European University Studies

Slavonic Languages and Literatures

Aleš Bičan

# Phonotactics of Czech

PL ACADEMIC RESEARCH

# Chapter One: INTRODUCTION

## 1.1 Goal

This book presents a phonotactic analysis of Czech. Czech is a language, a West Slavic one. Languages, in contrast to other systems of means of communication such as traffic lights or gestures, combine and order (i.e. arrange in a certain order) their means. It makes them highly effective and economic. Humans wanting to express an innumerable number of thoughts must cope with the inevitable fact that their lives and brain capacities are limited. Rather than expressing each message with a new and distinct means, it is more advantageous to have a limited stock of basic means which could be infinitely combined to produce new means for expressing novel messages. Of course, the ability to combine is not specific to languages only (and perhaps not even to human communication) because many other communication systems such as the Morse code make use of it, too. Still, languages are specific in taking advantage of *dual ordered combinability*. Each communication system has means for expressing some message or for conveying some meaning, and these means have an outward form or shape. In comparison to the other communication systems, languages combine and order both the meaning-expressing means and the formal means.

If one wants to understand how Czech (or any other language) functions as communication system, it is necessary to explain how its meaning-expressing and formal means are combined and arranged in combinations. Our goal is account for the combinability and the arrangement of formal means used by speakers of Czech. This area of research traditionally falls under the rubric of phonotactics which is a branch of linguistics that uses certain models to explain and to account for the mentioned characteristic property of languages. These models and other explanatory notions are formulated within some theoretical framework, and the one used in this work is based on the approach worked up by Jan Mulder and Sándor Hervey and known as Axiomatic Functionalism. The choice of such a framework is partly a result of an opinion shared with Mulder and Hervey and many others that science, the science of language included, should be an enterprise that enriches our knowledge by means of testable explanations and predictions about certain things. The testability implies the absence of speculation, which unfortunately diminishes the explanatory power of many current approaches to linguistics because they are purposely speculative. Although

speculations may enrich our knowledge, too, testability is superior in areas where it is an alternative to speculation. A synchronic description of a living language which our book wants to offer is one of such areas.

Axiomatic Functionalism is not the only non-speculative approach. Its adoption has also resulted from the conviction that it is capable to fulfill the goal of the linguistic science, that is, to explain how communication is achieved by languages, and that the explanation and the description it allows for are better than those offerred by other approaches. The property of being better is hard to measure, but if we agree that consistency (both internal and with data), exhaustiveness and simplicity are evaluation criteria of a good theory and a good description (an idea to be elaborated in §2.1), then the selection of Axiomatic Functionalism can be justified also from this respect. Its originators formulated it with these criteria in mind.

In particular, Axiomatic Functionalism is capable of coping effectively with the problem of the distribution and combinations of phonemes. In 1968 Mulder introduced a model he called *distributional unit* (= DU) the purpose of which was to account for how formal means are combined and arranged in Peking Chinese (Mulder 1968). Enabling a novel approach to phonotactics and a more efficient description of these phenomena, it has since been applied to various languages such as Quechua (Howkins 1972), Yulu (Gabjanda 1976), English (El-Shakfeh 1987), Sudanese Arabic (Dickins 2007), French or Russian (Rastall 1993). Although it may appear similar to models later used by other approaches, for example syllable templates, it is not preordained to apply to bundles of sounds known as syllables, even though the syllable is traditionally viewed as the most effective domain of distribution. The combinability of formal means may be more complex in some languages; by acknowledging this, Axiomatic Functionalism has become more effective in explaining phonotactics. The fundamental idea of the DU is simple—it is a set of positions which can be occupied or empty—, but the exact shape of the DU derives from the combinability of formal means in the language under description. And its purpose is always the same: to provide means for explaining the way formal semiotic entities are combined and arranged in the most exhaustive and comprehensive way.

Although utterances are in their essence continuous non-discrete chunks of sounds, it makes reason to interpret them as sequences of phonemes. Within such sequences each phoneme stands in a position of the DU. Whereas phoneme sequences are in principle unlimited, the number of positions and the number of phonemes standing in them are finite. By setting up these positions and their memberships, it is possible to explain regularities of the occurrence and the combinability of formal communicational means. The recourse to positions is not new in linguistics, but the axiomatic-functionalist approach is specific in two

respects. First, positions may be empty, but an empty position is not a mere nothingness; it is a simple way to explain why some formal means may be omitted from utterances with the result being still an admissible structure in the given language. Second, Axiomatic Functionalism recognizes that the syntagmatic difference between positions, which express distributional and combinatory properties of phonemes, may become irrelevant just as it recognizes that the paradigmatic difference between distinctive features, which express oppositional properties of phonemes, may be become irrelevant. The paradigmatic irrelevancy is sometimes recognized in phonological theories under the name of neutralization, but the syntagmatic irrelevancy has rather been neglected. Axiomatic Functionalism provides ways to account for both types of irrelevancy, which results in a more detailed and less redundant description of the use and the combinability of formal means in communication.

