many-to-many relationships in which different

prosodic cues might encode a single meaning, and a single prosodic cue might signal several meanings. As Cole (2015, p. 25) says, "All of this variability means that prosody is not a fully reliable cue to discourse meaning." Relating this current state of our understanding to the nature of cue phrases, it seems that in order to disambiguate cue phrases and to distinguish between various meanings one discourse marker may convey, the prosodic aspects of communication alone might not be sufficient. In addition to prosody, the context in which a cue word is used is a natural candidate for providing disambiguating cues. Hence, in this study we wish to compare the role of prosody with the role of context in cuing the pragmatic meanings of Slovak 'no'.

2.3.2 The role of context
Despite many issues in formulating a unified definition of context, it seems that there is one basic feature of context that is agreed upon among different studies: context might be conceptualized as any factor (linguistic, social, epistemic, and so on) which **influences the interpretation of particular content.** Linguistically, despite many differences, a common opinion is that context is the part of a discourse that surrounds a language unit and **helps to determine its interpretation.** Hirst (1997) proposes that "context is a source of information that can be used (is used, should be used, may be used, must be used) by a language processor to reduce (or completely eliminate) ambiguity, vagueness, or underspecification in its interpretations of the utterances that it processes. That's one of the effects of context. It constrains interpretation" (Hirst, 1997, p. 5). For example, context thus might disambiguate various structural ambiguities (e.g. "Visiting relatives might be nuisance", "I met Mary and Elena's mother", "I like Bill more than Mary"). In such examples, relevant context, even without access to prosody, can indicate what the utterance exactly means.

Based on this we assume that one of the functions of context at the linguistic level is participation in the process of disambiguation of the potentially ambiguous parts of the utterance. We hypothesize that context might be more tangible and accessible to interlocutors whereas prosody might be less directly accessible. Several studies showed that context and lexico-semantic information might provide better cues for various pragmatic or turn-management functions than prosody (e.g. Beňuš et al., 2007; de Ruiter et al., 2006). Moreover, prosody has many subsystems, is variable and depends on various factors (nationality, age, sex, current physical and psychological status). Context seems to be more independent of these factors and might thus be more reliable in the process of disambiguation of an utterance.

Based on these considerations, we formulated the following general hypothesis which is explicitly related to our objectives:

> H1: We expect that our subjects will be less successful in the identification of functions of *'no'* when having only prosodic cues available than they will be with lexical context available.

2.4 Biological sex and age
According to Allgood and Heaton (2015) and also Briton and Hall (1995) there is evidence that females are more sensitive to the cues that influence the disambiguation of utterances than males. These studies have shown that females and males differ in the way they express, experience and decode or encode emotions through prosody, with females outperforming males when completing tasks designed to identify prosodic cues.

Concerning age, Paulmann et al. (2008) showed that younger participants of oral communication are more sensitive to the prosodic cues that influence the disambiguation of utterances than older participants. Also, they found out that older subjects had as well greater delays in responding to their stimuli.

Although our experiment is different in exploring pragmatic meanings and not emotions, the results of the above studies warrant testing the effect of age and biological sex on the relationship of context and prosody. These results also serve as a motivation for us to test the influence of biological sex on the pragmatic perception of 'no'.

In addition to the general hypothesis in the previous section, we are thus also interested in the following additional questions:

> Q1: Do females and males differ in their sensitivity to prosodic cues in identifying the functions of *'no'*?

Q2: Does the age of subjects affect their sensitivity to prosodic cues in identifying the functions of *'no'*?

3. Methodology

The object of our research was the Slovak particle *'no'* and its pragmatic functions and meanings in relation to prosody. We wanted to test how people perceive the function of *'no'* when knowing the lexical context, in comparison to how the same people perceive this function when having just prosodic features available.

3.1 Stimuli preparation

Data comes from a corpus of task-oriented dyadic collaborative conversations (Beňuš, 2012). Interlocutors were recorded while playing an interactive game through a computer. Participants were seated in a quiet room in a way that they could not see each other but could hear each other. One of the participants played the role of the describer, the other one the role of the placer. During the spoken interaction they repeatedly changed these roles. Both players saw on their respective screen a collection of objects. On the placer's screen, however, there was one object missing. In each session one of the speakers (the describer) was asked to orally navigate his/her partner (the placer) to move the desired object from the list of unused objects to the same position as it was on the describer's screen. In each game there were fourteen tasks of object-placing with equally divided roles - both participants played the role of the describer and the role of the placer seven times.

