PROSODIC PHRASE IN SPOKEN CZECH

JAN **VOLÍN**
PAVEL **ŠTURM**
RADEK **SKARNITZL**
TOMÁŠ **BOŘIL**

KAROLINUM

Prosodic Phrase in Spoken Czech

Jan Volín
Pavel Šturm
Radek Skarnitzl
Tomáš Bořil

———

KAROLINUM PRESS
Karolinum Press is a publishing department of Charles University
Ovocný trh 560/5, 116 36 Prague 1, Czech Republic
www.karolinum.cz

Set in the Czech Republic by Karolinum Press
Layout by Jan Šerých
First edition

A catalogue record for this book is available from the National Library
of the Czech Republic.

The research presented in this book was carried out with the support of GAČR
(Czech Science Foundation), Project 21-14758S.

The original manuscript was reviewed by Alice J. Henderson
(Université Grenoble Alpes, France) and Tomáš Hoskovec
(University of South Bohemia in České Budějovice, Czech Republic).

ISBN 978-80-246-5798-1
ISBN 978-80-246-5827-8 (pdf)
ISBN 978-80-246-5828-5 (epub)

Charles University
Karolinum Press

www.karolinum.cz
ebooks@karolinum.cz

CONTENTS

FOREWORD

The monograph *Prosodic Phrase in Spoken Czech*, which you are beginning to read, has been dreamt of for decades. Thirty years ago, the then Director of the Institute of Phonetics in Prague advertised her personal plans for such a book. She claimed that the topic of prosodic structure was undeniably interesting, attractive even outside phonetics, and, quite importantly, there was absence of anything in this vein for the Czech language. Informal discussions at various academic gatherings have confirmed that a treatise of this sort was needed not only by phoneticians, but also by many other people who were interested in the sound patterns of Czech, whether for technological use in speech synthesis and automatic speech recognition or for didactic, forensic, therapeutic and other purposes. Overwhelmed with other duties, she never wrote the book.

The current authors joined their capacities to materialize the plans, at least partially. Although they definitely aspired to produce a useful treatise, they were not motivated by any specific technological applications. Those have grown dominant across scientific activities far and wide: commercialized societies have become reluctant to see the sheer joy of discovery as a true value. Yet, the desire to discover and the ability to revisit existing knowledge reach above and beyond consumerism. How magnificent it would be to replace the current plundering of the Earth's wealth with admiration for it! The authors of the present book believe that the complexity of every aspect of our lives deserves thorough study, and that the resulting knowledge should be shared.

Naturally, there needs to be demand for knowledge. The trouble is that when it comes to language, everybody 'has an opinion'. All humans typically use language to communicate, and many believe, by extension, that they are experts on language even if they never trained anything but spelling rules. One has to wonder if there is any scientific field more plagued with dilettantism and ignorance than linguistics.

These words sound perhaps a bit too harsh. Not everyone chooses ignorance willingly and deliberately. And it is only very recently that people have started accepting the fact that there is a massive gap between what we are able to capture by our limited conscious circuitry and what our whole brains really do. The way we are currently able to describe language analytically is not the way language really works. Recent research in adaptive, implicit thinking has brought evidence that we do not realize how

we function: how we process what we sense and how we plan our actions. Our consciousness cannot capture what we really feel before it is censored by various mental mechanisms. Yet, all this is extremely relevant to linguistics, and this is also why we currently witness a powerful shift from speculative to empirical methods.

Less self-satisfaction and more patient work are becoming typical for linguists. For instance, old introductory manuals used to amaze students with the claim that a finite number of small elements can be used to create an infinite number of utterances. We can leave the practical impact of this dazzling claim aside, but one thing is clear. It obscures the fact that people do not use infinite numbers of utterances in their everyday lives. This is because what we communicate is strictly grounded in what we live. Huge numbers are not infinity. The infinitely rich inner world of an immortal being is one of the pompous anthropocentric myths.

Therefore, let us focus with modesty on limited samples of true attempts to communicate. The seemingly pleonastic word 'spoken' in the title of this book was used purposefully: we wanted to emphasize that we do not estimate prosodic phrasing from written text as some scholars did in the past. Neither do we use invented contextless sentences where the communicative intent of speakers may be dubious or completely absent. On the other hand, we do not dare to use recordings of spontaneous conversations as yet: those are difficult to obtain legally, and they are difficult to analyze in our current linguistic framework. We opted for spoken texts that are produced with legitimate communicative intents and, yet, provide manageable and dependable data. We hope this approach may help us as authors and you as readers to find a way from biased or circular discussions towards the joy of discovery adequate to the times we live in.