## 1.2 Organization

We present here only the results of our analysis of Czech. Even though it may seem to be done in successive steps, the sequence is only a matter of presentation. A phonological description rises (and falls) in its entirety, and it is only for practical reasons that some areas are dealt with before others. In the course of our analysis we will establish for Czech the inventory of 33 phonemes and the DU of nine positions along which the phonemes are distributed. We will explain under what circumstances particular phonemes occur in those positions, how they combine with one another and what function they fulfill within such combinations. In combinations some phonemes must be obligatorily present, while others are optional. The former will be called *nuclear phonemes*, and the latter *peripheral phonemes* (traditionally called *vowels* and *consonants*, but we will make a finer classification of phonemes in Czech). In particular, our work will provide a comprehensive description of the occurrence and the combinability of nuclear and peripheral phonemes, both with each other and within their own class. The combinability of peripheral phonemes is especially rich in Czech. They form more than 500 distinct combinations within a DU, and many other combinations are possible across two DUs. The latter combinations are not dealt with here, but the DU will be established so as to have such a structure that the combinability of phonemes within larger complexes is derivable from and conditioned by the combinability of phonemes within a single DU.

The distribution of phonemes in Czech has already been described by Kučera (1961) upon the syllable. We offer an alternative analysis, but much more importantly, we provide a more detailed account of this topic. Combina-

tions of the Czech phonemes have also been described, but again besides being an alternative account, our work provides a more comprehensive treatise. Moreover, it deals with questions not addressed before such as whether the occurrence of some phonemes or their combinations before the nucleus of the DU constraints the occurrence of phonemes or their combinations after the nucleus. Another topic that has been insufficiently dealt with for Czech is the question of potentially possible combinations, that is, the question of recognizing phoneme combinations which are not actually attested, but which could function as potential forms of meaning-expressing means in Czech. The DU is as a device upon which this potentiality can be formalized and expressed.

The book is organized as follows. The rest of this introductory chapter briefly introduces the Czech language mentioning sources and data of our analysis. It is concluded with an outline of other works on the phonotactics of this language. Chapters 2, 3, 4 and 5 provide a necessary preamble to the main topic of this work. Chapter 2 *Outline of the Theory* lays out the theoretical and methodological background. In chapter 3 *Phonematics of Czech* we discuss the inventory of phonemes, and chapter 4 *Neutralization* then considers under what circumstances differences between the phonemes are valid and when they are not. The next chapter called *Prolegomena to the Phonotactic Analysis* deals with several issues relavant to the phonotactic analysis such as the analysis of phonological forms into sequences of DUs. It also explains why forms built of several phonotagms are not analyzed here because they belong to para-phonotactics where they must be considered together with features like accent, and how the occurrence of so-called non-syllabic prepositions will be dealt with.

Chapters 6 and 7 discuss the DU. The former explains the motivation behind this notion and its role as a domain of distribution of phonemes, while the latter applies it to Czech. We establish a unit of nine positions, determine what phonemes occur in them, and consider under which circumstances the differences between these positions become irrelevant. The next chapter called *Phonotagms* continues in the application and explication of the DU. Phonotagms are instances of the DU where every position of the DU is either occupied by a phoneme or empty. Two types of phonotagms are recognized: major-type phonotagms and minor-type phonotagms. The latter will be introduced as a special but limited group of phoneme combinations which may accompany major-type phonotagms.

Chapters 9, 10, 11 and 12 deal with the occurrence of phonemes within the positions of the DU. A position can be viewed as a sum of distributional and combinatory potentials of phonemes which are unique for each position. Sets of phonemes occurring in a given position are introduced in chapter 9 *Position Classes* where the idea of collocational restrictions is also explained. These are statements expressing limitations of phoneme occurrences. Chapter 10 *Distribu-*

*tion in the Nuclear Context* discusses distributional and combinatory properties of nuclear phonemes, that is, of vowels and of /r/ and /l/. The ability of the latter to function as nuclear entities is reviewed here, too. Chapter 11 *Distribution in the Pre-nuclear Context* and chapter 12 *Distribution in the Post-nuclear Context* examine properties of peripheral phonemes before and after the nucleus.