Within the almost four hours of speech that this corpus contains, the Slovak particle *'no'* was used 1067 times. Four annotators labelled all *no*-tokens. Their task was to identify the function of *'no'* from fifteen options. A synopsis of these options is given in *Table 1*:

Table 1. The synopsis of the options that describe various functions of the Slovak particle 'no' (Beňuš, 2012)

Label	Short description	Meaning
R	Acknowledgment	I acknowledge that I understand, I got it.
RP	Backchannel	I acknowledge that I understand, and please continue.
RZ	I got it, but…	I acknowledge that I understand, but I want to add something or express mild disagreement.
RN	Pivot from one topic to another	I acknowledge that I understand, and I want to start a new topic or a new discourse segment.
N	New segment	I want to start a new topic or a new discourse segment.
S	I agree, yes	I agree, also as an answer to question, usually meaning *yes*.
Z	But	I want to express an idea opposite to implied before, usually meaning *but* or *well*.
L	Primary (literary) meaning	Primary (literary) meaning, e.g. *'no'* as a conjunction.
H	Hesitation	Hesitation, I am stalling for time.
E	Repair	I want to repair/redo something I have just said or did.
PH	Assessment	Express assessment of something that just happened, usually on receiving a score.
J	Hedge	Softening of what is to follow, a hedge.
K	The end	Signal the end of the current topic or discourse segment.
D	Do something	Encourage some action, go on, do something.
?	I do not know	None of the labels correspond to the meaning I perceive.

These annotations were then analysed and the tokens that were labelled unanimously - with the same label from all four annotators - were identified. The decision to use this strategy was determined by the desire to maximize the chance that the function of the particle *'no'* in that particular communicational

situation is reliably perceived. The number of such tokens is 395. In these 395 tokens annotators used eleven different labels (R, RP, RZ, RN, N, S, H, E, PH, K, D).

3.2 Experiment

We decided to pick the three labels that were used most often: RP - backchannel (229 tokens), S - agreement (37 tokens), and N - cue beginning (33 tokens). After that, we semi-randomly[1] picked seven audio files of every label (21 audio files overall). We extracted the turn including *'no'* and as many additional turns (turns that precede or follow the *'no'* turn) as needed. For tokens labelled with S it was always necessary to extract a turn that preceded the *'no'* turn. For RP it was necessary to extract a turn that preceded the *'no'* turn (so it would be clear that the second speaker reacts to the previous turn) and as well the turn that followed or at least its part (so it would be clear that the second speaker continues in his/her speech after *'no'*). In some of the N tokens it was also needed to extract the *'no'* turn and the turn that preceded *'no'* (so it would be clear that the speaker does not react to the previous turn). In some of them (where one speaker talks for a longer time about one topic and in the same turn starts a new topic) it was enough to use just the *'no'* turn.

To study whether the perception of *'no'* is dependent on contextual information or not we decided to use these 21 audio files in two forms:

- The lexical context is masked (masked de-lexicalized audio).
- The lexical context is available in the textual form (transcribed).

For the first type of data it was necessary to edit the recordings so that the lexical context was not identifiable. The word *'no'* was not masked by the filter. Editing was done in the program Audacity. We used the "Low Pass Filter" option to mask the lexical context. The upper frequency boundary of the low pass filter was chosen subjectively and varied according to F0 from 300Hz to 400Hz. The higher the F0, the higher the upper boundary.

As this modification lowers the loudness of the whole sound, we used the option "Amplify" to increase the loudness. We increased the audibility to the same level (13 dB) for all 21 samples, but there were some samples in which we needed to increase the loudness to a higher level (about 15-16 dB - this was needed mainly among female interlocutors).

Next, it was necessary to ensure that the lexical context was unintelligible. Therefore, we asked three people to listen to these sound sequences and mark any words they recognized. There were no words recognized.

For the second form of stimuli (where the lexical context would be available in textual form) we prepared a standard written transcription of the previously chosen audio units.