Prague, August 2023
The Authors

1.
SPEECH PROSODY

1.1 RATIONALE

The value of language communication has been most probably appreciated already in pre-history. Even before people started recording their thoughts, important contracts, laws or observations in writing, they must have realized how important their speech behaviour was. From sheer admiration for the power of speech (often explicated as a magnificent gift from gods) people gradually moved to systematic observations, disciplined descriptions, and rigorous experimenting, i.e., to scientific treatment of speech. This approach embraced writing as well and, inevitably, the system that governs both of these communication means – language.

There is currently little opposition to the claim that language communication should be studied scientifically. Even though language use is omnipresent and quite pervasive in our everyday lives, when considered thoroughly it can hardly be seen as trivial. Unfortunately, there is still the awkward legacy of the past, by which linguistics is sorted into 'less precise' or 'less scientific' disciplines. However, this primordial figment is nowadays either just a superficial feeling of someone who has no capacity for deeper critical thinking or an unfair excuse in the relentless fight for resources in the current system of science financing. Be that as it may, modern linguistics can offer very interesting insights into the functioning of language. Phonetics, which is ultimately by its goals a linguistic discipline, has been doing so beyond any doubt.

One of the serious problems with the credibility of the past linguistics was the claim that human language is 'utterly unique' and, therefore, can be explained only from itself. This type of mysticism, within which scholars *a priori* decide not to see any connections with other communicative systems and not to seek any compatibility with other scientific disciplines, was self-destructive. Fortunately, it was prevalently abandoned before it made linguistics completely socially irrelevant. Most current linguists do not shy away from testing their hypotheses with methods that originated in psychology, neurophysiology, ethology, sociology, etc. In addition, a powerful boost to linguistic theories has been provided by the increased possibilities of cross-linguistic research.

A brief but patent example of the above-described developments can be found in the following quote: "... there are differences in delta responses across languages

due to the different usage of stress. In English and German, for example, phrasal responses are emphasized but in French, syllabicity remains dominant." (Ghitza, Giraud & Poeppel, 2013). The authors carry out cross-language comparisons and they use neurophysiological correlates of perceptual processes. The underlying message indicates that prosodic structure is of relevance. The phrasal responses mentioned in the quote are especially relevant to the chief topic of the present book. They point to a speech unit that, under various names and with various attributes across languages, seems to be always somehow present in the hierarchical structure: the PROSODIC PHRASE.

Already in 2001, Chafe observed that prosodic phrases in his corpus of conversational speech were "typically one to two seconds long", which to him was about the span of active consciousness. He also proposed semi-active consciousness for contextual information, but claimed that attention can only be devoted to smaller chunks of information (Chafe, 2001: 675). According to him, this 'information packaging' is essential for smooth speech processing by interacting individuals.

Two decades later, LaCroix and her colleagues listed a large number of studies that had resulted in the inference that "typical sentence prosody yields faster and more accurate sentence comprehension performance than atypical prosodic patterns" (LaCroix et al., 2020: 2). Their own experiment with brain-stroke patients and a group of typical language users added valid evidence to the claim.

Nowadays, it is widely accepted that division of larger stretches of speech into prosodic phrases is critical for effortless mental processing and, ultimately, even for correct recovery of intended meanings of spoken texts (be it the representational, conative, or affective components of the communicated messages). Individuals who intend to talk about objects and events of the surrounding world must always select only a limited choice of aspects to talk about at a time, while ignoring or backgrounding others. This is because taking all possible observables into account at once would be beyond human cognitive capacities and, also, beyond the potentials of natural language. Prosodic structure contributes to the process of focusing the addressee's attention on those aspects of reality that need to be communicated at the given moment, and provides a particular perspective on the state of affairs (in line with the personal intents of the speaker). Prosodic phrases – the central object of study in this monograph – belong among the devices that speakers use to guide listeners through spoken texts to desired conclusions.

From today's perspective, it would be quite naïve to imagine that listeners analyze linear flows of small units (phonemes, syllables) and glue them in their mind together to compose units that are meaningful. Phonemes and syllables need to be seen as mere features of meaningful items, and their qualities and positions only serve to recognize those meaningful objects as wholes. As long as these wholes fit into the given situation, they do not need to be analyzed into their parts. People do not have to focus on the exact qualities and the precise positions of 'features'; in fact, if they did, speech communication could not have developed into the fast process that it is. Recent development in this research area suggests quite convincingly that we even do not necessarily assemble our utterances from individual words. There is evidence from various sources

(e.g., young infants, aphasic patients, motoric patterns in typing, eye movement in reading, etc.) pointing to the use of MULTI-WORD UNITS (MWU) and the holistic storage of those in our brains (Lin, 2018: 48 and 53; Ellis, 2003: 75). It is hypothesized that speakers provide various cues about the boundaries of such meaningful units as they plan their utterances in them.