The topic of chapters 13, 14, 15 and 16 is combinations of phonemes in Czech. The first one called *Properties of Pre-nuclear and Post-nuclear Combinations* discusses the length, phonematic constituency, reducibility and resolvability of combinations of peripheral phonemes. They are inventoried and classified to subtypes in the next two chapters (chapter 14 *Pre-nuclear Combinations* and chapter 15 *Post-nuclear Combinations*). Finally, chapter 16 considers the ability of the nuclear phonemes to combine with the non-nuclear ones and the ability of the pre-nuclear phonemes to co-occur with the post-nuclear ones.

The last chapter (17) concludes this work. Instead of summarizing it, the chapter compares our analysis with the only phonotactic analysis of Czech worked up by Henry Kučera (1961). Kučera's analysis has its weak as well as strong points, and we show that the weaknesses overweight the strengths. However, the strengths can be taken advantage of, and we present an alternative analysis of Czech phonotagms inspired by Kučera's approach. Besides that, the final chapter shows how the DU can be used for predicting possibilities of the Czech phonotactics by means of so-called freight-yard schemes.

The book is supplemented with several appendices. Appendix A summarizes collocational restrictions introduced during the analysis. Appendix B provides a complete listing of pre-nuclear and post-nuclear combinations found in Modern Standard Czech. Appendix C provides a list of potential pre-nuclear and post-nuclear combinations. Finally, appendix D illustrates the distribution of nuclear phonemes, supplementing chapter 10.

## 1.3 Modern Standard Czech

Modern Czech is a West Slavic language spoken by the majority of inhabitants of the Czech Republic at the beginning of the 21st century and during the 20th century. By Standard Czech we mean a variety of Modern Czech which has been officially codified and which has a special status among all forms of Czech. In the local tradition the variety is known as *spisovná čeština*,[1] but we prefer the term *Standard Czech* without making any distinctions between them. Therefore, we describe one particular variety of Modern Czech, not Modern Czech as a

1 Although the term means "literary Czech", it is not confined to literature or even to written texts. It has a spoken form, too.

whole. Although such a description is possible in principle, it makes sense to view the language as a conglomeration of varieties with several intermixed systems. Standard Czech is one of such systems, one which is to a great extent clearly delimited by the codification. But this does not mean that Standard Czech is a closed set of linguistic items. In the course of time it may be enriched by new items or lose others. Still, it is relatively stable. Its pronunciation, grammar and vocabulary are well described in various handbooks, grammar books and dictionaries. Standard Czech serves as the norm for the other varieties of Czech, and it is in fact what a layman equates with *the* Czech language, in its purest form, even though it is not the variety normally used in everyday parlance.[2] The other varieties are judged against Standard Czech. The description of its phonotactics can serve as a point of reference for similar descriptions of the other varieties.

Once various varieties of Czech are recognized, they can be treated as separate communication systems differing from each other. The phonology of Standard Czech will not be the same as that of a non-standard variety. The latter may contain a different phoneme inventory as well as a different phonotactics. For example, Standard Czech does not allow word-final combinations of a fricative and /ř/, but such a combination exists in non-standard Czech: cf. *otevř* which is a non-standard variant of *otevři* "open (imper. sg.)". We will not also find there word-final combinations in which the velar occlusive is preceded by another occlusive, though again such combinations are encountered in non-standard Czech: cf. *nedotk'* which is a non-standard variant of *nedotkl* "he did not touch".

The pronunciation of Standard Czech has been codified and described in the official handbook *Výslovnost spisovné češtiny I* ("Pronunciation of literary Czech I" = VSČ, 2nd edition 1967) and its second volume *Výslovnost spisovné češtiny* (= Romportl et al. 1978).[3] It has later been outlined also in Hůrková (1995), Palková (1997), Krčmová (2008) and Zeman (2008). Several detailed phonetic descriptions of this variety exist as well, the most comprehensive and the most recent ones being Hála (1962) and Palková (*op. cit.*).

---

2 The language situation in the Czech Republic can be described as a diglossia between Standard Czech used in official situations (both in the spoken and written form) and so-called Common Czech ("obecná čeština") used in unofficial situations (cf. Bermel 2010). Common Czech derives from the dialect in and around the country's political and cultural center, Prague, and is mixed, mostly in Moravia, with other dialects. Standard Czech was modeled in the 19th century on the basis of older Czech literature. The codifiers ignored, for example, some sound changes which took place as early as 15th century and which are reflected in Common Czech. See Kučera (1961: 14 and references therein).

3 The book has the same title like the 1967 one except for the number I. Romportl et al. (1978) focuses on pronunciation of foreign words only.