3.3 Subjects and procedure

We used the internet and electronic communication with the subjects. We created an online questionnaire that consisted of two parts: a part with 21 masked de-lexicalized audio clips and a part with 21 textual transcriptions. For each unit there were three options that subjects were supposed to consider and to choose one from. These three options corresponded to the three labels (S, RP and N) and included a short description used in Table 1. Our goal was to collect responses from 36 subjects - 18 men and 18 women - of three age categories: 1. 18-30 years-old; 2. 31-50 years-old; 3. 51 years-old and older. We divided the subjects into two groups depending on the order of the questionnaire parts: 1. Subjects responding first to the masked de-lexicalized audio; 2. Subjects responding first to the textual transcription. All subjects were native speakers of Slovak.

---

[1] It was originally planned to pick audio examples as randomly as possible. However, some of the randomly picked audio files were not suitable for various reasons (e.g. the *'no'* token was not audible enough). Additionally, we wanted to have seven audio files from different interlocutor pairs. Therefore, it was necessary to loosen the strict random requirement and some of the less ideal random tokens were replaced by better suited tokens next in the random list.

4. Results

Out of 36 people who were asked to participate, only 30 actually completed the questionnaire. Four subjects reported that after opening the questionnaire and becoming familiar with the instructions, the questionnaire seemed to be too difficult. The remaining two people did not respond after the initial agreement. All these six people belonged to the age category 51 or older; four of them were men and two were women. Even though this outcome was not desirable for our research, it also suggests that the perception of our investigated data seemed to be more difficult for people belonging to the higher age category than it did for younger subjects.

Of the remaining 30 subjects, 14 men and 16 women, 15 listened to masked de-lexicalized audio data first (further "aud-first") and 15 started with the textual transcription of this data (further "txt-first"). Each subject answered all 42 questions (21 in the audio and 21 in the text part of the questionnaire) - making a total of 1260 responses.

4.1 Descriptive observations

The overall accuracy of subjects was 57%; i.e. 718 of 1260 answers agreed with the unanimous agreement among the four labellers. Considering that the subjects had three options to choose from, we can say that this percentage is considerably higher than the baseline of just guessing 33%. Hence, prosody and context provide useful cues for disambiguation of the discourse functions of *'no'*.

The overall accuracy of subjects in the de-lexicalized type of data (49.1%) was lower than their accuracy in the transcribed type of this data (64.9%). We can see that our subjects were able to identify the functions of *'no'* more accurately having lexical context available. However, the accuracy of subjects in the identification of *'no'* having only the prosodic cues was still considerably over the baseline of 33%. Hence, we can see that prosody and its aspects are deeply ingrained in our cognitive system, meaning that people are able to automatically and subconsciously identify the meanings of spoken communication even without any lexical context. Therefore both - context and prosody - influence the identification of various functions of *'no'*.

The greater observed accuracy in the text part of the questionnaire than in the masked audio might suggest that context provides more, or better, information than prosody alone. This would align with the results of the study of the English cue phrase 'okay' (Gravano et al., 2007) which showed that contextual influence is also prevalent in the identification of 'okay'. We might thus say that the Slovak discourse marker *'no'* behaves similarly to English discourse markers in terms of the relationship between context and prosody in cuing their functions. However, an alternative explanation, or additional consideration, includes the degree of naturalness of the two types of stimuli. As also pointed out by a reviewer, listening to the cue phrase embedded in an incomprehensible speech-like signal may have imposed a processing burden on the listeners that was not present in the text-only condition. This extra burden in terms of degraded and impoverished audio signal might have impeded accuracy in the task of selecting the cue-phrase meaning.

4.2 The role of experimental factors in the perception of 'no' meanings

In an effort to construct a model including all subjects and as many factors as possible and avoid testing the effects of individual factors separately we employed a generalized linear mixed effects model in R for the binary dependent variable accuracy (correct vs. incorrect), and tested the effects of MODALITY (masked audio vs. text), BIOLOGICAL SEX (males vs. females), ORDER (aud-first vs. txt-first), and their interactions, while including the random factors of SUBJECT, QUESTION and AGE. We included age as a random factor in this main model since six of the missing subjects all came from the oldest category, thus producing an imbalance in already sparsely populated data cells. Only random intercepts were used since the model did not converge with random slopes.