Although the above are various general conclusions based on empirical findings, Kohler emphasizes that there are profound differences in speakers' proficiency concerning phrasing (Kohler, 2018: 147). In other words, speakers are more or less adept at grouping words together in a way that is transparent and easy to follow for listeners. Kohler links the dexterity in prosodic phrasing to the more general cognitive skill of argumentation and perhaps even logical thinking. He demonstrates how, first, the placement of a boundary and, second, the relative strength of a boundary (expressed by means of prosodic markers) may influence the perceived information structure in the spoken text. The past decades produced quite a large number of studies that experiment with such semantic ambiguities. For instance, the following string of words forms an ordinary Czech sentence:

Czech:	*Dá se*	*tam*	*volat*	*v sobotu*	*a*	*v neděli*	*večer*
English:	*It's possible*	*there*	*to telephone*	*on Saturday*	*and*	*on Sunday*	*evening*

We demonstrate two different prosodic boundary placements that result in two different interpretations of the proper time for making the telephone call. If the phrasal boundary is placed as follows:

Dá se tam volat ‖ v sobotu a v neděli večer,

the intended meaning is that the phone call is possible on both days always in the evening.

On the other hand, if the phrasal break is in the following position:

Dá se tam volat v sobotu ‖ a v neděli večer,

the phone call is possible the whole day on Saturday, but on Sunday only in the evening.

Of course, it would be unwise to restrict linguistics only to such ambiguities. Even if the meaning is clear, but difficult to understand, language communication suffers. Excessive effort on the part of the recipient projects far beyond just a single utterance meaning. Struggling recipients are less likely to cooperate with the speaker, less likely to even trust the speaker, and unlikely to proliferate any positive feelings both within and outside the conversation.

Whether we focus on completely misleading outcomes or just cumbersome speech processing, it is unquestionable that the prosodic phrase is an important element of prosodic structure and of communication in general. As such, it deserves attention of anyone who is interested in human language. This book is our response to the demand.

1.2 DEFINING PROSODY BY FORM AND FUNCTION

Define your terms if you wish to converse with me, an enlightened thinker once allegedly said. Indeed, it is impossible to discuss problems with someone who ascribes unknown meanings to his or her words. Ideas can only be effectively examined if the discussants follow a common path, if one person's reaction to the other person's thought matches as much as possible the intended meaning. In other words, ideas and reactions to them must 'meet in a shared space'. There is a lot to discuss in the field of prosody research. What is prosody of speech, then?

This is a question that researchers in prosody do not particularly like. Not that it is uninteresting. On the contrary, if discussed by experts it often leads to exciting interactions. However, people who ask the question in this simple form usually expect a simple answer. Such an answer, regrettably, does not exist. If we try to provide an uncomplicated answer, substantial parts of the truth may remain unspoken.

Three types of definitions are heard most often: definition by negation, definition by parallel, and definition by listing.

The definition by negation is very common in communities that use alphabetic orthographies. In such communities, young children discover very early in their lives that individual letters of their alphabet refer to specific little sounds. Enormous care devoted to these little sounds at schools (even at pre-school establishments or within families) results in the idea that language consists only of those. They are highly activated in people's minds, while the larger sound phenomena remain produced or perceived with little explicit awareness. Relying on the common knowledge of little sounds represented by letters of the alphabet, speech prosody is then explained as 'everything in speech that is not represented by the letters of the alphabet, e.g., by phones or segmental phonemes'. (It should be added that this 'everything' is meant as communicatively functional sound phenomena.)

The definition by parallel exploits another known concept: that of music. Since music plays a very important role in human communities, it can be put forward as analogous with speech. Prosody is then explained as an aggregate of all the phenomena that speech shares with music: melody, rhythm, tempo, timbre and loudness.

Definition by listing skips the analogy with music and relies on the common knowledge of prosodic phenomena. The list might then be virtually the same as the above, i.e., melody, rhythm, tempo, voice quality patterns and distribution of loudness, or it might list some of the partial phenomena on their own. This is often the case of stress, which can be studied on its own but is an essential part of rhythm.

An interesting approach is taken by Nespor and Vogel (1983), who begin their consideration of what prosody is by listing various prosodic units. Their list comprises: "rhyme, syllable, foot, phonological word, phonological phrase, intonational phrase, and utterance" (Nespor & Vogel, 1983: 123). At this point, we can leave aside the question whether the syllabic rhyme (referred to as RIME here; see Section 2.5) is too artificial and whether prosody forms any structures above the utterance. What needs to be noticed is that by listing these units, Nespor and Vogel do not feel any further need

to define what prosody is. They say that whatever happens within these units and is explained with reference to these units is prosodic. That is a combination of listing and implicit negation, which leads to a conclusion that anything above the segment is prosodic.