Phonetic and orthoepic descriptions of Standard Czech are not by themselves sufficient if we aim to account for its phonotactics. Phonotactic units occur within phonological forms of meaningful units, hence if we want to describe the phonotactic structure of Czech, we must have in the first place a database of these meaningful units. We will focus on the phonotactic structure of words as the minimum syntactic entities. Most words are listed in dictionaries of Standard Czech the most important of which are *Slovník spisovného jazyka českého* ("Dictionary of the literary Czech language", 1960–1971, = SSJČ), and more recent *Slovník spisovné češtiny* ("Dictionary of literary Czech", 3rd edition 2008, = SSČ). The most comprehensive dictionary is *Příruční slovník jazyka českého* ("Concise dictionary of the Czech language", 1935–1957, = PSJČ), but it does not contain just words recognized as standard, but also words which are archaic or which have been created *ad hoc* by literary writers. If a word, and hence a certain phonotactic property, occurs only in PSJČ, it is explicitly acknowledged because there may be some disputes as to whether it belongs to the standard variety and whether the phonotactic peculiarity is a feature of this variety.

PSJČ and SSJČ have been digitalized and made available online (see the references at the end of this book) under the auspices of the Institute of the Czech Language of the Academy of Sciences of the Czech Republic.[4] The institute also maintains an online glossary of PSJČ, SSJČ, SSČ and other dictionaries of Czech as well as two unpublished lexical databases, *Heslář lexikálního standardu* ("Glossary of the lexical standard"), and *Heslář novočeského lexikálního archivu* ("Glossary of the New Czech lexical archive", = NLA). The second is particularly important because it provides a huge collection of words excerpted from literary texts. Some of them, and consequently some phoneme combinations, are attested only there. They are usually hapaxes legomena or markedly dialectal or archaic words, and although they are sometimes mentioned in our analysis, they are not the primary source of our data.

Besides the dictionaries and databases of the "normal" Czech vocabulary, we have also examined Czech place names and proper names because they contain phoneme combinations not attested elsewhere. By assumption, official names are part of Standard Czech and provide another piece of evidence of its potentials. For example, the word-final combination /jř/ is not found in the "normal" vocabulary, but is found in proper names, namely in the surname *Kejř*. Similarly, the word-initial combination /břv/ is attested only in the place names *Břve* and *Břvany*. It is always made clear when some phonotactic property is found exclusively in such words. Finally, it should be mentioned that no systematic

4 The Institute is the official workplace to look after the Czech language. It is also the publisher of some of the mentioned pronunciation handbooks and of grammar books.

analysis of onomatopoeic expressions has been attempted. Their phonotactic structure is sometimes deviant in comparison with the structure of "normal" vocabulary. These include expressions such as *bzz*, *kšc*, *pf* or *pšt*. Though they belong to Czech, it is better to view them as an auxiliary semiotic system with its own phonotactic rules.

Dictionaries and databases generally list words in their uninflected form. Yet many phoneme combinations are exclusive to inflected forms (such as /jSK/ in *vojsk*, gen. pl. of *vojsko*). Thus, we have also taken heed of rules of inflection and derivation; they are sufficiently described in various reference books such as *Česká mluvnice* ("Czech grammar", 6th edition, = Havránek – Jedlička 1986), *Mluvnice češtiny* ("Grammar of Czech", particularly the second volume, 1986, abbreviated MČ2), *Příruční mluvnice češtiny* ("Concise grammar of Czech", 2nd edition 1996, = PMČ), and *Mluvnice současné češtiny 1* ("Grammar of contemporary Czech 1", 2010, = MSČ). Another valuable resource is *Internetová jazyková příručka* ("Internet language reference book") available online and again maintained by the Institute of the Czech Language of the Academy of Sciences of the Czech Republic. We have also used *Český národní korpus* ("Czech National Corpus"), the synchronic base SYN, but this source does not contain only texts from Standard Czech and should be treated with caution.

## 1.4 Previous phonotactic descriptions of Czech

This work is not the first or the only phonotactic description of Modern Czech. Probably the earliest comprehensive treatise on phoneme combinations in Modern (or pre-Modern) Czech is Hattala (1870) which discusses word-initial consonantal combinations sorted into several classes. Some of the combinations mentioned are dubious or occurring in words probably no longer current in the Czech of Hattala's time, and some are missing (such as /vj/ in *věk* which is not a consonantal combination for him). Other descriptions of phoneme combinations stemmed from attempts to create shorthand for Czech. Summarizing these attempts, Trnka (1937, in particular pp. 31–41) discusses several combinatory restrictions on phonemes, and the ratio of occurrences of consonants with vowels and with other consonants; he also provides lists of attested consonantal combinations in a single morpheme and across morphological boundaries.

Phoneme combinations are discussed in some early articles by Mathesius (see below), but a properly phonological analysis of word-initial and word-final consonantal combinations was published as late as 1972 by Novotná (1972). Though her primary focus was spoken Standard Czech, and although she excludes combinations occurring in allegedly foreign-origin words, her work was