In this model, the strongest effect is due to MODALITY, as already observed in the previous section. In the text part of the questionnaire subjects were significantly more accurate in their identification of the *'no'* meaning than in the masked audio part ($\beta = 1.37$, SE = 0.24, z = 5.77, p < 0.001). This main effect can be observed in Figures 1 and 2 by comparing the left and right panels.
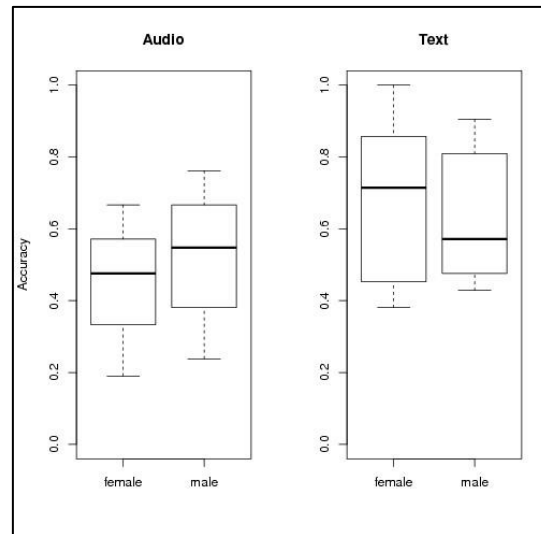
Figure 1. Accuracy of females and males split for the masked de-lexicalized and textual data

BIOLOGICAL SEX and ORDER did not show the main effect but they both interacted weakly, but significantly, with MODALITY. As shown in Figure 1, the higher accuracy of the text modality is primarily due to females who scored better than males in this condition while the performance in the masked audio condition was comparable between the sexes (β = -0.99, SE = 0.36, z = -2.72, p = 0.006). Hence, the difference between prosody and context in the quality of information they provide is greater for females than for males. While males performed comparably in the two modalities, females showed a significant difference among them in favour of the text modality.

Similarly, as seen in Figure 2, subjects were slightly more accurate in the text part of the questionnaire if it was the second part following masked audio while their performance in the first textual part was comparable to the masked audio condition (β = -0.73, SE = 0.35, z = -2.1, p = 0.037). Although we might have expected better results in the second part of the questionnaire due to the familiarity with the task; this main effect was negligible and non-significant overall. But subjects in the second text-based questionnaire benefited from the previous audio part more than the subjects in the second audio part benefited from the first textual part. The remaining two interactions in this main test were not significant.
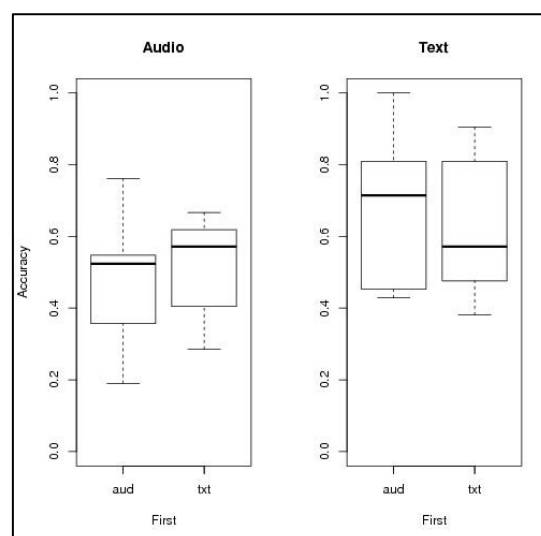

Figure 2. Accuracy of subjects depending on the order of stimuli presentation (aud-first vs. txt-first) females and males split for the masked de-lexicalized and textual data

Despite removing age from the tested factors above due to the imbalance caused by dropping many of the oldest subjects, we were interested if there was a difference between the young and the middle-aged categories in how they perceived the functions of *'no'*. Figure 3 again shows a possible interaction between MODALITY and AGE. We thus removed the remaining 6 older subjects and constructed a similar generalized linear mixed effects model just including MODALITY, AGE, including only young and middle-aged subjects, and their interaction in the subset of 24 subjects. The test showed no significant main effect of AGE, but, indeed, a significant interaction: the young subjects were significantly more accurate in the text condition than in the masked audio while the middle-aged subjects showed similar performance in the two ($\beta = 0.99$, SE = 0.42, z = 2.4, p = 0.018).
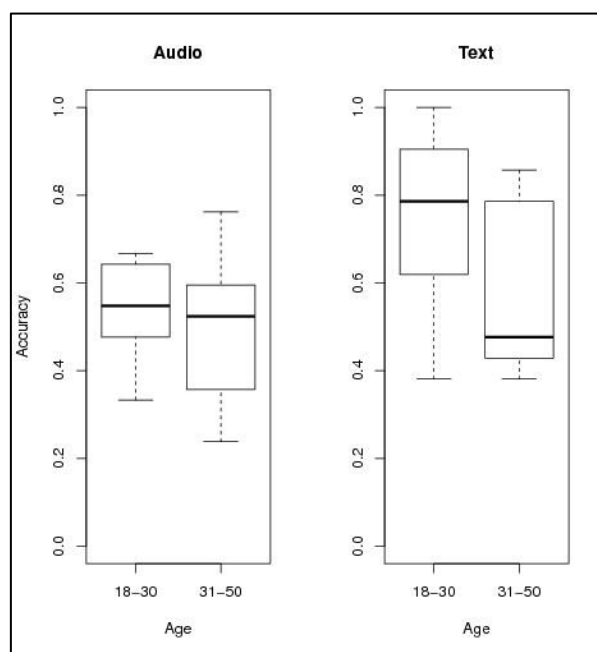