There are scholars who may claim that certain prosodic phenomena are only supralexical. This claim may simplify certain definitions but ultimately leads to trouble. On the one hand, it can differentiate the prosodic from the suprasegmental, for whatever reason felt as practical. On the other hand, if lexical stress can be materialized through melody (which it obviously can) and melody is ousted into the supralexical domain, we are trapped. However, if we realize that the natural use of language is actually supralexical (and this cannot be contradicted by the existence of one-word utterances), then the whole argument becomes quite unattractive. Indeed, we typically speak and write in utterances (see Section 2.1).

At this point, we could perhaps mention a colourful suggestion by Nick Ellis: "If words are the atoms of language function, then construction grammar provides the molecular level of analysis" (Ellis, 2003: 65). To elaborate on this figure, we could ascribe the phonemic level to atoms (and distinctive features to elementary particles), morphemes would be molecules, words could be cells, and utterances complex organs. Texts would then constitute organisms. Whether this simile is helpful or, on the contrary, confusing, can be decided by each reader individually. In any case, the prosody of speech (which we automatically project into written texts when reading) helps us to make sense of the intricate structures.

The last sentence of the preceding paragraph already invokes functions of prosody. Although it is easier to explain what speech prosody is by its appearance (i.e., form), the definition by function should be attempted as well. In the end, it is the function of things that makes them valuable.

In short, prosody serves to communicate meanings. This statement, however, is too vague and applies to segmental phonemes as well. Yet, even within this short proposition, there is a point to emphasize: linguistic functions should always be related to communication. Thus, if an author claims that the function of prosody is to create prominences and breaks, it is only true in the very general sense assigned to the word function. In the context of linguistics, however, such a statement would be misleading. Prominences and breaks are PROSODIC EVENTS or MEANS to fulfil (communicative) functions.

There have been many disparate attempts at listing the functions of prosody. They always reflect the frameworks within which their authors work and the purpose for which they were made. Ours will be no different in this sense. One of the many sources of our inspiration goes back almost a century to Karl Bühler (1934), who studied speech and language from the psychological point of view. He suggested that although the REPRESENTATIONAL COMPONENT of the meaning in language signs is the most obvious one, and hence might be considered primary, the AFFECTIVE and APPLICATIONAL COMPONENTS are nevertheless always present in utterances and, therefore, must be taken into consideration in any attempt to discuss language in more than

a fragmentary manner. For the sake of presentational clarity, the general cognitive supra-function (to serve the comprehension of the intended meaning) can be organized into various categories. In the following paragraphs, we will present our understanding of the affective function, discourse function, grammatical function, sociophonetic function, aesthetic function and lexical function.

The phylogenetically oldest function is the AFFECTIVE one. Affective processes can be observed also in non-human animals. They are adaptive evaluations that according to their strength, timing, and other parameters can be classified into emotions, moods, interactive stances, attitudes, and affective personal characteristics (Scherer, 2003). The affective component of meaning is present in all utterances with greater or smaller explicitness. Humans always signal their evaluations of the topics they talk about, of the addressees, or of the situations they are in. Even the so-called neutral style (which is extremely rare in real life) displays an implicit affective component: "I don't want to share my evaluation with you". If used, it is usually to keep distance between the speaker and the recipient(s). It is sometimes claimed, for instance, that lectures are presented without affective components, but we argue that good lecturers always display their involvement in the presented issues and their feelings about various aspects of their talk. If a lecturer is disinterested or monotonous, he or she displays clear negative inner evaluations. Prosody plays a key role in signalling affective processes, hence the common phrase "It is not what you said, it is how you said it". The affective function is sometimes labelled as attitudinal and sometimes as paralinguistic. The latter term is quite unfortunate, since it cripples our understanding of what the basis of language fundamentally is. By suggesting that the "how you said it" should be relegated outside of linguistics, one would strip the language of its essential core and make linguistics dangerously detached from reality.

Bühler's 'Appell' (1934) evolved in other traditions into an illocutionary force (Austin, 1962) from which we derive the applicative or conative component of utterance meaning. It refers to what the speaker wants to achieve through speaking. The reason why certain things are uttered is not necessarily obvious, and even the speakers themselves may not know why they are using certain discourse devices. This is because they often plan their speech production implicitly and the true motivations for certain lexical, syntactic and prosodic choices may stay hidden to their conscious mind. However, it is generally presupposed that healthy individuals act purposefully, even if the purpose is not accessible to their conscious observation. What should be achieved by individual utterances is related to the DISCOURSE FUNCTION of prosody.

Since human intentions and motivations are immensely varied, the discourse function is difficult to circumscribe. In the narrow sense, it is defined as managing solely the spoken interaction. By certain prosodic forms, the speaker can express what he or she expects from the recipient. Prosody may signal a wish to receive more information, to elicit the recipient's opinion, to ask for approval, to give permission to talk, etc. The speaker does not have to use explicit words for that. Consider the following exchange.

A: You may need my ID card, though

B: OK (↗)