Figure 3. Accuracy of subjects depending on the age (young vs. middle) split for the masked de-lexicalized and textual data

The three figures and the associated statistical modelling shows that all three investigated factors had some effect on the role of context and prosody in perceiving the pragmatic functions of Slovak *'no'*. While the generally higher accuracy in text than in the masked audio stimuli is a strong and robust effect, its interactions with biological sex, the order of stimuli presentation and age showed that females, younger adults, and subjects filling the text questionnaire after the modal audio all did better in the text than in the audio. By the same token, this sensitivity to modality (text vs. audio) was not observed for males and middle-aged subjects.

5. Discussion
We set out to test the role of context and prosody on the interpretation of discourse pragmatic meanings of Slovak *'no'* in general and presented two additional hypotheses regarding the biological sex and age of the participants affecting this perception.

We start with a reminder that the task of identifying these pragmatic meanings is a genuinely difficult one. Recall that from the original corpus of 1067 tokens only 395 (37%) were labelled unanimously for one of the pragmatic functions when the annotators had all the information available. Over 60% of the tokens that were not labelled unanimously suggests a subjective nature of the prosody-meaning relationship. Of course, context and prosody are relevant and important, but they are not the only cues. We have to keep in mind that when it comes to the interpretation of various meanings of speech, there is no such thing as a "correct" answer. We are all influenced by different circumstances, life experiences, social backgrounds, and we perceive speech with all its aspects in various ways. That is the reason why

we selected our stimuli only from those unanimously labelled tokens in which the pragmatic meaning seems to be the most transparent.

This complex task of identifying the pragmatic meanings of *'no'* tokens was partly simplified for our subjects by limiting the number of functions to three and selecting only the most transparent tokens, but also further complicated by either obscuring prosody completely or masking the lexical context. The main result suggests that even in this complex task, subjects use both prosody and context in identifying the pragmatic meaning of *'no'* tokens. We found out that lexical context provides better cues in identifying the function of the Slovak particle *'no'* than de-lexicalized prosody included in low-pass filtered speech. The accuracy of subjects in the identification of the functions of *'no'* with only prosodic cues available was significantly lower than with just lexical context available. However, subjects' accuracy in the identification of the functions of *'no'* according to prosodic cues was still significantly over the baseline. Both - context and prosody - to a great extent influence the disambiguation of *'no'*, but context provides more information about the function of Slovak *'no'* than prosody does.

The first of the additional questions explored the accuracy of females when compared to males in the perception of prosody. The expectation that females would be more sensitive to prosody was not supported in our data. First, there was no main effect of biological sex on accuracy. Second, the interaction between biological sex and modality (text vs. masked audio) showed increased accuracy of females in the text condition while males performed comparably in the two modalities. The second additional question concerned the potential effect of age on the sensitivity of subjects to prosodic cues in disambiguating pragmatic functions. Although we investigated this question in a limited fashion due to exclusion of the oldest-category subjects, the results are similar to those of biological sex. Young adults (18-30) were not better in the audio in general than middle-aged subjects (31-50), and the interaction with modality showed increased accuracy of young adults in the text-only condition rather than the audio condition.

We speculate that these two observations can be conceptualized along similar lines. In both females and young adults, we expected greater sensitivity to prosody, and in both we saw text providing better cues than low-pass filtered audio. We see two potential explanations. First, the degradation of the audio stimuli in low-pass filtering acted against assumed female and younger-adult sensitivity to the prosodic cues. It might be that the information females and young adults rely on was filtered, or the degradation lowered the sensitivity of these two groups in general. The second, and we think more plausible, explanation is that it might be that increased accuracy of females and young adults in the text condition over males and middle-aged subjects stems from the ability of the former to reconstruct implicit prosody from text better than the latter.

6. Study limitations and future research

This is a small-scale pilot exploratory study and as such includes multiple limitations. The first one is the nature of lexical masking using low-pass filtering. This is not an ideal method since some prosodic features are masked as well. Hence, such audio definitely does not represent the complete prosodic information available in an unfiltered sound file. Additionally, as already mentioned in section 4.1, this way of eliminating lexical context likely imposed a processing burden on the listeners, impeding their accuracy in the task of function identification when compared to text-based identification. Moreover, this burden might have been greater for older subjects and thus influence their results more. We partially alleviated this last point by excluding older subjects from the statistical testing of the age factor on the perception of *'no'* meanings. However, an approach addressing some of the points above, and also suggested by reviewers, would be to use also full audio stimuli without masking. It would be interesting to see the upper boundary of the listeners' abilities in this task, which could provide a better basis of comparison. For example, it would allow for testing the added effect of prosody by comparing text-only with full audio, and also for assessing the processing burden imposed by low-pass filtering. We intend to implement this condition in data collection in our future research.

The second set of limitations includes the analysis presented. In addition to the general accuracy-based findings, we plan to look at two additional areas. First, the length of discourse context in individual stimuli varied across items. Therefore, we plan to treat it as a predictor when we analyse the results based on stimuli items. Second, we did not analyse error patterns both in terms of individual stimuli tokens and the factors of interest. In future, we would like to explore question such as whether females

or younger adults were particularly accurate (or inaccurate) on certain stimuli items and how the nature of the stimuli might explain such patterns.

7. Conclusion

The goal of this work was to analyse the relationship between context and prosody and the influence of this relationship on the perception of the Slovak particle *'no'*. We were interested in whether there is a significant difference in the perception of *'no'* when having the lexical context available and in the perception of *'no'* when having only prosodic cues available. Furthermore, we were interested in whether there are significant differences between men and women in identifying the functions of this particle. We studied whether there was a difference in perception of *'no'* between different age categories. We also analysed if the order of stimuli played a significant role in the identification of the functions of the particle *'no'*.

The core finding is that subjects were generally more accurate in identifying pragmatic functions in text than in the masked audio stimuli and that this robust effect interacted with biological sex, the order of stimuli presentation and age, revealing that females, younger subjects and those assessing functions based on transcripts after assessment based on masked audio all did better in the text condition than in the audio condition.

These results presented in Section 4 and discussed in Section 5 bring valuable knowledge and provide a basis for better understanding of natural human communication. We hope to provide useful data that adds to our understanding of how communication is ingrained in the human brain and how prosodic and contextual cues are interpreted within different age and biological sex categories.

Acknowledgment

References

Allgood, R. and Heaton, P., 2015. Developmental change and cross-domain links in vocal and musical emotion recognition performance in childhood. *onlinelibrary.wiley.com* [Accessed 18 February 2019]. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/bjdp.12097

Andersen, G., 2001. Pragmatic Markers and Sociolinguistic Variation: A Relevance-theoretic Approach to the Language of Adolescents. Philadelphia: John Benjamins Publishing.

Austin, L. J., 1975. How to Do Thing with Words. New York: Clarendon Press.

Beňuš, Š., 2012. Prosodic forms and pragmatic meanings: the case of the discourse marker 'no' in Slovak. Paper presented at *3rd IEEE International Conference on Cognitive Infocommunications, CogInfoCom.* Košice, Slovakia, 2-5 December.

Beňuš, Š., Gravano, A. and Hirschberg, J., 2007. Prosody, emotions, and… whatever. Paper presented at *8th Annual Conference of the International Speech Communication Association.* Antwerp, Belgium, 27-31 August.

Blake, L.M., 2017. The Right Hemisphere and Disorders of Cognition and Communication: Theory and Clinical Practice. San Diego, CA: Plural Publishing.

Brinton, J. L., 1996. Pragmatic Markers in English: Grammaticalization and Discourse Functions. New York: Mouton de Gruyter.

Briton, N. J. and Hall, J. A., 1995. Beliefs about female and male nonverbal communication. *link.springer.com* [Accessed 18 February 2019]. Available at: https://link.springer.com/article/10.1007/BF01544758

Cole, J., 2015. Prosody in Context: A review. *tandfonline.com* [Accessed 17 February 2019]. Available at: https://www.tandfonline.com/doi/full/10.1080/23273798.2014.963130?scroll=top&needAccess=true

De Ruiter, J. P., Mutterer, H. and Enfield, N. J., 2006. Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, vol. 82, no. 3, pp. 515-535.

Dupuis, K. and Pichora-Fuller, M. K., 2010. Use of affective prosody by young and older adults. *psycnet.apa.org* [Accessed 03 March 2019]. Available at: https://psycnet.apa.org/record/2010-04859-002

Duranti, A., 1992. Rethinking Context: Language as an Interactive Phenomenon. Cambridge: Cambridge University Press.

Findra, J., 2004. Štylistika slovenčiny. Martin: Osveta.

Gravano, A., Beňuš, Š., Chávez, H., Hirschberg, J. and Wilcox, L., 2007. On the role of context and prosody in the interpretation of ´okay´. Paper presented at *45th Annual Meeting of the Association of Computational Linguistics.* Prague, Czech Republic, 23-30 June.

Golato, A. and Fagyal, Z., 2008. Comparing Single and Double Sayings of the German Response Token ja and the Role of Prosody: A Conversation Analytic Perspective. *Research on Language and Social Interaction,* vol. 41, no. 3, pp. 241-270.

Hirschberg, J. and Litman, D., 1993. Empirical Studies on the Disambiguation of Cue Phrases. *dl.acm.org* [Accessed 30 January 2019]. Available at: https://dl.acm.org/citation.cfm?id=972490

Hirst, G., 1997. Context as a Spurious Concept. *ftp.cs.toronto.edu.* [Accessed 12 February 2019]. Available at: http://ftp.cs.toronto.edu/pub/gh/Hirst-Context-97.pdf

Hockey, A. B., 1993. Prosody and the role of the okay and uh-huh in discourse. *citeseerx.ist.psu.edu* [Accessed 12 February 2019]. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.31.2337&rep=rep1&type=pdf

Hölker, K., 1988. Zur Analyse Von Markern: Korrektur- und Schlussmarker des Französischen. Stuttgart *:* F. Steiner Verlag Wiesbaden.

Jucker, A. H. and Ziv, Y., 1998. Discourse Markers: Descriptions and Theory. Philadelphia: John Benjamins Publishing.

Laussen, A. and Schacht, A., 2018. Biological sex Differences in the Recognition of Vocal Emotions. *ncbi.nlm.nih.gov* [Accessed 03 March 2019]. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5996252/

Lee, L., Bartkova, K., Jouvet, D., Dargnat, M. and Keromnes, Y., 2019. Can prosody meet pragmatics? Case of discourse particles in French. Paper presented at *International Congress of Phonetic Sciences*. Melbourne, Australia, 5-9 August.

Mistrík, J., 1997. Štylistika. Bratislava: SPN.

Paulmann, S., Pell, M. D. and Kotz, S.A., 2008. How aging affects the recognition of emotional speech. *mcgill.ca* [Accessed 18 February 2019]. Available at: https://www.mcgill.ca/pell_lab/files/pell_lab/paulmann_pell__kotz_2008b.pdf

Récanati, F., 1987. Meaning and Force: The Pragmatics of Performative Utterances. Cambridge: Cambridge University Press Archive.

Romero, P., 2010. Phantom Stress: Brain Training to Master Relationship Stress. Bloomington, IN, USA: Xlibris Corporation.

Sridhar, V. K. R., Bangalore, S. and Narayanan, S., 2009. Combining lexical, syntactic and prosodic cues for improved online dialog act tagging. *ict.usc.edu* [Accessed 02 March 2019]. Available at: http://ict.usc.edu/pubs/Combining%20lexical,%20syntactic%20and%20prosodic%20cues%20for%20improved%20online%20dialog%20act%20tagging.pdf

Svartvik, J. and Stenström A. B., 1985. Words, words, words: The rest is silence? In: S. Bäckman and G. Kjellmer, eds. *Papers on language and literature. Presented to Alvar Ellegärd and Erik Frykman.* Göteborg : Humanities Pr, pp. 342-353.

Volín, J., Weingartová, L. and Niebuhr, O., 2016. The Prosody of the Czech Discourse Marker 'Jasně': An Analysis of Forms and Functions. *Phonetica* vol. 73, no. 3-4, pp. 314-337.